

GENETICS

Chlamydial contribution to anaerobic metabolism during eukaryotic evolution

Courtney W. Stairs^{1*}, Jennah E. Sharamshi^{1*}, Daniel Tamarit^{1,2,3*}, Laura Eme^{1,4}, Steffen L. Jørgensen⁵, Anja Spang^{1,6}, Thijs J. G. Ettema^{1,2†}

The origin of eukaryotes is a major open question in evolutionary biology. Multiple hypotheses posit that eukaryotes likely evolved from a syntrophic relationship between an archaeon and an alphaproteobacterium based on H₂ exchange. However, there are no strong indications that modern eukaryotic H₂ metabolism originated from archaea or alphaproteobacteria. Here, we present evidence for the origin of H₂ metabolism genes in eukaryotes from an ancestor of the Anoxychlamydiales—a group of anaerobic chlamydiae, newly described here, from marine sediments. Among Chlamydiae, these bacteria uniquely encode genes for H₂ metabolism and other anaerobiosis-associated pathways. Phylogenetic analyses of several components of H₂ metabolism reveal that Anoxychlamydiales homologs are the closest relatives to eukaryotic sequences. We propose that an ancestor of the Anoxychlamydiales contributed these key genes during the evolution of eukaryotes, supporting a mosaic evolutionary origin of eukaryotic metabolism.

INTRODUCTION

Hydrogen transfer is a central aspect of many syntrophic relationships across the tree of life and represents the basis for various hypotheses on the origin of eukaryotes (1–5). The current prevailing theory for the origin of eukaryotes involves the symbiosis of an archaeon and an alphaproteobacterium. Although there are multiple proposals for the nature of the symbiosis, many hypotheses suggest that eukaryotic life evolved in an oxic-anoxic transition zone (OATZ) and that syntrophic interactions may have played a key role (1–5). By studying syntrophies in present-day OATZs, such as those found in marine sediments, we may be able to reconstruct interactions that were critical for eukaryogenesis.

Marine sediments are typically characterized by stable conditions and a well-defined OATZ (6, 7). The anoxic layers immediately below this transition zone often host representatives of the Asgard archaea, members of a superphylum of diverse lineages that are thought to represent the closest living relatives of the archaeal ancestor of eukaryotes (8, 9). Asgard archaea are predicted to have distinct metabolic features, with several members likely dependent on syntrophic interactions for growth based on H₂ exchange (3, 4). H₂ metabolism is widespread in prokaryotes, including Asgard archaea (3, 10), and H₂ production and/or utilization is featured in various eukaryogenesis models (1–5). However, the absence of genes encoding eukaryotic-type H₂ production [i.e., [FeFe]-hydrogenases (HYDA)] in most modern archaea (including Asgard archaea) or alphaproteobacteria has made it difficult to pinpoint the origin of this metabolism in eukaryotes.

The ability to produce H₂ has been demonstrated in numerous eukaryotes known to experience transient or permanent anoxia (11, 12). H₂-producing eukaryotes often have drastically different mito-

chondria compared to aerobic eukaryotes. These organelles are collectively referred to as “mitochondrion-related organelles” (MROs) and include H₂-producing mitochondria, hydrogenosomes, and mitosomes. These organelles can vary in their complement of respiratory complexes (complexes CI to CV) ranging from complete retention in the H₂-producing mitochondria of *Acanthamoeba castellanii* to complete loss in the hydrogenosomes of *Trichomonas vaginalis* [Fig. 1; see (11, 12)]. In eukaryotes and some anaerobic bacteria, pyruvate oxidation is coupled to H₂ production by the concerted action of pyruvate:ferredoxin oxidoreductase (PFO) and HYDA, ultimately generating H₂ and acetyl-coenzyme A (CoA). The proper assembly of the Fe-S cluster of HYDA relies on three maturase proteins (HYDE, HYDF, and HYDG). In some bacteria, HYDA functions with two other proteins, HYDB and HYDC—homologs of the NUOF and NUOE subunits of CI of the respiratory chain, respectively—in a trimeric confurcating hydrogenase complex. This complex couples the favorable and unfavorable oxidations of ferredoxin (Fd; $E_0' \approx -450$ mV) and NADH (reduced form of nicotinamide adenine dinucleotide) ($E_0' \approx -320$ mV), respectively, to reduce protons ($E_0' \approx -420$ mV) (13). In some anaerobic and microaerophilic microbial eukaryotes, PFO-mediated acetyl-CoA production and HYDA-mediated H₂ production occur in the cytoplasm (14), plastid (15), and/or mitochondrial compartment (16, 17). While many of these eukaryotes have lost all genes that encode the respiratory chain (i.e., CI to CV), some have retained two nuclear-encoded genes of MRO-localized CI subunits (i.e., *nuoE* and *nuoF* genes). These two subunits can oxidize both NADH or Fd (18), suggesting that, within the MRO, they could interact with HYDA to form a trimeric electron-confurcating hydrogenase, analogous to the bacterial complex discussed above (13).

Unlike in bacteria, genes encoding proteins for H₂-producing metabolism (i.e., HYDA, HYDE-G, and PFO) are sparsely distributed across the tree of eukaryotes (11), and the origins of this pathway remain hotly debated: Was HYDA-mediated H₂ production present in the alphaproteobacterial endosymbiont that gave rise to the mitochondrion (1, 19), or was it acquired by different eukaryotic lineages via horizontal gene transfer (HGT) (20)? One obstacle for inferring the evolutionary origin of these genes is the overall lack of resolution

Copyright © 2020
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123 Uppsala, Sweden. ²Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6708 WE Wageningen, Netherlands. ³Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden. ⁴Unité d'Ecologie, Systématique et Evolution, CNRS, Université Paris-Sud, Orsay, France. ⁵Department of Earth Science, Centre for Deep Sea Research, University of Bergen, N-5020 Bergen, Norway. ⁶Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, and Utrecht University, NL-1790 AB Den Burg, Netherlands.

*These authors contributed equally to this work.

†Corresponding author. Email: thijs.ettma@wur.nl

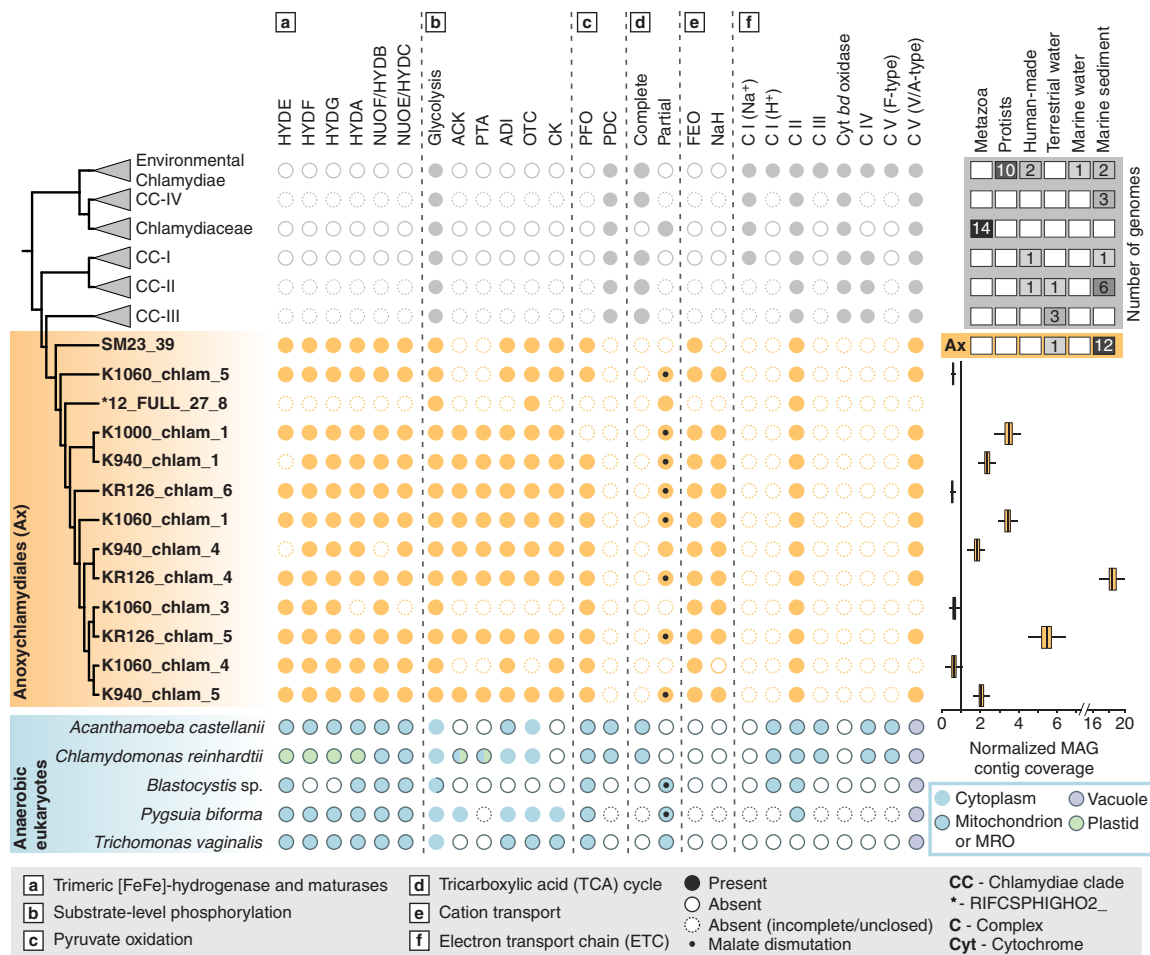


Fig. 1. Anoxychlamydiales and anaerobic eukaryotes use similar anoxic metabolic strategies. The presence of genes encoding components of the indicated energy metabolism pathways (left), encoded in the genomes of Chlamydiae (gray), Anoxychlamydiales (orange), and select anaerobic eukaryotes (blue), is indicated by a filled circle. For eukaryotic representatives, the subcellular location of each gene product or pathway is indicated by a filled blue circle (cytoplasm) or a filled circle with an outline (organelles). Outlined blue, green, and purple filled circles indicate the mitochondrion or MRO, the plastid (green), and the vacuole (purple), respectively (see legend). Eukaryotic representatives *A. castellanii* and *Blastocystis* sp. have hydrogen-producing mitochondria, *Pygusua biforma* and *T. vaginalis* have hydrogenosomes, and *Chlamydomonas reinhardtii* has anaerobically functioning mitochondria with plastidal hydrogen production. The number of genomes obtained from the indicated environmental sources for each Chlamydiae clade is shown in the top right. The boxplot on the lower right shows the normalized coverage of contigs from the corresponding Anoxychlamydiales MAGs obtained from Loki’s Castle marine sediments (data S1). HYDA-G, hydrogenase subunits A to G. ADI, arginine deiminase; OTC, ornithine transcarbamoylase; CK, carbamate kinase; PFO, pyruvate:ferredoxin oxidoreductase; PDC, pyruvate dehydrogenase complex; FEO, ferrous iron transport; ACK, acetate kinase; PTA, phosphotransacetylase; NaH, Na⁺:H⁺ antiporter. The Chlamydiae species tree is a representation of species relationships derived from Dharamshi *et al.* (28).

of individual protein phylogenies (21–23). Moreover, while the origins of most proteins related to aerobic metabolism in mitochondria [e.g., respiratory chain, pyruvate dehydrogenase, and some components of the tricarboxylic acid (TCA) cycle] display clear affinity to genes from extant relatives of the endosymbiont (i.e., modern alphaproteobacteria) (24, 25), the origins of all anaerobiosis-associated proteins are unknown. Previous phylogenetic analyses of these proteins failed to recover a clear consistent prokaryotic group as the sister lineage of the eukaryotic sequences (21–23). Here, we provide the first evidence for a specific bacterial donor of eukaryotic H₂ metabolism from an order, “*Candidatus* Anoxychlamydiales,” newly described here, that represents the first identified anaerobic members of the phylum Chlamydiae.

RESULTS AND DISCUSSION

Chlamydiae are described as obligate intracellular bacteria of eukaryotes that display varying levels of auxotrophy for cofactors, amino

acids, and nucleotides, and often import these compounds from their hosts (26). Although chlamydiae can acquire nucleotide triphosphates directly from their hosts using adenosine triphosphate (ATP)/adenosine diphosphate (ADP) transporters (27), they also conserve energy by glycolysis and aerobic respiration (26). With the retrieval of diverse and abundant chlamydial metagenome-assembled genomes (MAGs) from anoxic marine sediments (28) sampled near Loki’s Castle hydrothermal vent field (29), we have recently challenged the view that aerobiosis and an obligate eukaryotic intracellular lifestyle are unifying features of the phylum Chlamydiae (29).

Anoxychlamydiales are the first Chlamydiae shown to encode anaerobic metabolism

Using phylogenetic analyses of conserved single-copy marker proteins (28), we observed that 11 of the 24 Loki’s Castle chlamydial MAGs and two MAGs from estuary sediment and groundwater (30, 31) consistently form a distinct order hereafter referred to as Anoxychlamydiales

(Fig. 1). In our marine sediment samples, Anoxychlamydiales are dominant community members as determined by read coverage of both the conserved ribosomal protein operon and each MAG (Fig. 1, fig. S1, and data S1). Because these genomes appear to derive from cells that are actively replicating (28), and because all cultured chlamydiae can only replicate intracellularly inside eukaryotic hosts (26), we expected to find eukaryotes in these samples. However, the low number of eukaryote-derived sequences in these samples could not explain the observed high relative abundance of Anoxychlamydiales (Fig. 1 and fig. S1) (28). We therefore suspect that Anoxychlamydiales are not obligate intracellular symbionts of eukaryotes.

When examining the predicted metabolic potential of the Anoxychlamydiales, we observed notable differences in their gene content compared to other chlamydiae (Fig. 1, fig. S2, and data S1). Reconstruction of Anoxychlamydiales metabolism strongly supports the hypothesis that this group is composed of anaerobes, in contrast to previously identified members of the phylum Chlamydiae (figs. S2 to S4). Phylogenetic analyses revealed that many anaerobiosis-associated proteins found in Anoxychlamydiales are closely related to homologs from diverse groups of anaerobic prokaryotes (fig. S3 and data S2). Like other facultative and obligate anaerobic fermentative bacteria, the Anoxychlamydiales are predicted to produce ATP by substrate-level phosphorylation via glycolysis, the arginine deimination pathway, and the concerted action of acetate kinase (ACK) and phosphate acetyltransferase (PTA), resulting in the concomitant production of acetate (Fig. 1, figs. S3 and S4, and data S1). Some chlamydiae can generate a Na^+ gradient using CI (i.e., a Na^+ -transporting NADH dehydrogenase), allowing for Na^+ -driven synthesis of ATP via a V-type adenosine triphosphatase (ATPase) (Fig. 1) (32). While Anoxychlamydiales do not encode a Na^+ -transporting CI, they can likely still generate a Na^+ gradient using a $\text{Na}^+:\text{H}^+$ antiporter (NaH), the gene of which appears to have been acquired by HGT from anaerobic deltaproteobacteria (fig. S3F). We therefore suspect that these bacteria also have Na^+ -driven ATP synthesis via a V-type ATPase (Fig. 1 and figs. S3 and S4). We also identified a number of distinguishing anaerobiosis-associated characteristics of Anoxychlamydiales compared to other chlamydiae, e.g., ferrous iron import and metabolism (FeoB), and the absence of a complete respiratory chain and TCA cycle (Fig. 1).

The Anoxychlamydiales MAGs also encode a collection of metabolic modules that are common in anaerobic eukaryotes (Fig. 1, figs. S3 and S4, and data S1). Most unexpectedly, Anoxychlamydiales have the capacity for H_2 production, a feature not previously assigned to representatives of this phylum. In addition to encoding a *pfo* gene, most of these MAGs have a gene cluster encoding the [FeFe]-hydrogenase maturases (i.e., *hydE*, *hydF*, and *hydG*) and a second cluster encoding a trimeric [FeFe]-hydrogenase (i.e., *hydA*, *nuoE/hydC*, and *nuoF/hydB*) (fig. S6). The domain architecture of both the HYDA and NUOF/HYDB proteins (figs. S5 and S6) resembles those of other trimeric H_2 -producing hydrogenase systems in bacteria (10). As [FeFe]-hydrogenases are bidirectional enzymes, it remains to be assessed whether these chlamydial proteins produce or consume H_2 . However, considering that we could not identify genes typically involved in H_2 -consuming pathways (e.g., terminal electron acceptors or carbon fixation pathways; data S1), we find it likely that Anoxychlamydiales produce H_2 during acetogenic fermentation (Fig. 1, fig. S4, and data S1). Collectively, these features indicate that the Anoxychlamydiales might be H_2 and acetate producers that live syntrophically within a microbial consortium with H_2 and/or acetate consumers (e.g., hydrogenotrophic methanogens

and homoacetogens). Given their high relative abundance in anoxic marine sediments (Fig. 1 and fig. S1) and the importance of H_2 and acetate in the carbon cycle (33), these Anoxychlamydiales may play a previously unrecognized role in global biogeochemical cycling in anoxic environments. The punctate distribution of anaerobiosis-associated metabolism in other unclassified chlamydial lineages (data S1) suggests that further exploration of Chlamydiae in anoxic environments will reveal anaerobiosis to be more widespread within this phylum than is currently recognized.

Anoxychlamydiales encode eukaryotic-like hydrogen metabolism

To investigate the evolutionary history of the proteins mediating H_2 metabolism in Anoxychlamydiales and their relationship to eukaryotic homologs, we carried out phylogenetic analyses of each of the seven proteins involved in the PFO-HYDA system. In these phylogenies, the Anoxychlamydiales formed a sister clade to eukaryotes (HYDE, HYDF, and HYDG) or branched close to them (PFO), with high support (Fig. 2). The relationships between eukaryotes within the eukaryotic clades display patterns of both vertical (e.g., plastid-bearing lineages and chytrid fungi; see data S2) and horizontal inheritance (data S2). Furthermore, the topologies within the Anoxychlamydiales clades (data S2) are largely congruent with known organismal relationships (Fig. 1), suggesting vertical inheritance of these genes in this chlamydial order. This finding was unexpected because until now, there was no clear evidence for a consistent prokaryotic lineage branching sister to eukaryotes in those four protein phylogenies. For example, if we exclude the Anoxychlamydiales, the closest prokaryotic taxa to the eukaryotic sequences are non-Asgard archaea (PFO), firmicutes and aquificae (HYDE), or bacteria of mixed taxonomic assignment (HYDF and HYDG). The observation of the Anoxychlamydiales sequences representing a sister clade to (and not emerging from within) eukaryotes strongly suggests that Anoxychlamydiales did not receive these genes by HGT from eukaryotes but that, on the contrary, an Anoxychlamydiales ancestor might have contributed these genes to eukaryotes.

We next assessed the relationships between trimeric H_2 -producing hydrogenase components and related proteins of CI from Anoxychlamydiales, other chlamydiae, and eukaryotes. In phylogenetic analyses of both NUOE/HYDC and NUOF/HYDB, eukaryotic sequences form a clade with alphaproteobacterial sequences, a topology expected given their mitochondrial ancestry (25). The NUOE/HYDC and NUOF/HYDB Anoxychlamydiales sequences branch distantly from CI subunits of eukaryotes and other chlamydiae (fig. S6). In the HYDA phylogeny, there are two groups of eukaryotic sequences, both of which branch distantly from Anoxychlamydiales homologs (fig. S5). Contrary to other components of H_2 metabolism mentioned above (e.g., HYDE-G and PFO), HYDA appears to have different evolutionary origins in Anoxychlamydiales and eukaryotes. In HYDA, NUOE/HYDC, and NUOF/HYDB phylogenies, the closest sister taxa to the Anoxychlamydiales sequences are a collection of diverse and distantly related bacteria isolated from anoxic environments (34–36). The congruence of these phylogenies is mirrored by the conserved operon organization and domain structure of the bacterial sequences that branch closest to the Anoxychlamydiales (figs. S5 and S6). Collectively, this implies that the Anoxychlamydiales NUOE/HYDC and NUOF/HYDB subunits have a distinct evolutionary history compared to the homologous CI subunits from other chlamydiae. These findings also suggest that the *nuoE/hydC*, *nuoF/hydB*, and

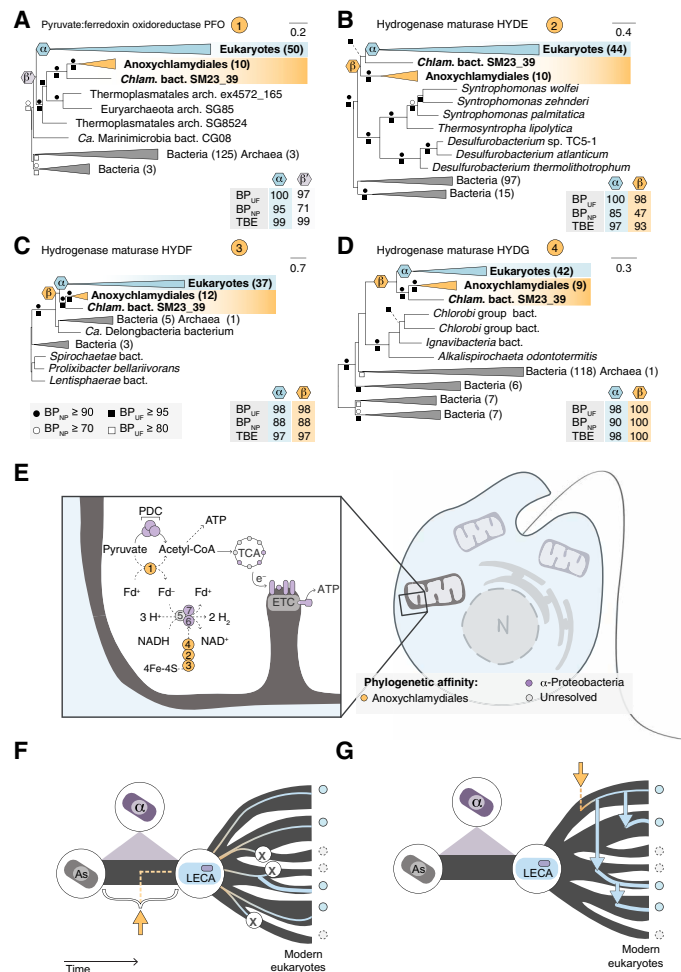


Fig. 2. Anoxychlamydiales genes are the closest prokaryotic relatives of eukaryotic genes encoding anaerobic metabolism. (A to D) Maximum likelihood (ML) phylogenies for H₂ metabolism proteins. Circles [nonparametric (NP)] and squares [ultrafast (UF)] summarize bootstrap support values (BP, bootstrap percentage) for each bipartition mapped onto the best-scoring ML phylogeny. BP_{UF}, BP_{NP}, and transfer bootstrap expectation (TBE) for monophyly of eukaryotes (α) and eukaryotes with Anoxychlamydiales (β) or Anoxychlamydiales and other taxa (β') are indicated. Eukaryotes and Anoxychlamydiales are shaded blue and orange, respectively (see data S1 for model parameters and alignment features and data S2 for full phylogenies). (E) The origin of mitochondrial metabolism in eukaryotes is an evolutionary mosaic. In aerobic mitochondria (solid lines), pyruvate is oxidized by PDC to acetyl-CoA, which is fed into the TCA cycle to produce reducing equivalents for the ETC that fuels ATP synthesis by oxidative phosphorylation. In MROs (dashed lines), pyruvate is oxidized by PFO (1) to acetyl-CoA, which is used for ATP synthesis by substrate-level phosphorylation (12). Hydrogen is produced by a trimeric confurcating hydrogenase (5 to 7) using electrons from Fd⁻ and NADH. The 4Fe-4S cluster of HYDA (5) is assembled by maturases (2 to 4). In this hypothetical cell, colors represent the proposed origin of each component [data S2 and figs. S5 and S6; (21–23)]. 1 to -5 function in the plastids of some algae (15). (F and G) Scenarios for the timing of acquisition of hydrogen metabolism relative to major events in eukaryotic evolution. Relative timings of gene acquisitions from an Anoxychlamydiales ancestor (orange arrows) (F) before or immediately after the emergence of last eukaryotic common ancestor (LECA) or (G) after the radiation of eukaryotes, mediated by HGT into and between eukaryotes (blue arrows). Uncertainty regarding the timing of mitochondrial integration is depicted with a purple triangle. Gene losses are shown with an “X.” Ancestral Asgard archaeon (As; gray) and the alphaproteobacterial ancestor of the mitochondrion (α; purple).

hydA genes have been acquired independently from the hydrogenase maturases and *pfo* (Fig. 2).

Anoxychlamydiales ancestor contributed to the mosaic origins of eukaryotic metabolism

To date, the origin of the proteins mediating H₂ metabolism in eukaryotes has been elusive, and it was unclear whether any of these proteins were present in the last eukaryotic common ancestor (LECA). Two prevailing hypotheses for the origin of this pathway suggest that modern H₂ metabolism derives from HGT events into and between eukaryotes or that this pathway was present in the alphaproteobacterial endosymbiont that gave rise to the mitochondrion (1, 20). This work identifies an ancestor of the Anoxychlamydiales as the likely donor of genes coding for components of H₂ metabolism in eukaryotes, as opposed to this metabolism having been derived from the alphaproteobacterial ancestor of the mitochondria (1, 37). The timing of the acquisition of these genes from an ancestor of Anoxychlamydiales relative to major events in eukaryotic evolution remains unclear (Fig. 2, F and G). However, the lack of evidence for genes encoding eukaryotic-type H₂ metabolism in modern representatives of archaea (including the Asgard archaea) and alphaproteobacteria strongly suggests that this metabolism was acquired after the emergence of the first eukaryotic common ancestor (FECA). The punctate distribution of H₂ metabolism across the tree of eukaryotes suggests that vertical inheritance (potentially dating back to LECA), differential loss, and HGT between eukaryotes have played a role in the evolution of these genes in eukaryotes. These findings highlight the contribution of genes originating from outside the archaeal host-ancestor and alphaproteobacterial mitochondrial-ancestor during eukaryotic evolution, a possibility that has been explicitly part of several eukaryogenesis scenarios (2, 5, 38). Future investigations might reveal additional chlamydial homologs, or those from other prokaryotes, that are more closely related to the eukaryotic proteins than those from Anoxychlamydiales, and additional data of this sort could allow for the refinement of evolutionary scenarios underpinning the origin and early evolution of eukaryotes.

Chlamydiae have previously been invoked in other hypotheses detailing major evolutionary transitions in eukaryotes, such as the emergence of plastids (39, 40). In addition, Pittis and Gabaldon (38) found evidence for the late acquisition of genes encoding mitochondrial components (i.e., of alphaproteobacterial origin) in the eukaryotic lineage before LECA, alongside a similarly timed acquisition of genes from Chlamydiae and the related phylum Verrucomicrobia. We therefore propose that a chlamydial ancestor contributed key components of H₂ metabolism during eukaryogenesis and potentially before the diversification of eukaryotes. The acquisition of an [FeFe]-hydrogenase system and its integration into the metabolic networks of an early proto-eukaryote may have allowed for the loss of the presumed archaeal and/or alphaproteobacterial H₂ metabolism of the host cell. While H₂ exchange is featured in many eukaryogenesis models, our work provides the first evidence for a hypothesis whereby the archaeal-type H₂ production was replaced by a non-alphaproteobacterial H₂ system during eukaryogenesis.

Conclusion

Extant chlamydiae (e.g., Anoxychlamydiales) and the closest prokaryotic relatives of eukaryotes (i.e., Asgard archaea) inhabit similar anoxic environments, in close vicinity to OATZs where eukaryogenesis is thought to have occurred. Although we do not know whether

Anoxychlamydiales and Asgard archaea interact, their co-occurrence within the same microbial communities suggests that HGT between their ancestors could have facilitated the transfer of genes related to H₂ metabolism. However, questions regarding the origins of the remaining components of H₂ production and other anaerobic strategies in eukaryotes remain open. The polyphyletic relationship of eukaryotic HYDA implies multiple distinct origins and suggests that, if HYDA was in LECA, it has been replaced numerous times in the ancestors of modern HYDA-containing eukaryotes. In contrast, eukaryotic HYDA maturases and PFO have a single evolutionary origin. The exact timing of these contributions and their importance for the origin of eukaryotes remain as avenues for future exploration. The mosaic origin of various mitochondrial components, including respiration and H₂ production, strongly suggests that prokaryotes from at least three major phyla of life—Alphaproteobacteria, Chlamydiae, and Asgard archaea—contributed to the emergence of the first eukaryotic cell.

MATERIALS AND METHODS

Metagenomic relative abundance

To investigate the relative abundance of Anoxychlamydiales among microbial community members in the marine sediment samples, we surveyed the four metagenomes for “RP15 contigs,” i.e., contigs encoding at least 5 of 15 ribosomal proteins found in an operon conserved across prokaryotes. These were identified in the metagenomic assemblies (GS08_GC12_126, GS10_PC15_940, GS10_PC15_1000, and GS10_PC15_1060) using a previously described workflow (41). Contig coverage was estimated by mapping sequence reads from each metagenome to corresponding assembled contigs using Bowtie2 (42). Contigs longer than 20 kb were split into 10-kb fragments before coverage estimation. Averaged coverages of RP15 contigs from each marine sediment metagenome were then compared (fig. S1 and data S1).

Maximum likelihood (ML) phylogenies of concatenated proteins extracted from RP15 contigs were inferred for each marine sediment sample (GS08_GC12_126 in fig. S1, all in data S2), alongside a set of phylogenetically diverse prokaryotic reference taxa (41, 43) using RAxML version 8.2.4 (44) under the PROTCATLG model of evolution with 100 rapid bootstraps.

We normalized the coverage of contigs (and contig fragments) from each Loki’s Castle Anoxychlamydiales MAG to the average metagenome coverage (data S1). This allowed us to compare MAG abundance between samples and indicated the magnitude of over- or underrepresentation of each MAG in the corresponding metagenome (Fig. 1).

Annotation and orthology clustering

Protein sequences were annotated according to the strategy laid out by Dharamshi *et al.* (28). In short, annotations include: top hits against National Center for Biotechnology Information nonredundant (*nr*) protein database using DIAMOND aligner v0.9.19.120 (45) blastp (“--more-sensitive”) alongside taxonomic classification of sequences [lowest common ancestor (LCA) algorithm, “-f 102”], assignment of protein domain annotations using InterProScan (46) version 5.22-61.0 [e.g., Protein Families (Pfam) (47) and InterPro (IPR) (48) domains], and assignment to both root-level “-d NOG” and bacterial-level “-d BACT” non-supervised orthologous groups (NOGs) using eggNOG-mapper (49) with the eggNOG database (50).

The presence of KEGG (Kyoto Encyclopedia of Genes and Genomes) (51, 52) modules connected to various metabolic pathways

related to energy generation and electron transfer (Fig. 1 and fig. S4) were assessed across Chlamydiae species representatives (data S1) through the assignment of KEGG orthology (KO) numbers by GhostKOALA (53). Pfam domains (47) allowed the identification of several additional proteins of interest (data S1), including CIII of the electron transport chain (PF00033). The presence of specific proteins across Chlamydiae included in phylogenetic trees (see below) is also indicated in data S1.

The presence of specific pathways related to energy generation found across Chlamydiae was also assessed in several anaerobic eukaryote representatives (e.g., *A. castellani*, GCF_000313135.1; *T. vaginalis*, GCF_000002825.2; *Chlamydomonas reinhardtii*, GCF_000002595.1; *Blastocystis* sp. ST 1, GCA_001651215.1; and *Pygysuia biforma*, GCRY00000000.1) using GhostKOALA (53) or manual inspection (see data S1).

Gene content unique to or enriched in Anoxychlamydiales relative to other chlamydiae was assessed by mapping protein sequences to NOGs. eggNOG-mapper (49) version 4.5 was used to map all protein sequences from Chlamydiae species representatives (data S1) to root-level NOGs (“-d NOG”). NOGs unique to Anoxychlamydiales in comparison to other chlamydiae and found across the group (in at least five members) are outlined in fig. S2 (data S1) and are ordered by Clusters of Orthologous Genes (COG) category (54). Those enriched in Anoxychlamydiales, in at least five members and up to three other chlamydiae, can be found in data S1.

Phylogenetic analysis of genes encoding anaerobiosis-associated proteins

We selected orthologous groups related to anaerobiosis that were found in the Anoxychlamydiales MAGs and typically absent from other chlamydiae. To generate the phylogenetic datasets, we used each of the Anoxychlamydiales sequences as a query against GenBank *nr* (January 2019), the Marine Microbial Eukaryote Transcriptome Sequencing Project [MMETSP; (55)], and other sequencing projects (22, 56) using BLAST (57) to retrieve the best 2000 hits with an *e* value lower than 1×10^{-5} . For HYDA phylogenies, we retrieved 2000 additional hits using the large HYDA domain of *T. vaginalis* (XP_001322682.1), *Nyctotherus ovalis* (CAA76373.1), and *Thalassiosira pseudonana* (XP_002295160.1). All HYDA sequences were classified with HYDDDB (58), and only group 1A HYDAs were retained. HMMer version h3.1b2 (www.hmm.org) (59) and the hidden Markov model (HMM) for the large HYDA domain (PF02906) were used to identify and extract only the large HYDA domain from each sequence for further analysis. For all datasets, taxonomy was assigned to each sequence using ETE version 3.0.0 (60), and CD-HIT version 4.6 (61) was used to reduce the dataset at the 80% amino acid sequence identity level.

MAFFT v7 (a multiple alignment program for amino acid or nucleotide sequences) (62) using the “--auto” flag was used to build initial alignments, and BMGE (Block Mapping and Gathering with Entropy) version 1.12 (63) was used to mask ambiguously aligned residues (i.e., entropy scores greater than 0.6) using the BLOSUM30 matrix. Initial trees were generated using FastTree version 2.1.9 SSE3 (64), distantly related sequences were removed, and clades of closely related organisms (e.g., same class) were trimmed to contain only representative sequences. With the final datasets, T-COFFEE v11 using M-Coffee mode (65, 66) was used to generate a consensus alignment of the eight default alignment softwares. Alignments were manually refined when necessary. Ambiguously aligned residues were

removed using BMGE version 1.12 (63) with the indicated thresholds (data S1).

ModelFinder (67) implemented in IQ-TREE v 1.6.10 (68) was used to select the most appropriate evolutionary model [including the C-series mixture models (69)] for each protein (data S1). A total of 1000 ultrafast (70) and SH-like approximate likelihood ratio test (SH-aLRT) (71) bootstraps were calculated using the best-scoring evolutionary model for each protein. The resulting ML tree was used as the guide tree for rapid approximation of posterior mean site frequency (PMSF) of the C-series of mixture models (72) and 100 non-parametric bootstraps. For HYDE, HYDF, HYDG, and PFO phylogenies, in addition to the nonparametric bootstraps discussed above, we calculated the transfer bootstrap expectation (73) implemented in IQ-TREE v 2.0 (74). Rogue taxa were selected with RogueNaRok (75) using the nonparametric bootstrap trees and ML tree from these analyses under the default settings. Taxa identified as rogue were removed from the original dataset, and the alignments and phylogenies were recomputed as described above.

In initial phylogenetic trees for HYDF (see “HYDF family” in data S2), a clade composed of eukaryotes, Anoxychlamydiales, and additional mixed prokaryotic taxa (HYDF clade I) resolved on a long branch apart from other prokaryotic sequences (HYDF clade II), indicating two paralogs. We therefore also analyzed HYDF clade I as described above (Fig. 2 and data S2).

Topologies constraining eukaryotic sequences with various prokaryotic clades were tested and are outlined in data S1. In all cases, the branching of Anoxychlamydiales sister to eukaryotes could not be rejected. Topology tests were performed for HYDE, HYDF, HYDG, and PFO (data S1) using IQ-TREE v 2.0 (74). Briefly, ML trees of the constrained topologies were calculated using the “-g” option under the model of evolution previously selected by ModelFinder (67). The approximately unbiased (AU) test (76) was performed on the ML constrained trees and the 100 bootstrap trees (“-au,” “-z,” “-n 0,” “-zb 10000,” and “-zw”) and with model parameters estimated using the ML tree (“-te”).

Summarized and full phylogenies can be found in Fig. 2; figs. S3, S5, and S6; and data S2. For each gene, the sequence dataset (*.fasta), sequence alignment (*.tcoffee.fasta), BMGE masked alignment (*.bmgc.fasta), PMSF phylogeny (*.pmsf.tre), and ultrafast bootstrap phylogeny (*.uf.tre) are provided in the following data repository (Figshare DOI: 10.6084/m9.figshare.12387980). Additional datasets with the TBE values (*.tbe.tree) and rogue taxa removed (*.RR_removed.*) for HYDE, HYDF, HYDG, and PFO are also provided. Accession numbers including “CAMPEP” or “CAMNT” derived from the MMETSP database (55).

Data visualization

In R v.3.2.2 (R Development Core Team, 2008), the package ggplot2 (77) was used for plots in Fig. 1 and fig. S1, while the package genoPlotR (78) was used to generate a synteny plot for visualization of the genome organization of HYDA, NUOE, and NUOF (fig. S6). iTOL (79) was used to visualize protein domains and to map them to the corresponding sequences in phylogenetic trees (figs. S5 and S6). Figtree v1.4.2 (80), iTOL (79), and Adobe Illustrator were used to visualize and edit phylogenetic trees.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/35/eabb7258/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

- W. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- D. Moreira, P. Lopez-Garcia, Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
- A. Spang, C. W. Stairs, N. Dombrowski, L. Eme, J. Lombard, E. F. Caceres, C. Greening, B. J. Baker, T. J. G. Ettema, Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* **4**, 1138–1148 (2019).
- H. Imachi, M. K. Nobu, N. Nakahara, Y. Morono, M. Ogawara, Y. Takaki, Y. Takano, K. Uematsu, T. Ikuta, M. Ito, Y. Matsui, M. Miyazaki, K. Murata, Y. Saito, S. Sakai, C. Song, E. Tasumi, Y. Yamanaka, T. Yamaguchi, Y. Kamagata, H. Tamaki, K. Takai, Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020).
- P. López-García, D. Moreira, The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).
- P. N. Froelich, G. P. Klinkhammer, M. L. Bender, N. A. Luedtke, G. R. Heath, D. Cullen, P. Dauphin, D. Hammond, B. Hartman, V. Maynard, Early oxidation of organic-matter in pelagic sediments of the eastern equatorial Atlantic - suboxic diagenesis. *Geochim. Cosmochim. Acta* **43**, 1075–1090 (1979).
- S. L. Jørgensen, B. Hannisdal, A. Lanzén, T. Baumberger, K. Flesland, R. Fonseca, L. Øvreas, I. H. Steen, I. H. Thorseth, R. B. Pedersen, C. Schleper, Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *PNAS* **109**, E2846–E2855 (2012).
- A. Spang, J. H. Saw, S. L. Jørgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A. E. Lind, R. van Eijk, C. Schleper, L. Guy, T. J. G. Ettema, Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
- K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, M. B. Stott, T. Nunoura, J. F. Banfield, A. Schramm, B. J. Baker, A. Spang, T. J. G. Ettema, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
- C. Greening, A. Biswas, C. R. Carere, C. J. Jackson, M. C. Taylor, M. B. Stott, G. M. Cook, S. E. Morales, Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).
- C. W. Stairs, M. M. Leger, A. J. Roger, Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140326 (2015).
- M. Muller, M. Mentel, J. J. van Hellemond, K. Henze, C. Woehle, S. B. Gould, R. Y. Yu, M. van der Giezen, A. G. M. Tielens, W. F. Martin, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).
- G. J. Schut, M. W. Adams, The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: A new perspective on anaerobic hydrogen production. *J. Bacteriol.* **191**, 4451–4457 (2009).
- D. Lloyd, J. R. Ralphs, J. C. Harris, *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen. *Microbiology* **148**, 727–733 (2002).
- T. Happe, J. D. Naber, Isolation, characterization and N-terminal amino acid sequence of hydrogenase from the green alga *Chlamydomonas reinhardtii*. *Eur. J. Biochem.* **214**, 475–481 (1993).
- D. G. Lindmark, B. L. Eckenrode, L. A. Halberg, I. D. Dinbergs, Carbohydrate, energy and hydrogenosomal metabolism of *Trichomonas foetus* and *Trichomonas vaginalis*. *J. Protozool.* **36**, 214–216 (1989).
- E. Nyvltova, R. Sutak, K. Harant, M. Sedivova, I. Hrdy, J. Paces, C. Vlcek, J. Tachezy, Iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7371–7376 (2013).
- I. Hrdy, R. P. Hirt, P. Dolezal, L. Bardonová, P. G. Foster, J. Tachezy, T. Martin Embley, *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432**, 618–622 (2004).
- W. F. Martin, Too much eukaryote LGT. *Bioessays* **39**, (2017).
- M. M. Leger, L. Eme, C. W. Stairs, A. J. Roger, Demystifying eukaryote lateral gene transfer. *Bioessays* **40**, e1700242 (2018).
- L. A. Hug, A. Stechmann, A. J. Roger, Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. *Mol. Biol. Evol.* **27**, 311–324 (2010).
- C. W. Stairs, L. Eme, M. W. Brown, C. Mutsaers, E. Susko, G. Dellaire, D. M. Soanes, M. van der Giezen, A. J. Roger, A Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsoia*. *Curr. Biol.* **24**, 1176–1186 (2014).
- M. M. Leger, L. Eme, L. A. Hug, A. J. Roger, Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. *Mol. Biol. Evol.* **33**, 2318–2336 (2016).

24. C. Schnarrenberger, W. Martin, Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur. J. Biochem.* **269**, 868–883 (2002).
25. V. V. Emelyanov, Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch. Biochem. Biophys.* **420**, 130–141 (2003).
26. A. Omsland, B. S. Sixt, M. Horn, T. Hackstadt, Chlamydial metabolism revisited: Interspecies metabolic variability and developmental stage-specific physiologic activities. *FEMS Microbiol. Rev.* **38**, 779–801 (2014).
27. S. Schmitz-Esser, N. Linka, A. Collingro, C. L. Beier, H. E. Neuhaus, M. Wagner, M. Horn, ATP/ADP translocases: A common feature of obligate intracellular amoeba related to Chlamydiae and Rickettsiae. *J. Bacteriol.* **186**, 683–691 (2004).
28. J. E. Dharamshi, D. Tamarit, L. Eme, C. W. Stairs, J. Martijn, F. Homa, S. L. Jørgensen, A. Spang, T. J. G. Ettema, Marine sediments illuminate Chlamydiae diversity and evolution. *Curr. Biol.* **30**, 1032–1048.e7 (2020).
29. R. B. Pedersen, H. T. Rapp, I. H. Thorseth, M. D. Haft, F. J. A. S. Barriga, T. Baumberger, K. Flesland, R. Fonseca, G. L. Früh-Green, S. L. Jørgensen, Discovery of a black smoker vent field and vent fauna at the Arctic Mid-Ocean Ridge. *Nat. Commun.* **1**, 126 (2010).
30. B. J. Baker, C. S. Lazar, A. P. Teske, G. J. Dick, Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**, 14 (2015).
31. K. Anantharaman, C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas, A. Singh, M. J. Wilkins, U. Karaoz, E. L. Brodie, K. H. Williams, S. S. Hubbard, J. F. Banfield, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
32. P. Liang, M. Rosas-Lemus, D. Patel, X. Fang, K. Tuz, O. Juárez, Dynamic energy dependency of *Chlamydia trachomatis* on host cell metabolism during intracellular growth: Role of sodium-based energetics in chlamydial ATP generation. *J. Biol. Chem.* **293**, 510–522 (2018).
33. W. D. Orsi, Ecology and evolution of seafloor and seafloor microbial communities. *Nat. Rev. Microbiol.* **16**, 671–683 (2018).
34. W. Liesack, F. Bak, J. U. Kreft, E. Stackebrandt, *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch. Microbiol.* **162**, 85–90 (1994).
35. G. B. Slobodkina, O. L. Kovaleva, M. L. Miroshnichenko, A. I. Slobodkin, T. V. Kolganova, A. A. Novikov, E. van Heerden, E. A. Bonch-Osmolovskaya, *Thermogutta terrifontis* gen. nov., sp. nov. and *Thermogutta hypogea* sp. nov., thermophilic anaerobic representatives of the phylum Planctomycetes. *Int. J. Syst. Evol. Microbiol.* **65**, 760–765 (2015).
36. H. Zheng, A. Brune, Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (Phylum Elusimicrobia). *Genome Announc.* **3**, e00679-15 (2015).
37. M. Degli Esposti, D. Cortez, L. Lozano, S. Rasmussen, H. B. Nielsen, E. Martinez Romero, Alpha proteobacterial ancestry of the [Fe-Fe]-hydrogenases in anaerobic eukaryotes. *Biol. Direct* **11**, 34 (2016).
38. A. A. Pittis, T. Gabaldon, Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
39. A. Moustafa, A. Reyes-Prieto, D. Bhattacharya, Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* **3**, e2205 (2008).
40. S. Ball, C. Colleoni, U. Cenci, J. N. Raj, C. Tirtiaux, The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J. Exp. Bot.* **62**, 1775–1801 (2011).
41. J. Martijn, J. Vosseberg, L. Guy, P. Offre, T. J. G. Ettema, Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
42. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
43. K. Raymann, C. Brochier-Armanet, S. Gribaldo, The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6670–6675 (2015).
44. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
45. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
46. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
47. R. D. Finn, A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
48. R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H. Y. Chang, Z. Dosztányi, S. el-Gebali, M. Fraser, J. Gough, D. Haft, G. L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka, C. A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N. D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, I. Xenarios, L. S. Yeh, S. Y. Young, A. L. Mitchell, InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
49. J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
50. J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M. C. Walter, T. Rattei, D. R. Mende, S. Sunagawa, M. Kuhn, L. J. Jensen, C. von Mering, P. Bork, eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
51. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
52. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
53. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
54. R. L. Tatusov, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. Sridhar Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, D. A. Natale, The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
55. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, M. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
56. M. M. Leger, M. Kolisko, R. Kamikawa, C. W. Stairs, K. Kume, I. Čepička, J. D. Silberman, J. O. Andersson, F. Xu, A. Yabuki, L. Eme, Q. Zhang, K. Takishita, Y. Inagaki, A. G. B. Simpson, T. Hashimoto, A. J. Roger, Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitochondria. *Nat. Ecol. Evol.* **1**, 0092 (2017).
57. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
58. D. Sondergaard, C. N. S. Pedersen, C. Greening, HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
59. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
60. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
61. Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
62. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
63. A. Crisuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
64. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
65. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
66. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: Combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
67. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
68. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
69. S. Quang, Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323 (2008).
70. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

71. S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
72. H. C. Wang, B. Q. Minh, E. Susko, A. J. Roger, Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235 (2018).
73. F. Lemoine, J. B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. de Oliveira, O. Gascuel, Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
74. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
75. A. J. Aberer, D. Krompass, A. Stamatakis, Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).
76. H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
77. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009).
78. L. Guy, J. Roat Kultima, S. G. E. Andersson, genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).
79. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
80. A. Rambaut, *FigTree v1.3.1* (Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, 2010).

Acknowledgments

Funding: This work was supported by grants from the Swedish Research Council (VR grant 2015-04959), the European Research Council (ERC starting and consolidator grants 310039 and 817834, respectively), and the Swedish Foundation for Strategic Research (SSF-FFL5) to T.J.G.E. C.W.S. was supported by the European Molecular Biology Organization long-term fellowship (ALTF-997-2015) and the Natural Sciences and Engineering Research Council of

Canada (PDF 487174-2016). L.E. was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 704263 and by funding from the European Research Council (ERC starting grant no. 803151). D.T. was supported by the Swedish Research Council International Postdoc grant (2018-06609). A.S. was supported by the Swedish Research Council (VR starting grant 2016-03559) and the NWO-I Foundation of the Netherlands Organization for Scientific Research (WISE fellowship). S.L.J. was supported by the Trond Mohn starting grant BFS2017REK03. **Author contributions:** Conceptualization: T.J.G.E., J.E.D., C.W.S., and A.S. Data curation: J.E.D. and C.W.S. Formal analysis: J.E.D., C.W.S., and T.J.G.E. Investigation: J.E.D., C.W.S., L.E., D.T., and T.J.G.E. Validation: all authors. Resources: S.L.J. and T.J.G.E. Supervision: T.J.G.E. Visualization: J.E.D. and C.W.S. Writing—original draft: J.E.D., C.W.S., and T.J.G.E. Writing—reviewing and editing: all authors. The contributions of D.T. and L.E. should be regarded as equal. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** In addition to data available in the Supplementary Materials, files containing sequence datasets, alignments, and phylogenetic trees in Newick format are archived at the digital repository Figshare DOI: 10.6084/m9.figshare.12387980. Whole Genome Shotgun projects for metagenome assemblies GS08_GC12_126, GS10_PC15_940, GS10_PC15_1000, and GS10_PC15_1060 have been deposited at DDBJ/ENA/GenBank under the accessions SDBU000000000, SDBV000000000, SDBS000000000, and SDBT000000000, respectively. The versions described in this paper are versions SDBU010000000, SDBV010000000, SDBS010000000, and SDBT010000000, with MAGs generated from each linked to BioProject PRJNA504765. Accessions for genomes analyzed in this study can be found in data S1.

Submitted 13 March 2020

Accepted 13 July 2020

Published 26 August 2020

10.1126/sciadv.abb7258

Citation: C. W. Stairs, J. E. Dharamshi, L. Eme, D. Tamarit, S. L. Jørgensen, A. Spang, T. J. G. Ettema, Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. *Sci. Adv.* **6**, eabb7258 (2020).