bioRχiv

An Asgard archaeon with internal membrane compartments

Fraser I. MacLeod[a,*], Andriko von Kügelgen[b,*], Magdalena K. Lechowska[a], Joe Parham[a,†], Iain A. Richard[a,†], Emily J. Aguilar-Pine[c], Brett J. Baker[c,d], Tanmay A.M. Bharat[b,], and Buzz Baum[a,**]**

[a] Cell Biology Division, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom.
[b] Structural Studies Division, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom.
[c] Department of Integrative Biology, University of Texas Austin, Texas, USA.
[d] Department of Marine Science, University of Texas Austin, Marine Science Institute, Port Aransas, Texas, USA
[*] These authors contributed equally to this work.
[†] These authors contributed equally to this work.
[**]Correspondence should be addressed to B.B. (bbaum@mrc-lmb.cam.ac.uk) and T.A.M.B. (tbharat@mrc-lmb.cam.ac.uk)

The emergence of eukaryotes from a merger between an archaeon and a bacterial cell ~two billion years ago involved a profound change in cellular organisation. While the order in which different features of the eukaryotic cell arose remains a matter of controversy[1-3], close archaeal relatives of eukaryotes have recently been identified that possess homologues of eukaryotic trafficking machinery[4,5] and a complex cell architecture[6,7]. The members of this phylum, the Asgard archaea (*syn.* Prometheoarchaeota) described so far, however, lack internal membrane-bound compartments, and therefore have shed little light on origins of the hallmark eukaryotic endomembrane system. Here we report the cell biological analysis of a member of the Heimdallarchaeia, *Candidatus* 'Yibarchaeum umbracryptum' in enriched mixed microbial communities. Possessing a small genome encoding few homologues of eukaryotic membrane remodelling machinery[8], *Ca.* Y. umbracryptum cells in late-stage cultures resemble previously described Asgard archaea with extensive cellular protrusions. Surprisingly, during early stages of culture growth *Ca.* Y. umbracryptum cells have fewer protrusions but possess numerous intracellular vesicles, most of which have a luminal surface that morphologically resembles the outer coat of the plasma membrane. These data alter our view of eukaryogenesis by identifying a close archaeal relative of eukaryotes with a regulated endomembrane system.

Introduction

The emergence of eukaryotes about two billion years ago was associated with a profound change in cellular organisation as comparatively, ultrastructurally simple archaea and bacteria gave rise to complex eukaryotic cells that possess numerous different internal membrane-bound compartments[2]. These include the nucleus, mitochondria, the endoplasmic reticulum, the Golgi apparatus, lysosomes, along with intracellular vesicles that mediate the transport of lipids and protein cargo between these different organelles. In addition, subsets of these internal vesicles traffic to and from the plasma membrane[9], mediating communication between the cell interior and the external world.

The origins of this complex architecture of the eukaryotic cell and the order in which the different compartments arose remain unknown[2]. Nevertheless, as the result of phylogenetic analyses[10,11] and structural data[12], it is now widely accepted that eukaryotes emerged from a process of endosymbiosis when a cell related to modern Alphaproteobacteria took up residence inside a host cell to give rise to mitochondria. While the identity of the host cell long remained a point of contention, recent phylogenetic studies using a range of marker genes of archaeal origin firmly place eukaryotes within the Asgardarchaeota (*syn.* Prometheoarchaeota). More specifically, most studies place eukaryotes within[4,8,13] or as sisters to[14] the Heimdallarchaeia. In support of this conclusion, the genomes of many Asgard archaea also contain genes with homologues in eukaryotes that function in membrane remodelling[8,15]. However, with the possible exception of ESCRT-III polymers[4,16-19] and the actin machinery from Asgard archaea[20,21], which exhibit structurally and functionally similar properties to their eukaryotic counterparts when studied *in vitro*, it is unclear whether these perform analogous roles in archaea. Similarly, while the first images of Asgard lineage cells revealed a number of cellular hallmarks of eukaryotes, including protrusions rich in an actin cytoskeleton[7], the Asgard archaea imaged thus far lack clear evidence of internal membrane-bound compartments. As a result, it has not yet been possible to distinguish between models of the evolution of eukaryotes that differ in their predictions about the relative timing of the emergence of different parts of the endomembrane system[1,3,22,23].

In this article, we report the cell biological analysis of a member of the Heimdallarchaeia, *Candidatus* 'Yibarchaeum umbracryptum', which was cultivated as part of a mixed microbial community from a sample taken from microbial mats in Shark Bay, Western Australia. Strikingly, *Ca.* Y. umbracryptum cells possess a diverse set of intracellular vesicles, most of which appear empty at the resolution of cryogenic electron tomography (cryo-ET), which are lined with a surface coat that is morphologically and ultrastructurally indistinguishable from the coat on the external face of the plasma membrane. Furthermore, *Ca.* Y. umbracryptum cells

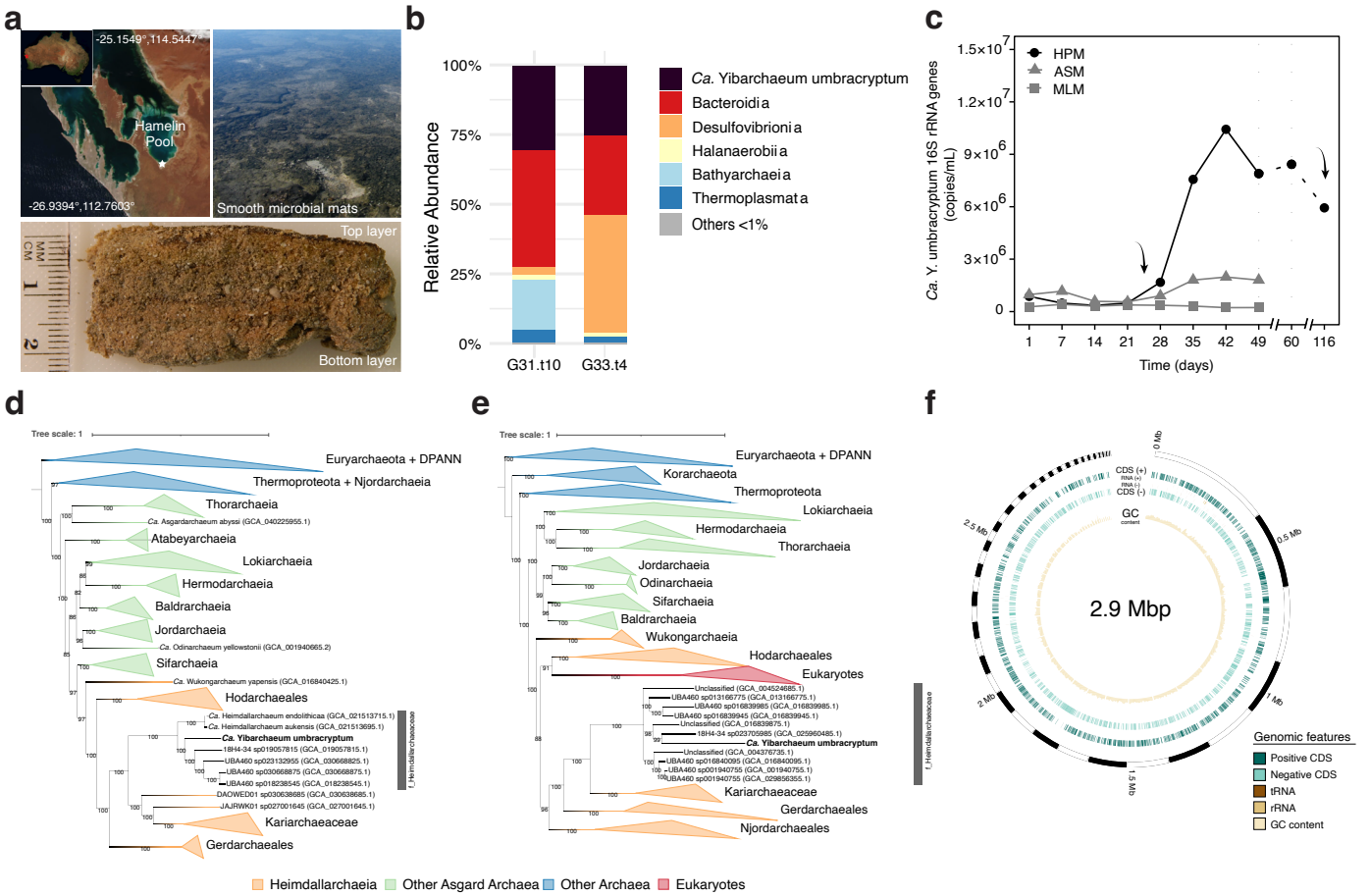

Figure 1: Isolation and cultivation of Heimdallarchaeaceae *Candidatus* Y. umbracryptum.
(a) Satellite image of Australia showing the location of Shark Bay and satellite image of Shark Bay showing the location of the sample site, taken on day of sampling (top left). Satellite images were retrieved from NASA's Earth Science Data and Information System (ESDIS). Image of the smooth microbial mats in the Hamelin Pool inter-tidal zone (top left). Image of the smooth microbial mat sample used as the initial inoculum (bottom). **(b)** Relative abundance of *Ca.* Y. umbracryptum in late stage (G31.t10) and early stage (G33.t4) enrichment cultures, as determined by amplicon sequencing using the universal primer set 926F-1392wR. **(c)** Growth of *Ca.* Y. umbracryptum as determined by qPCR under different media conditions. Hamelin Pool Medium (HPM), Archaeal Salt Medium (ASM), Minimal Loki Medium (MLM)[7]. Early and late sampling timepoints for light-microscopy and cryo-ET experiments are indicated by curved black arrows. **(d)** Maximum-likelihood phylogenetic placement of *Ca.* Y. umbracryptum within the archaea using ar53 marker genes from GTDB-tk[26]. **(e)** Maximum-likelihood phylogenetic placement of *Ca.* Y. umbracryptum including the placement of eukaryotes using NM57 marker genes[8]. **(f)** Schematic of the *Ca.* Y. umbracryptum draft genome, illustrating genome size, GC content, the set of tRNA and rRNA included in the genome, and the binning of the genome into 55 contigs.

were observed wrapping extracellular material, and in one case another organism, in the mixed culture. Since vesicles were only present in high numbers in cells taken from early-stage cultures that possess markedly fewer protrusions, their formation and loss appears to be something that cells regulate depending on environmental circumstances. Taken together, these data suggest the possibility that *Ca.* Y. umbracryptum has a dynamic endomembrane system and can likely both internalise material from the environment, a prerequisite for the final steps in the process leading to the acquisition of mitochondria, and traffic intracellular material to the cell surface. Thus, by reporting the cell biology of a member of the Heimdallarchaeia, an archaeon with an abundance of internal vesicles, this study further narrows the divide that separates archaea and eukaryotes[4] and sheds new light on the emergence of the endomembrane system, which a hallmark of eukaryotes.

Results

Heimdallarchaeia enrichment cultures

Microbial mat samples **(Fig. 1a)** rich in Asgard archaea (between approximately 2.5% and 5% of DNA sequencing reads) were collected from the Nilemah tidal flats of Hamelin Pool in Shark Bay, Western Australia as described previously[24,25], and used to inoculate anaerobic cultures (Materials and Methods). By passaging these cultures over six years in different growth media, mixed cultures were identified that, based on 16S rRNA gene sequences, contained a single Heimdallarchaeia lineage at a stable level of 25-30% of the population, along with a set of eight other archaea and bacteria, including uncharacterised species of Bacteroidia, Desulfovibrionia, Halanaerobiia, Bathyarchaeia, and Thermoplasmata **(Fig. 1b)**. We named the Heimdallarchaeia lineage Asgard archaeon, *Candidatus* 'Yibarchaeum umbracryptum'. Based upon 16S rRNA qPCR, we estimate that *Ca.* Y. umbracryptum has a minimum doubling time of about three days and an average doubling time of seven and a half days during log phase growth **(Fig. 1c).** Short-read DNA sequencing of these cultures enabled the assembly of a near-complete Heimdallarchaeia genome in 55 contigs **(Supplementary Table 1)**. A maximum-likelihood phylogenetic analysis using recently published marker sets[8,26] identified this organism as a member of the class

Heimdallarchaeia and family Heimdallarchaeaceae **(Fig. 1d, Supplementary Fig. 1)** and a close relative of eukaryotes **(Fig. 1e and Supplementary Fig. 1)**.

The genome of *Ca.* Y. umbracryptum, while typical for a member of the family Heimdallarchaeaceae, was markedly smaller than those reported for the recently cultivated Lokiarchaeia species *Prometheoarchaeum syntrophicum*[6], *Candidatus* 'Lokiarchaeum ossiferum'[7] and Hodarchaeales[27] **(Fig. 1f and Supplementary Fig. 2a)**. In line with this, its genome encodes fewer proteins associated with a host of functions from information processing to metabolism, cellular ultrastructure and division than the other cultivated Asgard archaea **(Supplementary Fig. 2b)**. Most strikingly, the reconstruction of its metabolic pathways **(Supplementary Fig. 3a-b)** revealed that *Y. umbracryptum* is only capable of synthesizing six amino acids (alanine, methionine, glutamic acid, serine, glutamine, glycine). Since it also lacks degradation pathways for most amino acids, these data suggest the possibility that raw material for protein synthesis in *Ca.* Y. umbracryptum is supplied from external sources of amino acids or peptides. The *Ca.* Y. umbracryptum genome also encodes multiple ATP-binding cassette transporters that includes a system for small peptide uptake. In other respects, however, its complement of genes seemed largely complete. The genome assembly encodes 45 tRNAs **(Fig. 1f and Supplementary Table 1)**, along with the genes required to perform tRNA aminoacylation for most amino acids **(Supplementary Fig. 3a)**, as well as the genes required to perform basic nucleic acid interconversions, and key steps of nucleic acid biosynthesis/degradation. A nearly complete archaeal-type mevalonate (MVA) pathway was also present **(Supplementary Fig. 3a)**, suggesting that *Ca.* Y. umbracryptum can synthesize necessary precursors for any isoprenoid-based compounds, including archaeal membrane lipids[28]. These data support the idea that *Ca.* Y. umbracryptum depends on partners or extracellular material for growth as was suggested to be the case for other Asgard archaea in mixed cultures[6,7]. We also noted that while the *Ca.* Y. umbracryptum genome encodes homologues of eukaryotic actin and ESCRT-I, -II, and -III proteins **(Supplementary Fig. 3b and 3c)** and 14 Arf/SAR type GTPase domain proteins[4], it lacks clear homologues of proposed membrane remodelling systems reported in other Asgard genomes[8,13], such as Sec23/24, TRAPP, SNAREs and arrestins **(Supplementary Fig. 3c)**.

The paucity of "eukaryotic signature proteins" (ESPs) present in its genome relative to the genomes of Lokiarchaeia species *P. syntrophicum*[6], *Ca.* 'Lokiarchaeum ossiferum'[7] and Hodarchaeales[27] led us to examine the cell biology of *Ca.* Y. umbracryptum. To study its morphology using light microscopy, cells from late-stage cultures were visualised under anaerobic conditions on a confocal microscope. Under these conditions, alongside filamentous bacteria and rare cells with distinct morphologies **(Supplementary Fig. 4)**, we observed cells similar in appearance to the Asgard archaea described in recent reports[6,7,25,27,29,30], with a central cell body containing the DNA and long branched protrusions **(Fig. 2a)**.

To confirm the identity of *Ca.* Y. umbracryptum in these mixed cultures, we modified the imaging protocol for hybridised chain reaction fluorescent *in situ* hybridisation (HCR-FISH). Fixed and permeabilised cells were labelled with HCR-FISH probes specific to the *Ca.* Y. umbracryptum 16S rRNA gene and a fluorescent NHS-ester to mark the contour of all the cells in the community **(Figs. 2b-c)**. As expected, the *Ca.* Y. umbracryptum-specific 16S rRNA probe specifically labelled cells with long protrusions, many of which appeared to be part of extended networks, but did not label cells in Heimdallarchaeia-free cultures **(Supplementary Fig. 4b)**. Conversely, the Heimdall cultures were not labelled by an HCR-FISH probe designed to complement Lokiarchaeia rRNA. Having confirmed the identity of these cells as *Ca.* Y. umbracryptum **(Fig. 2b and Supplementary Fig. 4)**, to image the localisation pattern of rRNA at higher resolution, we processed the *Ca.* Y. umbracryptum-containing multicellular cultures for expansion microscopy by adapting a previously described protocol[31] that uses a crosslinking agent to enhance the RNA retention, and observed cells after a 4X expansion using the same Heimdall-specific HCR-FISH probes. Again, this revealed networks of cells with elaborate protrusions rich in rRNA **(Fig. 2d)**, similar those seen in other Asgard archaeal cultures[7,29].

Ultrastructural analysis of *Ca.* Y. umbracryptum

Having used HCR-FISH to confirm the identity of *Ca.* Y. umbracryptum cells in mixed communities, cryo-ET was employed to examine their ultrastructure. For this analysis, cells from late-stage cultures were collected from anaerobic serum bottles, applied directly onto electron microscopy grids and vitrified. In line with data from the light microscopy experiments, cells on the grids tended to have a small cell body with long undulating cell-surface protrusions **(Fig. 2e and Supplementary Fig. 5).** In tomograms, ribosomes could be seen that were abundant in protrusions, as was previously observed in *Ca.* L. ossiferum cultures[7] but tended to be excluded from cell bodies that contained a dense nucleoid **(Fig. 2e)**, in agreement with our HCR-FISH data (**Fig. 2b-d**). Using subtomogram averaging[32,33] we generated a 10 Å-resolution structure of the ribosome in these cells **(Fig. 3f, Supplementary Fig. 6 and Supplementary Table 2)** that could be compared with previously reported atomic structures of an archaeal ribosome from *Thermococcus kodakerensis*[34] and a bacterial ribosome[35]. The comparison confirmed the identity of these cells as archaea **(Supplementary Fig. 6)**, in agreement with the HCR-FISH data. These cells also contained numerous linear cytoskeletal elements, many of which were bundled **(Fig. 2g)**, especially within the thin necks connecting protrusions **(Fig. 2g and Supplementary Fig. 5).** These resembled the bacterial and archaeal actin filaments that have been observed by cryo-ET previously[36,37]. and subtomogram averaging revealed a double-protofilament architecture with a 5.9 nm repeat close to that of prokaryotic

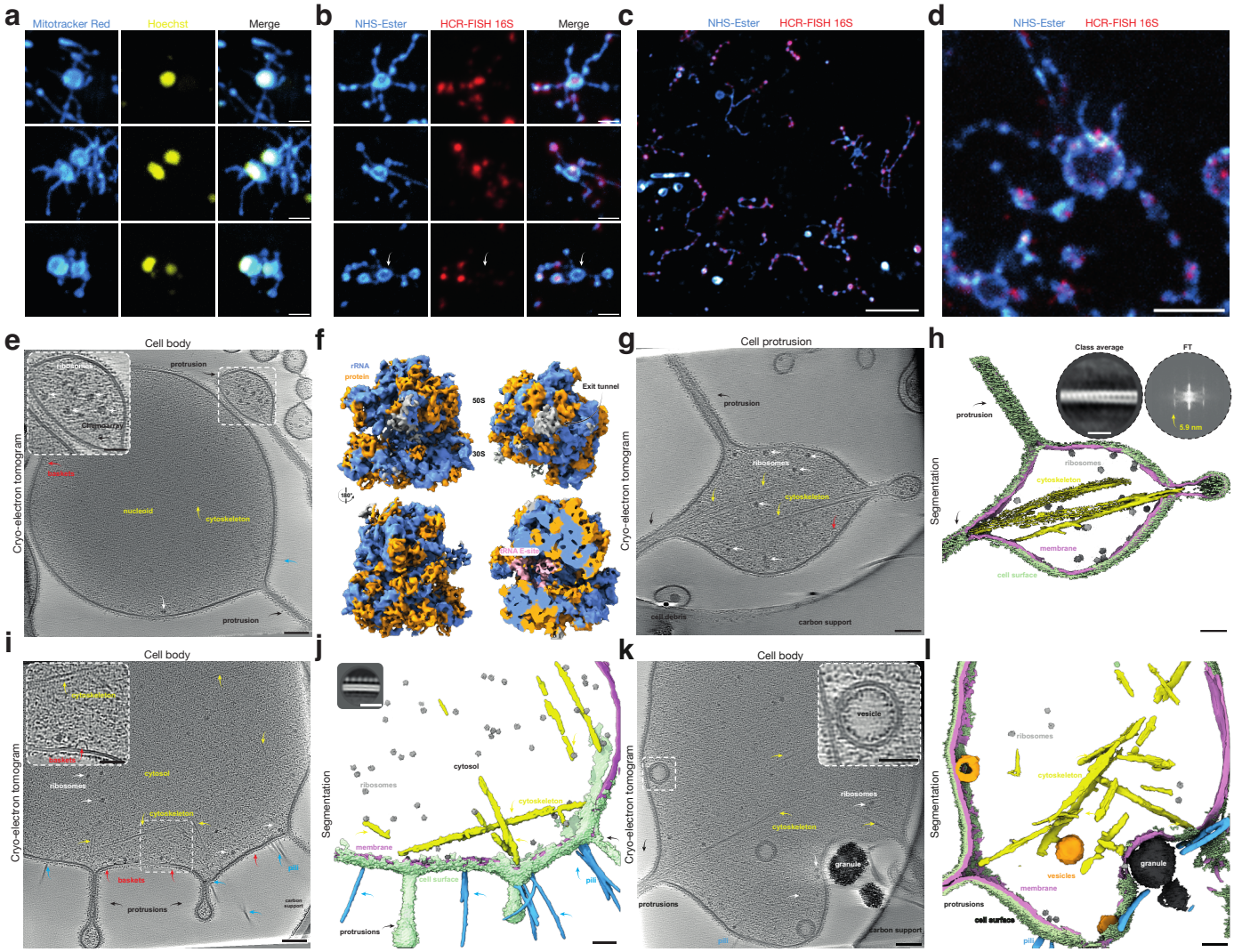

Figure 2: Light microscopy and cryo-ET of stationary phase *Ca.* Y. umbracryptum cultures.
(a) Examples of cells from live-cell imaging of the stationary culture. MitoTracker Red CMXRos (depicted in cyan hot) was used as a cell contour stain and Hoechst 33342 (yellow) as a DNA stain in two biological replicates. Scale bars: 1 μm. **(b)** Example images of cells from cultures labelled with *Ca.* Y. umbracryptum-specific HCR-FISH 16S rRNA probes (red) and with NHS-ester (cyan hot) as a generic cell contour stain from two biological replicates. White arrow depicts cell with HCR-FISH signal in protrusions but not in the cell body. Scale bars: 1 μm. **(c)** Overview image of HCR-FISH data shown in **b**, Scale bar: 5 μm. **(d)** Expansion HCR-FISH image of *Ca.* Y. umbracryptum with HCR-FISH 16S probes (red) with NHS-ester (cyan hot) from two biological replicates. Scale bar: 5 μm (see also Supplementary Fig. 4). **(e)** Cryo-ET slice through the cell body of (putative) *Ca.* Y. umbracryptum with few ribosomes (white arrows) and cytoskeletal filaments (yellow arrows). Inset shows close up of cellular protrusion filled with multiple ribosomes (white). Scale bar: 100 nm; Inset scale bar: 50 nm (see also Supplementary Fig. 5). **(f)** Subtomogram averaging structure of the ribosome from these cells at ~10 Å (see also Supplementary Fig. 6a-c). **(g)** Cryo-ET slice through a cell protrusion showing multiple cytoskeletal filaments (yellow arrows). Scale bar: 100 nm. **(h)** Semantic segmentation of the tomogram shown in panel g) depicting the cell membrane (pink), cell-surface (pale green), ribosomes (grey) and cytoskeletal filaments (yellow). Inset shows two-dimensional class average of cytoskeletal filaments and the corresponding Fourier Transform (FT, power spectrum shown) with a repeat at ~5.9 nm. Scale bar: 100 nm; Inset scale bar: 25 nm (see also Supplementary Fig. 6d). **(i)** Cryo-ET slice through the cell body of *Ca.* Y. umbracryptum with multiple cytoskeletal filaments (yellow). Inset shows close up of cytoskeletal filaments and a membrane attached basket. Scale bar: 100 nm; Inset scale bar: 50 nm. **j**, Semantic segmentation of the tomogram shown in panel i) with similar color schemes as panel h), but including pili (blue). Inset shows a two-dimensional class average of the cell membrane. Scale bar: 100 nm. Inset scale bar: 25 nm (see also Supplementary Video 1). **(k)** Cryo-ET slice through a cell body of a *Ca.* Y. umbracryptum cell with multiple intracellular vesicles. Inset shows a close-up of one intracellular vesicle with clear separation of the two leaflets of the lipid bilayer. Scale bar: 100 nm; Inset scale bar: 50 nm. (see also Supplementary Fig. 7 and Supplementary Video 2). **(l)** Semantic segmentation of the tomogram shown in panel k) with similar colour schemes as panel h) but including versicles (orange) and extracellular granules (black). Tomograms: n=135, from two biological replicates.

actin[37] **(Figs. 2g-h).** At the same time, these filaments had only a weakly discernible long-range protofilament twist that is characteristic eukaryotic actin[38]. Since *Ca.* Y. umbracryptum is the only species in the mixed community containing a close actin homologue in its genome, these data, when taken together with the ribosome subtomogram averaging, and with their characteristic morphology observed in HCR FISH, confirmed the identity of these cells.

By cryo-ET we were also able to observe the bounding membrane of cells and membrane associated structures in detail. Like the surface of *P. syntrophicum* and *Ca.* L. ossiferum[6,7] cells, the surface of *Ca.* Y. umbracryptum cells appeared coated with an amorphous proteinaceous layer rather than a continuous, paracrystalline S-layer lattice that is characteristic of many archaea[39-41]. The plasma membrane was also decorated with occasional pili of one type **(Fig. 2i-j).** Additionally, in many tomograms we observed basket-

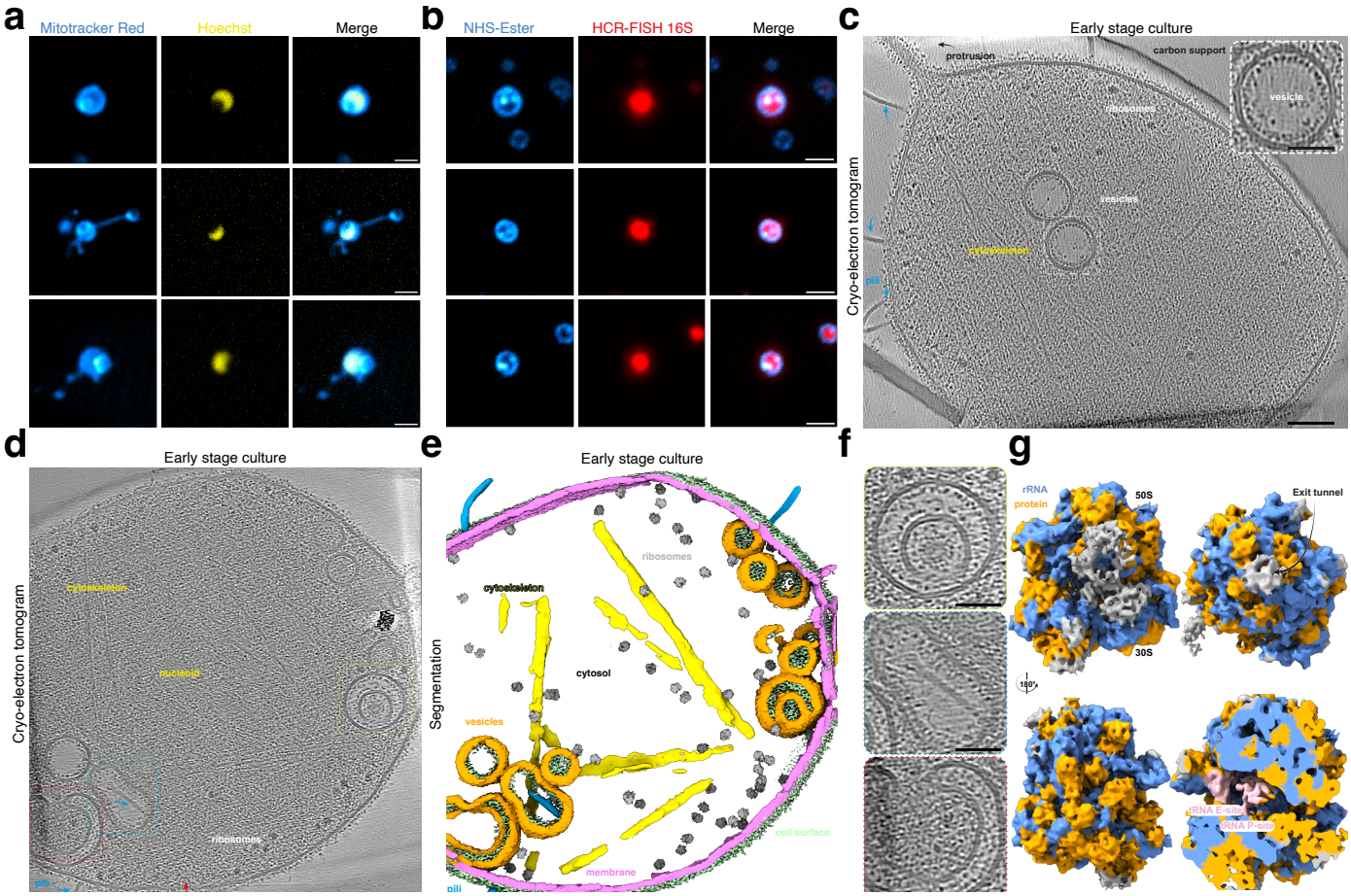

Figure 3: Light and electron microscopy of early phase *Ca.* Y. umbracryptum cultures.
(a) Examples from live-cell imaging of the early log phase culture. MitoTracker Red CMXRos (depicted in cyan hot) was used to mark the *Ca.* Y. umbracryptum cell contour and Hoechst (yellow) the DNA in two biological replicates. MitoTracker Red CMXRos staining revealed large intracellular spheres of staining, which excluded Hoechst signal. Scale bars: 1 μm. **(b)** Example images of *Ca.* Y. umbracryptum cells labelled with specific HCR-FISH probes targeting the 16S rRNA (red) and NHS-ester (cyan hot) as a generic stain to mark the cell contour from two biological replicates (see also Supplementary Fig. 4). Cells identified as *Ca.* Y. umbracryptum by HCR-FISH labelling appeared to contain intracellular NHS-ester foci. **(c)** Cryo-ET slice through a *Ca.* Y. umbracryptum cell with few ribosomes, cytoskeletal filaments and intracellular vesicles marked. Inset shows close-up of one intracellular vesicle with clear separation of the two leaflets of the lipid bilayer. Scale bar: 100 nm; Inset scale bar: 50 nm. (see also Supplementary Figs. 9-10) **(d)** Cryo-ET slice through a cell body of early-log phase *Ca.* Y. umbracryptum with multiple intracellular vesicles. Scale bar: 100 nm. **(e)** Semantic segmentation of the tomogram shown in panel d) depicting the cell membrane (pink), cell-surface proteins (pale green), pili (blue), ribosomes (grey), cytoskeletal filaments (yellow) and intracellular vesicles (orange). Scale bar: 100 nm. (see also Supplementary Fig. 10 and Supplementary Video 3). **(f)** Close-up views of cryo-ET slices shown in panel d) depicting 1) closed multi-lamellar vesicle; 2) internalized pilus; and 3) open multi-lamellar vesicle. Scale bars: 50 nm. **(g)** Subtomogram averaging structure of the ribosome at ~10.7 Å resolution. Tomograms: *n*=42, from two biological replicates.

like structures attached to the internal face of the *Ca.* Y. umbracryptum plasma membrane **(Figs. 2e,g,i and Supplementary Fig. 7)**, often in clusters (**Supplementary Fig. 7b).** Since similar structures are formed by SPFH family of proteins in bacteria and eukaryotes[42,43], we searched the genome of *Ca.* Y. umbracryptum for SPFH homologues and identified two SPFH-family proteins that contain the extended alpha-helical domain required to form a basket **(Supplementary Fig. 7)**.

Intracellular vesicles in the cytosol are enriched in early-stage cultures

By cryo-ET we also identified several cells containing a small number of internal vesicles **(Figs. 2k-l, Supplementary Figs. 8a-h).** To ensure that this was not an experimental artefact of our cryo-ET vitrification in ambient conditions, we performed vitrification under anaerobic conditions with identical results **(Supplementary Fig. 8i-l).** Since these vesicles were rare in high-density, stationary, late-stage cultures, we then repeated the analysis, looking at cells taken from cultures at earlier stages in the microbial community growth cycle. For this analysis we examined cultures that were just entering log phase growth (3-4 weeks after inoculation) **(Fig 1c).** While these contained ~30% Heimdallarchaeia as measured by 16S amplicon sequencing, when imaged live or following fixation we observed few cells with protrusions **(Fig. 3a and Supplementary Fig. 4a).** Instead, these early-stage cultures contained a high proportion of cells that were spherical, slightly larger than the *Ca.* Y. umbracryptum cell bodies seen in late-stage cultures; a proportion of which had poorly resolved spot-like structures that labelled with the membrane probe **(Figs. 3a-b).** Using cryo-ET to examine the ultrastructure of these cells, we were able to confirm their identity as *Ca.* Y. umbracryptum based on the presence of actin-like filaments, their characteristic plasma membrane surface coat, and the ribosomal structure generated by subtomogram averaging that matched the structure from the late stage cells (**Fig. 3 and**

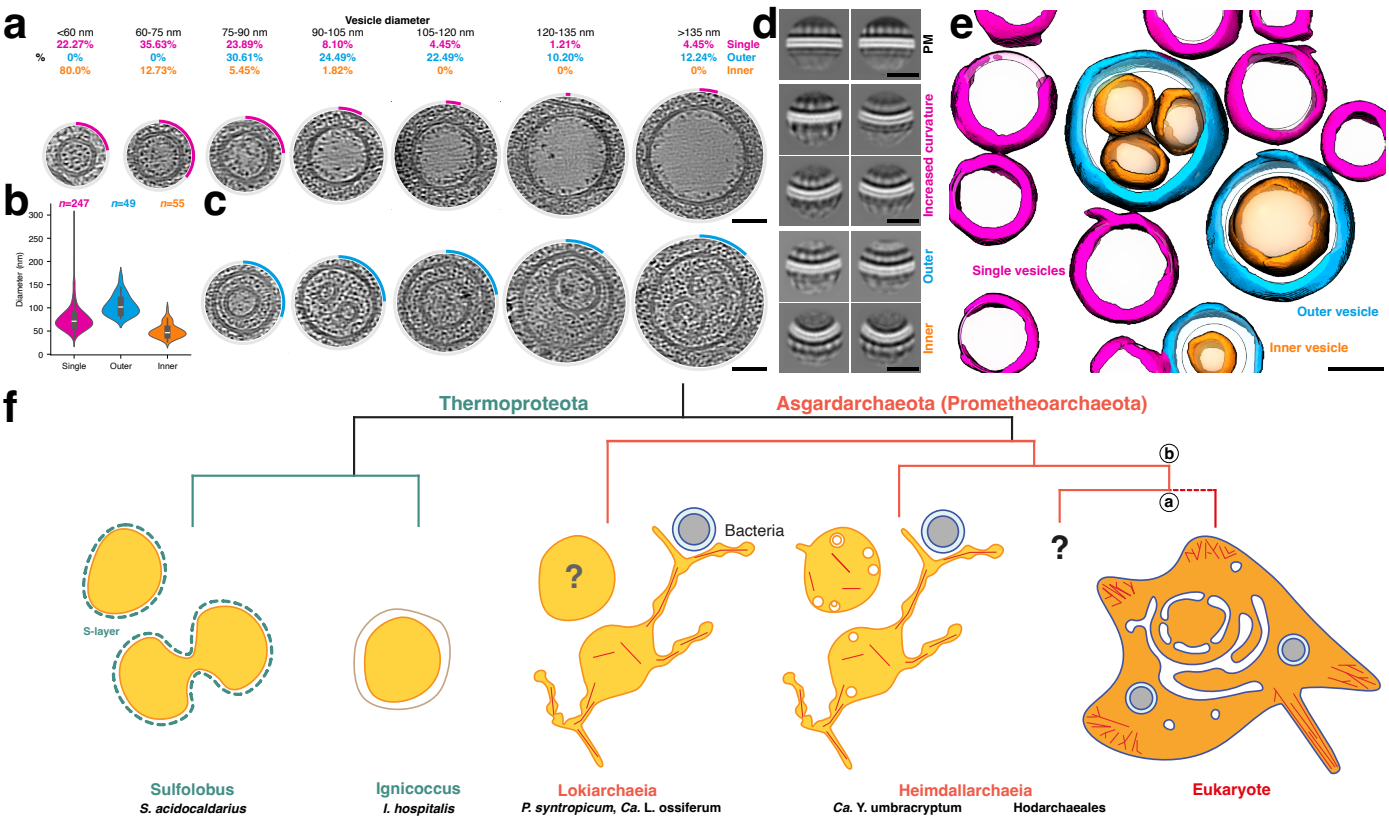

Figure 4: Intracellular vesicles of *Ca.* Y. umbracryptum.
(a) Gallery of cryo-ET slices showing single vesicles arranged by increasing diameter from left to right. The percentages of intracellular vesicles are measured in groups from <60 nm to >135nm in bin-sizes of 15 nm. The percentage of each bin-size is additionally visualized as an arc (magenta) surrounding a representative vesicle of the corresponding diameter. Scale bar: 100 nm. **(b)** Violin plot of measured vesicle diameter (nm) of single (magenta), outer filled vesicles (cyan) and inner vesicles (orange). **(c)** Gallery of multi-lamellar vesicles ordered by diameter with the same group sizes and percentages displayed in panel a). Scale bar: 100 nm. **(d)** Subtomogram analysis of the plasma membrane (PM) and internalized membranes from vesicles produced two-dimensional averages at different membrane curvatures. The lipid bilayer (white) is clearly resolved in the averages, which is covered by an amorphous, proteinaceous surface coat in all cases. Two-dimensional averages of outer filled vesicles show an additional fuzzy lipid bilayer within its lumen, while averages of inner vesicle show inverted topology of the proteinaceous coat (i.e. coat on the outside, with positive curvature of the underlying membrane). Scale bars: 25 nm. **(e)** Semantic segmentation of a *Ca.* Y. umbracryptum cell with single (magenta), outer filled vesicles (cyan) and inner vesicles (orange), overlaid with a fitted transparent sphere in the same color. Scale bar: 50 nm. Tomograms: *n*=35, from two biological replicates. **(f)** Cartoon representation of cell morphology of Thermoproteota (light green), Asgardarchaeota (*syn.* Prometheoarchaeota - light red), diderm bacteria (pale grey) and Eukaryotes (red). Eukaryotes are depicted to branch within (a)[4,8,13] or sister (b)[14] to Heimdallarchaeia. The architecture of early-stage Lokiarchaeial cells remains unresolved (grey question mark). Cells are not depicted to scale.

Supplementary Fig. 9). Strikingly, most cells contained vesicles, often in very large numbers **(Figs. 3c-f and Supplementary Fig. 10).** While these vesicles had a range of morphologies, the major class seen across multiple biological replicates resembled those observed in late-stage cultures in shape, size and architecture **(Fig. 2k and Supplementary Fig. 8).** The majority were spherical and had diameters of between 60 and 90 nm (**Figs. 4a-b).** Most were positioned close to the cytoplasmic membrane without being topologically connected to it, and were devoid of dense contents. On their luminal side, the vesicular coat appeared morphologically indistinguishable from the cell surface (**Figs. 4a-d and 2j**). This characteristic coat ultrastructure was confirmed when we segmented and extracted all membranes (including the plasma membrane) observed in tomograms, and performed a subtomogram analysis by adapting previously described techniques[44] **(Fig. 4d)**. The presence of this internal coat was consistent with these vesicles being involved in material exchange with the surface membrane. In line with this hypothesis, in one instance a vesicle was identified that contained a pilus that was morphologically indistinguishable from pili observed on the cell surface **(Figs. 3d-f)** - implying that the vesicle was either in the process of delivering material to the surface or contained cell surface material that had been internalised.

Although the lumen of most vesicles was largely empty when viewed at this resolution, we also observed a population of intermediate-sized vesicles (75-150 nm) that contained smaller vesicles inside **(Figs. 4b,c,e).** In some cases, the inner membrane was connected to the outer vesicular membrane by three-way junctions **(Supplementary Fig. 10a,e).** In addition to this, we observed a number of larger vesicles which had a non-spherical shape. These included vesicles with donut-shaped form and, in one case, a double membrane that formed a C-shape **(Fig. 3f and Supplementary Fig. 10c)**, reminiscent of an early stage intermediate of multivesicular body formation or autophagosome assembly in eukaryotes[45]. In this instance, it was clear that the inner compartment contained cytoplasm and ribosomes as would be expected if

this process were catalysed by ESCRT-III dependent membrane remodelling[46]. We also observed several examples of extracellular material, including another archaeal cell, being wrapped within surface invaginations of the *Ca*. Y. umbracryptum cell **(Supplementary Fig. 11)**. Finally, we also observed a cell with large internal compartments containing electron-dense material similar to that seen in the medium, which may have been internalised **(Supplementary Fig. 11f)**.

Discussion

This paper describes the cell biology of *Ca.* Y. umbracryptum, the first member of the Heimdallarchaeia to be cultivated and imaged. Like the other Asgard archaea that have been subjected to a detailed cell biological analysis[6,7,29], when taken from dense, late-stage cultures, these cells were characterised by a central cell body that contains the genome from which long, branched protrusions extend. *Ca.* Y. umbracryptum (like other Asgard archaea) appears to lack many key enzymes in several core metabolic pathways consistent with suggestions[6] that the protrusions facilitate syntropic interactions with partners in the mixed community in which they live, or assist in the absorption of metabolites from the environment. Unlike previously enriched or co-cultured Lokiarchaeia, however, these Asgard archaea were found to possess large numbers of intracellular vesicles. While vesicles were only present in a small number of cells in late-stage cultures, they were common in early-stage cultures in cells that lacked protrusions and which were entering a period of rapid growth. Thus, *Ca.* Y. umbracryptum cells have a large surface to volume ratio at all stages of growth, and appear to be able to use vesicles and protrusions as interchangeable condition-specific membrane reservoirs. While an understanding of the triggers that cause cells taken from cultures at two different stages in their growth cycle to differ in their morphology and internal cell architecture will require a much more detailed future analysis, growth to stationary phase is likely accompanied by a host of changes in the chemical and physical extracellular environment, all of which may contribute to this effect – as has been shown to be the case for changes in the shape of other archaea[47,48].

The majority of the vesicles present in *Ca.* Y. umbracryptum cells were lined on their luminal side by a proteinaceous inner coat that resembled the external surface of the plasma membrane **(Fig. 4)**. Additionally, we observed a single instance of a pilus being present inside a non-spherical vesicle **(Fig. 3d-f)**, implying that its contents were being trafficked either to or from the cell surface. Since vesicles appear to be formed and lost according to environmental conditions, we expect this traffic to flow in both directions. Note that vesicles in eukaryotic cells bud off from both the plasma membrane and the ER and Golgi membranes via similar but mechanistically distinct processes as they traffic cargo into or out of the cell, complicating direct comparisons with *Ca.* Y. umbracryptum cells that have a single continuous outer membrane.

We were unable catch intermediates in the process of vesicle fusion or fission. This would be expected if the process was fast. However, we cannot exclude the possibility that many of the vesicles were generated via a distinct archaeal-specific process. The ability of these cells to generate vesicles in an unusual way is also suggested by structures seen in these cells that are rarely seen in bacteria or eukaryotic cells, including vesicle membranes connected by three-way junctions **(Figs. 3d-e and Supplementary Fig. 10)**, along with large numbers of vesicles within vesicles **(Figs. 3d-f and 4c-d)**. At the same time, there were rare instances in which tomograms appeared to catch a new membrane compartment in the process of forming in a way that resembled the way multivesicular bodies are formed in eukaryotes[49] - with cytoplasm and ribosomes becoming enclosed on the inside (**Figs. 3d-f**). In addition, we also observed a very few cases in which extracellular material was seen in a deep invagination in the external membrane of a cell in tomograms, and one case of a *Ca.* Y. umbracryptum cell ensheathing another cell in the mixed culture **(Supplementary Fig. 11a-c)**. Moreover, while most vesicles were small, we observed rare instances in which cells had an internal compartment that contained dense material resembling abiotic material seen in the extracellular space **(Supplementary Fig. 11f)**, which suggested the possibility that they may be able to undergo a process similar to phagocytosis. Thus, it seems possible that *Ca.* Y. umbracryptum can generate a range of internal vesicles of different types, perhaps via distinct processes.

The genome of *Ca.* Y. umbracryptum provides few clues as to nature of the molecular machinery underpinning vesicle formation. Despite the abundance of intracellular membrane-bound compartments observed in some cells, *Ca.* Y. umbracryptum has a small genome in which we were unable to identify homologues of Sec13/31, Clathrin, COP-II, all of which are known to function in membrane trafficking in eukaryotes[50,51]. Furthermore, *Ca.* Y. umbracryptum does not contain obvious homologues of dynamin, which drives membrane scission in eukaryotes, but which has been reported to be of bacterial origin[52], nor did we observe any evidence of branched actin in the vicinity of vesicles or coat-like proteins in tomograms. In addition, *Ca.* Y. umbracryptum does not encode homologues of the proposed membrane trafficking associated complexes that are encoded in the genomes of other Asgard/Prometheoarchaeota[8,13]. Therefore, either the homologues of these proteins are far too divergent to be detected by standard bioinformatics or *Ca.* Y. umbracryptum cells use alternative machinery to generate internal vesicles. Nevertheless, the cells possess a large set of small GTPases[53] and, despite the lack of classical endocytotic machinery in *Ca.* Y. umbracryptum **(Supplementary Fig. 3b-c)**[50], we were able to identify multiple SPFH-family proteins in the genome that could plausibly form membrane-attached basket-like structures observed on the inner face of the plasma membrane **(Supplementary Fig. 7)**. These proteins are ubiquitous across the tree of life[54,55]

and are implicated in a variety of functions, such as protease regulation, membrane signalling and, in case of eukaryotic flotillin, clathrin-independent endocytosis[43]. Whether or not they could contribute to membrane remodelling in *Ca.* Y. umbracryptum remains to be explored.

While an understanding of vesicle formation in this system will have to await improvements in our ability to perform live cell imaging of these small anaerobes at subcellular resolution, the discovery of *Ca.* Y. umbracryptum challenges established views of archaeal cell biology and the timeline of events leading to the origin of eukaryotes. Since abundant internal vesicles were not seen in recent cell biological studies of the Lokiarchaeia[7,29], it is possible that intracellular membranes emerged in the Heimdallarchaeia which, under most phylogenetic analyses, are closer relatives of eukaryotes than Lokiarchaeia[8,13,14] **(Fig. 4f).** Alternatively, it is possible that vesicles were missed in ultrastructural studies of Lokiarchaeal cells because they only form under specific conditions, for example, in cells that have yet to use up the membrane reservoir in the form of protrusions (depicted by a question mark in **Fig. 4f)**.

Under either scenario, the capacity of *Ca.* Y. umbracryptum to generate and lose intracellular membrane-bound compartments in a way that depends on conditions supports a hypothesis in which the archaeal ancestor of eukaryotes had a dynamic endomembrane system prior to its association with the bacterial ancestor that gave rise to mitochondria. If the vesicles in *Ca.* Y. umbracryptum cells are used to exchange material with its environment to facilitate syntrophic partnerships, it is possible that an archaeal ancestor of eukaryotes engulfed the ancestor of mitochondria, perhaps aiding its ability to migrate into oxygen-rich environments[56]. In addition, if, as previously suggested[57], the central cell body of cells that contains the genome can functionally separate transcription from sites of translation and protein activity at the cell periphery, these cells may also possess a proto-nuclear compartment[57]. In these and other ways, the cell biological analysis of different members of the Asgard archaea is closing the gulf that was long assumed to separate bacteria and archaea from eukaryotes.

Etymology

Candidatus 'Yibarchaeum umbracryptum'. *Yiba* (Malgana); sister, archaeum, from *archaea,* (Latin, Greek Origin); primitive. *Umbra* (Latin); shadow, *cryptum* (Latin, Greek Origin); vault. The name describes an archaeon that is a sister lineage to eukaryotes containing hidden compartments.

Sample site description

Candidatus 'Yibarchaeum umbracryptum' was enriched from poorly lithified microbial mats collected from a tidal flat of a hypersaline pool in Shark Bay, Western Australia.

References

1 Vosseberg, J. *et al.* Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100 (2021). https://doi.org/10.1038/s41559-020-01320-z

2 Vosseberg, J. *et al.* The emerging view on the origin and early evolution of eukaryotic cells. *Nature* **633**, 295–305 (2024). https://doi.org/10.1038/s41586-024-07677-6

3 Baum, D. A. A comparison of autogenous theories for the origin of eukaryotic cells. *Am. J. Bot.* **102**, 1954–1965 (2015). https://doi.org/10.3732/ajb.1500196

4 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015). https://doi.org/10.1038/nature14447

5 Dacks, J. B. *et al.* The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *J. Cell. Sci.* **129**, 3695–3703 (2016). https://doi.org/10.1242/jcs.178566

6 Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020). https://doi.org/10.1038/s41586-019-1916-6

7 Rodrigues-Oliveira, T. *et al.* Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* **613**, 332–339 (2023). https://doi.org/10.1038/s41586-022-05550-y

8 Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023). https://doi.org/10.1038/s41586-023-06186-2

9 Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116**, 153–166 (2004). https://doi.org/10.1016/s0092-8674(03)01079-1

10 Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999). https://doi.org/10.1126/science.283.5407.1476

11 Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Phylogenetic affiliation of mitochondria with Alpha-II and Rickettsiales is an artefact. *Nat. Ecol. Evol.* **6**, 1829–1831 (2022). https://doi.org/10.1038/s41559-022-01871-3

12 Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967). https://doi.org/10.1016/0022-5193(67)90079-3

13 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017). https://doi.org/10.1038/nature21031

14 Zhang, J. *et al.* Deep origin of eukaryotes outside Heimdallarchaeia within Asgardarchaeota. *Nature* **642**, 990–998 (2025). https://doi.org/10.1038/s41586-025-08955-7

15 Klinger, C. M., Spang, A., Dacks, J. B. & Ettema, T. J. Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Mol. Biol. Evol.* **33**, 1528–1541 (2016). https://doi.org/10.1093/molbev/msw034

16 Makarova, K. S., Yutin, N., Bell, S. D. & Koonin, E. V. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* **8**, 731–741 (2010). https://doi.org/10.1038/nrmicro2406

17 Hatano, T. *et al.* Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat. Commun.* **13**, 3398 (2022). https://doi.org/10.1038/s41467-022-30656-2

18 Souza, D. P. *et al.* Asgard archaea reveal the conserved principles of ESCRT-III membrane remodeling. *Sci. Adv.* **11**, eads5255 (2025). https://doi.org/10.1126/sciadv.ads5255

19 Melnikov, N. *et al.* The Asgard archaeal ESCRT-III system forms helical filaments and remodels eukaryotic-like membranes. *EMBO J.* **44**, 665–681 (2025). https://doi.org/10.1038/s44318-024-00346-4

20 Akıl, C. & Robinson, R. C. Genomes of Asgard archaea encode profilins that regulate actin. *Nature* **562**, 439–443 (2018). https://doi.org/10.1038/s41586-018-0548-6

21 Akıl, C. *et al.* Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. *Proc Natl Acad Sci U S A* **117**, 19904–19913 (2020). https://doi.org/10.1073/pnas.2009167117

22 Gould, S. B., Garg, S. G. & Martin, W. F. Bacterial Vesicle Secretion and the Evolutionary Origin of the Eukaryotic Endomembrane System. *Trends Microbiol.* **24**, 525–534 (2016). https://doi.org/10.1016/j.tim.2016.03.005

23 Baum, D. A. & Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014). https://doi.org/10.1186/s12915-014-0076-2

24 MacLeod, F. I. *Ancient form: archaea take the shape of reduced syntrophs and eukaryotic progenitors*, UNSW Sydney, (2023).

25 Nobs, S.-J. *et al.* An Asgard archaeon from a modern analog of ancient microbial mats. *bioRxiv*, 2025.2007.2022.663070 (2025). https://doi.org/10.1101/2025.07.22.663070

26 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022). https://doi.org/10.1093/bioinformatics/btac672

27 Imachi, H. *et al.* Eukaryotes' closest relatives are internally simple syntrophic archaea. *bioRxiv*, 2025.2002.2026.640444 (2025). https://doi.org/10.1101/2025.02.26.640444

28 Zhu, P. *et al.* Expanded Archaeal Genomes Shed New Light on the Evolution of Isoprenoid Biosynthesis. *Microorganisms* **12** (2024). https://doi.org/10.3390/microorganisms12040707

29 Wollweber, F. *et al.* Microtubules in Asgard archaea. *Cell* **188**, 2451–2464 e2426 (2025). https://doi.org/10.1016/j.cell.2025.02.027

30 Avci, B. *et al.* Peculiar morphology of Asgard archaeal cells close to the prokaryote-eukaryote boundary. *mBio* **16**, e0032725 (2025). https://doi.org/10.1128/mbio.00327-25

31 Cui, Y. *et al.* Expansion microscopy using a single anchor molecule for high-yield multiplexed imaging of proteins and RNAs. *PLoS One* **18**, e0291506 (2023). https://doi.org/10.1371/journal.pone.0291506

32 Burt, A. *et al.* An image processing pipeline for electron cryo-tomography in RELION-5. *FEBS Open Bio.* **14**, 1788–1804 (2024). https://doi.org/10.1002/2211-5463.13873

33 Zivanov, J. *et al.* A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11** (2022). https://doi.org/10.7554/eLife.83724

34 Sas-Chen, A. *et al.* Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* **583**, 638–643 (2020). https://doi.org/10.1038/s41586-020-2418-2

35 Jha, V. *et al.* Structural basis of sequestration of the anti-Shine-Dalgarno sequence in the Bacteroidetes ribosome. *Nucleic Acids Res.* **49**, 547–567 (2021). https://doi.org/10.1093/nar/gkaa1195

36 Izore, T., Duman, R., Kureisaite-Ciziene, D. & Löwe, J. Crenactin from *Pyrobaculum calidifontis* is closely related to actin in structure and forms steep helical filaments. *FEBS Lett.* **588**, 776–782 (2014). https://doi.org/10.1016/j.febslet.2014.01.029

37 Bharat, T. A., Murshudov, G. N., Sachse, C. & Löwe, J. Structures of actin-like ParM filaments show architecture of plasmid-segregating spindles. *Nature* **523**, 106–110 (2015). https://doi.org/10.1038/nature14356

38 Stoddard, P. R., Williams, T. A., Garner, E. & Baum, B. Evolution of polymer formation within the actin superfamily. *Mol. Biol. Cell.* **28**, 2461–2469 (2017). https://doi.org/10.1091/mbc.E15-11-0778

39 von Kügelgen, A., Alva, V. & Bharat, T. A. M. Complete atomic structure of a native archaeal cell surface. *Cell Rep.* **37**, 110052 (2021). https://doi.org/10.1016/j.celrep.2021.110052

40 Gambelli, L. *et al.* Structure of the two-component S-layer of the archaeon *Sulfolobus acidocaldarius*. *Elife* **13** (2024). https://doi.org/10.7554/eLife.84617

41 Foo, S., Caspy, I., Cezanne, A., Bharat, T. A. M. & Baum, B. A self-assembling surface layer flattens the cytokinetic furrow to aid cell division in an archaeon. *Proc Natl Acad Sci U S A* **122**, e2501044122 (2025). https://doi.org/10.1073/pnas.2501044122

42 Fu, Z. & MacKinnon, R. Structure of the flotillin complex in a native membrane environment. *Proc Natl Acad Sci U S A* **121**, e2409334121 (2024). https://doi.org/10.1073/pnas.2409334121

43 Bodin, S., Elhabashy, H., Macdonald, E., Winter, D. & Gauthier-Rouviere, C. Flotillins in membrane trafficking and physiopathology. *Biol. Cell* **117**, e2400134 (2025). https://doi.org/10.1111/boc.202400134

44 Bharat, T. A. M., Hoffmann, P. C. & Kukulski, W. Correlative Microscopy of Vitreous Sections Provides Insights into BAR-Domain Organization In Situ. *Structure* **26**, 879–886 e873 (2018). https://doi.org/10.1016/j.str.2018.03.015

45 Eskelinen, E. L. Novel insights into autophagosome biogenesis revealed by cryo-electron tomography. *FEBS Lett.* **598**, 9–16 (2024). https://doi.org/10.1002/1873-3468.14726

46 Babst, M., Katzmann, D. J., Estepa-Sabal, E. J., Meerloo, T. & Emr, S. D. Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting. *Dev. Cell.* **3**, 271–282 (2002). https://doi.org/10.1016/s1534-5807(02)00220-4

47 Schiller, H. *et al.* Identification of structural and regulatory cell-shape determinants in *Haloferax volcanii*. *Nat. Commun.* **15**, 1414 (2024). https://doi.org/10.1038/s41467-024-45196-0

48 Rados, T. *et al.* Tissue-like multicellular development triggered by mechanical compression in archaea. *Science* **388**, 109–115 (2025). https://doi.org/10.1126/science.adu0047

49 Klumperman, J. & Raposo, G. The complex ultrastructure of the endolysosomal system. *Cold Spring Harb. Perspect. Biol.* **6**, a016857 (2014). https://doi.org/10.1101/cshperspect.a016857

50 Kumari, S., Mg, S. & Mayor, S. Endocytosis unplugged: multiple ways to enter the cell. *Cell Res.* **20**, 256–275 (2010). https://doi.org/10.1038/cr.2010.19

51 Dacks, J. B. & Field, M. C. Evolutionary origins and specialisation of membrane transport. *Curr. Opin. Cell. Biol.* **53**, 70–76 (2018). https://doi.org/10.1016/j.ceb.2018.06.001

52 Low, H. H., Sachse, C., Amos, L. A. & Löwe, J. Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell* **139**, 1342–1352 (2009). https://doi.org/10.1016/j.cell.2009.11.003

53 Vargová, R. *et al.* The Asgard archaeal origins of Arf family GTPases involved in eukaryotic organelle dynamics. *Nat. Microbiol.* **10**, 495–508 (2025). https://doi.org/10.1038/s41564-024-01904-6

54 Rivera-Milla, E., Stuermer, C. A. & Malaga-Trillo, E. Ancient origin of reggie (flotillin), reggie-like, and other lipid-raft proteins: convergent evolution of the SPFH domain. *Cell. Mol. Life Sci.* **63**, 343–357 (2006). https://doi.org/10.1007/s00018-005-5434-3

55 Hinderhofer, M. *et al.* Evolution of prokaryotic SPFH proteins. *BMC Evol. Biol.* **9**, 10 (2009). https://doi.org/10.1186/1471-2148-9-10

56 Craig, J. M., Kumar, S. & Hedges, S. B. The origin of eukaryotes and rise in complexity were synchronous with the rise in oxygen. *Front. Bioinform.* **3**, 1233281 (2023). https://doi.org/10.3389/fbinf.2023.1233281

57 Baum, B. & Spang, A. On the origin of the nucleus: a hypothesis. *Microbiol. Mol. Biol. Rev.*, e0018621 (2023). https://doi.org/10.1128/mmbr.00186-21

58 Allen, M. A., Goh, F., Burns, B. P. & Neilan, B. A. Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. *Geobiology* **7**, 82–96 (2009). https://doi.org/10.1111/j.1472-4669.2008.00187.x

59 Widdel, F., Kohring, G. W. & Mayer, F. Studies on Dissimilatory Sulfate-Reducing Bacteria That Decompose Fatty-Acids .3. Characterization of the Filamentous Gliding Desulfonema-Limicola Gen-Nov Sp-Nov, and Desulfonema-Magnum Sp-Nov. *Arch Microbiol* **134**, 286–294 (1983). https://doi.org/Doi 10.1007/Bf00407804

60 Janssen, P. H., Schuhmann, A., Morschel, E. & Rainey, F. A. Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl. Environ. Microbiol.* **63**, 1382–1388 (1997). https://doi.org/10.1128/aem.63.4.1382-1388.1997

61 FastQC: a quality control tool for high throughput sequence data (Bioinformatics, Babraham, 2011).

62 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016). https://doi.org/10.1093/bioinformatics/btw354

63 Wick, R. rrwick/Porechop. (2025). https://doi.org/https://github.com/rrwick/Porechop

64 Nygaard, A. B., Tunsjo, H. S., Meisal, R. & Charnock, C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Sci Rep* **10**, 3209 (2020). https://doi.org/10.1038/s41598-020-59771-0

65 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018). https://doi.org/10.1093/bioinformatics/bty560

66 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014). https://doi.org/10.1186/gb-2014-15-3-r46

67 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–596 (2013). https://doi.org/10.1093/nar/gks1219

68 Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3** (2017). https://doi.org/10.7717/peerj-cs.104

69 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). https://doi.org/10.1093/bioinformatics/btu170

70 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012). https://doi.org/10.1089/cmb.2012.0021

71 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017). https://doi.org/10.1101/gr.213959.116

72 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* (2013). https://doi.org/10.48550/arXiv.1303.3997

73 Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014). https://doi.org/10.1038/nmeth.3103

74 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019). https://doi.org/10.7717/peerj.7359

75 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016). https://doi.org/10.1093/bioinformatics/btv638

76 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018). https://doi.org/10.1038/s41564-018-0171-1

77 Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023). https://doi.org/10.1038/s41592-023-01940-w

78 Seemann, T. tseemann/barrnap. (2025). https://doi.org/https://github.com/tseemann/barrnap

79 Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021). https://doi.org/10.1093/nar/gkab688

80 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). https://doi.org/10.1186/1471-2105-11-119

81 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**, 1079–1086 (2020). https://doi.org/10.1038/s41587-020-0501-8

82 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). https://doi.org/10.1093/molbev/mst010

83 Criscuolo, A. & Gribaldo, S. 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). https://doi.org/10.1186/1471-2148-10-210

84 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). https://doi.org/10.1093/molbev/msaa015

85 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018). https://doi.org/10.1093/molbev/msx281

86 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017). https://doi.org/10.1038/nmeth.4285

87 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024). https://doi.org/10.1093/nar/gkae268

88 Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017). https://doi.org/10.1093/molbev/msx148

89 Sillitoe, I. *et al.* CATH: increased structural coverage of functional space. *Nucleic Acids Res.* **49**, D266–D273 (2021). https://doi.org/10.1093/nar/gkaa1079

90 Lewis, T. E. *et al.* Gene3D: Extensive prediction of globular domains in proteins. *Nucleic Acids Res.* **46**, D435–D439 (2018). https://doi.org/10.1093/nar/gkx1069

91 Marchler-Bauer, A. *et al.* CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **35**, D237–240 (2007). https://doi.org/10.1093/nar/gkl951

92 Pedruzzi, I. *et al.* HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Research* **43**, D1064–D1070 (2015). https://doi.org/10.1093/nar/gku1002

93 Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22 (2022). https://doi.org/10.1002/pro.4218

94 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021). https://doi.org/10.1093/nar/gkaa913

95 Wu, C. H. *et al.* PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.* **32**, D112–114 (2004). https://doi.org/10.1093/nar/gkh097

96 Attwood, T. K. *et al.* The PRINTS database: a fine-grained protein sequence annotation and analysis resource--its status in 2012. *Database (Oxford)* **2012**, bas019 (2012). https://doi.org/10.1093/database/bas019

97 Sigrist, C. J. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344–347 (2013). https://doi.org/10.1093/nar/gks1067

98 Akiva, E. *et al.* The Structure-Function Linkage Database. *Nucleic Acids Res.* **42**, D521–530 (2014). https://doi.org/10.1093/nar/gkt1130

99 Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**, D458–D460 (2021). https://doi.org/10.1093/nar/gkaa937

100 Wilson, D., Madera, M., Vogel, C., Chothia, C. & Gough, J. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* **35**, D308–313 (2007). https://doi.org/10.1093/nar/gkl910

101 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). https://doi.org/10.1093/bioinformatics/btu031

102 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016). https://doi.org/10.1016/j.jmb.2015.11.006

103 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w

104 Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* **53**, D444–D456 (2025). https://doi.org/10.1093/nar/gkae1082

105 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021). https://doi.org/10.1002/pro.3943

106 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). https://doi.org/10.1016/j.jsb.2005.07.007

107 Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017). https://doi.org/10.1016/j.jsb.2016.06.007

108 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). https://doi.org/10.1038/nmeth.4193

109 Zheng, S. *et al.* AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *J. Struct. Biol. X* **6**, 100068 (2022). https://doi.org/10.1016/j.yjsbx.2022.100068

110 Buchholz, T.-O., Jordan, M., Pigino, G. & Jug, F. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019).* 502–506.

111 Buchholz, T.-O. *et al.* in *Methods Cell Biol.* Vol. 152 *Three-Dimensional Electron Microscopy* (eds Thomas Müller-Reichert & Gaia Pigino) 277–289 (Academic Press, 2019).

112 Chaillet, M. L., Roet, S., Veltkamp, R. C. & Förster, F. pytom-match-pick: A tophat-transform constraint for automated classification in

template matching. *J. Struct. Biol. X* **11**, 100125 (2025). https://doi.org/10.1016/j.yjsbx.2025.100125
113 Ermel, U. H., Arghittu, S. M. & Frangakis, A. S. ArtiaX: An electron tomography toolbox for the interactive handling of sub-tomograms in UCSF ChimeraX. *Protein Sci.* **31**, e4472 (2022). https://doi.org/10.1002/pro.4472
114 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996). https://doi.org/10.1006/jsbi.1996.0013
115 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). https://doi.org/10.1038/nmeth.2019
116 Dragonfly 3D World (Version 2024.1) v. 2024.1 (Comet Technologies Canada Inc., Montreal, QC, 2024).
117 Heebner, J. E. *et al.* Deep Learning-Based Segmentation of Cryo-Electron Tomograms. *J. Vis. Exp.* (2022). https://doi.org/10.3791/64435
118 Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sc* **9351**, 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
119 Zeiler, M. D. ADADELTA: An Adaptive Learning Rate Method. *arXiv* (2012). https://doi.org/10.48550/arXiv.1212.5701
120 Lamm, L. *et al.* MemBrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography. *bioRxiv*, 2024.2001.2005.574336 (2025). https://doi.org/10.1101/2024.01.05.574336
121 Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021). https://doi.org/10.21105/joss.03021
122 Valentin-Alvarado, L. E. *et al.* Asgard archaea modulate potential methanogenesis substrates in wetland soil. *Nat. Commun.* **15**, 6384 (2024). https://doi.org/10.1038/s41467-024-49872-z

Acknowledgements

B.B. and T.A.M.B. groups received support from the Medical Research Council, as part of United Kingdom Research and Innovation (also known as UK Research and Innovation) [Programme MC_UP_1201/31 to T.A.M.B] and [Programme MC_UP_1201/27 to B.B., which supported J.P. and I.R.]. B.B. also received support from the Wellcome Trust (222460/Z/21/Z), the Moore-Simons Project on the Origin of the Eukaryotic Cell, Simons Foundation 735929LPI and from the Gordon and Betty Moore Foundation's Symbiosis in Aquatic Systems Initiative (9346). In addition, B.B. would like to thank the Allen Distinguished Investigator program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation for support for F.I.M. T.A.M.B. would like to thank the Wellcome Trust (grant 225317/Z/22/Z) and the Lister Institute for Preventative Medicine for support. M.L. was supported by the Herchel Smith Fund PhD Scholarship. E.J.A-P. was supported by the Moore-Simons Project on the Origin of the Eukaryotic Cell, Simons Foundation grant 73592LPI to B.J.B., and a Simons Foundation investigator award LI-SIAME-00002001 to B.J.B. The authors also wish to thank Brendan Burns, Pieter Visscher, Tony Larkum, and Therese Morris for collecting the original microbial mat samples, as previously described[24,25]. This study was also supported by the MRC Laboratory of Molecular Biology Electron Microscopy and Scientific Computing Facility. This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. The authors thank Jan Löwe, Aurélien Roux, David Baum, Roni Levin-Konigsberg, Sean Munro and Gautam Dey for constructive suggestions on the manuscript. For the purpose of open access, the MRC Laboratory of Molecular Biology has applied a CC BY public copyright license to any Author Accepted Manuscript version arising. This preprint was typeset using a modified word template originally available at www.github.com/chrelli/bioRxiv-word-template.

Author contributions

F.I.M. performed cell cultivation, genome assembly, live cell and HCR FISH imaging; A.v.K. carried out the cryo-ET experiments; J.P. and I.R. performed the ExM FISH experiments; F.I.M., A.v.K., T.A.M.B., B.B., M.L., J.P. and I.R. analysed experimental data; F.I.M., M.L., E.J.A-P. and B.J.B analysed the genomic data; F.I.M., A.v.K., T.A.M.B. and B.B. designed the research; F.I.M., A.v.K. T.A.M.B. and B.B. wrote the manuscript with the support of all authors. B.B., T.A.M.B. and B.J.B raised the funds to support the work.

Competing interests

The authors declare no competing financial interests and no conflict of interest.

Data availability

Subtomogram averaging maps will be deposited in the Electron Microscopy Data Bank (EMDB). For more details see Supplementary Table 2.

Correspondence and requests for materials

Correspondence and requests for materials should be addressed to the corresponding authors T.A.M.B. (tbharat@mrc-lmb.cam.ac.uk) or B.B. (bbaum@mrc-lmb.cam.ac.uk).

Materials and Methods

Growth of *Ca.* Y. umbracryptum

Poorly lithified smooth microbial mat samples collected in 2019 from the Nilemah tidal flats of Hamelin Pool in Shark Bay (26°27'336''S, 114°05'762''E) were used as the initial inoculum to establish enrichment cultures. Briefly, serum bottles containing 50 mL of salt solution reflective of Hamelin Pool seawater (50.7 g/L NaCl, 13.3 g/L $MgSO_4{\cdot}7H_2O$, 7.23 g/L $MgCl_2{\cdot}6H_2O$, 2.7 g/L $CaCl_2{\cdot}2H_2O$, 1.4 g/L KCl)[58] were supplemented with 0.05% (w/v) yeast extract, 0.05% (w/v) casamino acids and 100 µg/mL resazurin. Serum bottles were sparged with 80:20 N_2:CO_2 (v/v) gas and sealed with butyl rubber stoppers and crimp sealed. The solution was further reduced by the addition of 0.3 g/L L-cysteine HCl and 0.3 g/L $Na_2S{\cdot}9H_2O$. Serum bottles were autoclaved prior to the addition of 1 mL/L SL10 trace element solution[59], 3 mL/L Vitamin 10 Solution[60] and the pH was adjusted to 7.4. To inoculate cultures, serum bottles were opened and continuously flushed with 80:20 N_2:CO_2 (v/v) gas and 2 g of homogenized microbial mat was added to each bottle. Each bottle was then resealed with butyl rubber stoppers and crimp sealed. The headspace was reestablished by flushing with 80:20 N_2:CO_2 (v/v) gas. Bottles were incubated at 37 °C for 90 days. For subsequent transfers over the next 270 days, 1 mL of enrichment culture was added to 49 mL of medium. The medium was unchanged bar the addition of 4 g/L microbial mat sample to the salt solution which was sterilized through repeated autoclaving and the final pH adjustment was performed by adjusting to pH 7.5 with a 10% (w/v) solution of Na_2CO_3. For subsequent transfers after 360 days post inoculation, media was adjusted to include 100 µg/mL (w/v) ampicillin, 50 µg/mL (w/v) kanamycin, and 100 µg/mL (w/v) streptomycin which was added after autoclaving and prior to pH adjustment. For three rounds of the enrichment, the media was additionally supplemented with 64 µg/mL of ciprofloxacin.

After a further four years of continuous cultivation under the previously described media, the media was adjusted to an alternative media herein referred to as Hamelin Pool Media (HPM, see Supplementary Table 3). HPM contained 50.7 g/L NaCl, 13.3 g/L $MgSO_4{\cdot}7H_2O$, 7.23 g/L $MgCl_2{\cdot}6H_2O$, 2.7 g/L $CaCl_2{\cdot}2H_2O$, 1.4 g/L KCl and was supplemented with 0.05% (w/v) yeast extract and 0.05% (w/v) casein hydrolysate and 100 µg/mL resazurin in serum bottles. The bottles were again reduced by sparging with 80:20 N_2:CO_2 (v/v) gas and further reduced by the addition of 0.3 g/L L-cysteine HCl and 0.3 g/L $Na_2S{\cdot}9H_2O$. The headspace was then established by flushing with 80:20 N_2:CO_2 (v/v) gas and setting final pressure to 0.5 bar. Bottles were then autoclaved. After autoclaving, 100 µg/mL (w/v) ampicillin, 50 µg/mL (w/v) kanamycin, and 100 µg/mL (w/v) streptomycin was added to each serum bottle along with 1 mL/L SL10 trace element solution, 3 mL/L V10 solution, 1 mL/L Se/W/Mo solution containing 0.021 g/L Na_2SeO_3, 0.037 g/L Na_2WO_4, 0.021g/L Na_2MoO_4 and 50 µL of 1 M KH_2PO_4 solution (pH 7.5). Finally, the pH was adjusted to pH 7.5 with a 10% (w/v) solution of Na_2CO_3. For subculturing in HPM, 5 mL of cell culture was added to 45 mL of HPM media and cultures were transferred every 60 days.

The enrichment cultures were regularly checked for their abundance, both absolute and relative, through qPCR and amplicon sequencing, respectively. DNA was extracted from 1 mL of enrichment cultures using DNeasy PowerBiofilm Kits, according to manufacturer's instructions. For qPCR, the copy number of 16S was determined by running extracted DNA on a ViiA 7 Real-Time PCR system using the PowerTrack SYBR Green Master Mix and the *Ca.* Y. umbracryptum 16S specific primers HEI245F (5' - TCGGTAGGGGCTATGAGAGT -3') and HEI416R (5' - TTGCTGCCGGAGTTTTACAC - 3') against a standard curve of known template concentrations. For relative abundance determination, PCR was performed using the universal primer pair 926F (5' - AAACTYAAAKGAATTGRCGG - 3') and 1392wR (5' - ACGGGCGGTGWGTRC - 3'). PCR products were cleaned

using a Monarch PCR & DNA Cleanup Kit. A long-read sequencing library was prepared using a v14 library and sequenced by Oxford Nanopore Technologies. Quality checks were performed on fastq reads with FastQC and MultiQC[61,62]. Adaptor sequences and chimeras were removed from fastq reads using Porechop[63]. Reads were filtered to meet quality standards using fastp[64,65]. Read quality was then reevaluated using FastQC and MultiQC[61,62]. Taxonomy was assigned to reads using Kraken2 using the SILVA SSU 138 database[66,67]. Relative abundance of the Kraken 2 assigned reads was determined using Bracken[68].

Genome assembly, binning, and taxonomic placement

For characterization of the *Ca*. Y. umbracryptum genome, DNA was extracted from 5 mL of 90 day culture in triplicate using a potassium ethyl xanthogenate/sodium dodecyl sulfate (SDS) lysis buffer, as described previously[24]. DNA was purified from cell lysate using a phenol:chloroform:isoamyl alcohol extraction followed by an ethanol precipitation with the inclusion of glycogen. DNA quality and concentration was assessed using Xpose (Trinean, Belgium), NanoDrop (ThermoFisher Scientific, U.S.A) and Qubit (ThermoFisher Scientific, U.S.A) spectrophotometers. Short read library preparation was performed using a Nextera DNA Flex Library Prep Kit (Illumina, U.S.A). DNA was sequenced on a NovaSeq 6000 using a NovaSeq 6000 SP 2x150bp flowcell (Illumina, U.S.A). DNA sequencing was performed at the Ramaciotti Centre for Genomics (The University of New South Wales, Australia).

The quality of fastq reads was assessed with FastQC[61] and low-quality bases were removed with Trimmomatic[69]. The resulting triplicate trimmed fastq reads were co-assembled into contigs using SPAdes v.3.15.5 with the -meta option[70,71]. Assembled contigs were mapped against the unassembled DNA sequences using BWA-MEM[72]. Contigs shorter than 1000 bp were removed and resulting contigs were used to generate bins with Metabat v.2.12.1, CONCOCT v.1.0 and MaxBin2 v.2.2.3[73-75]. Bins generated were dereplicated and refined using Das Tool[76]. Genome quality checks for completeness and contamination were performed with CheckM2[77]. Barrnap v.0.9 was used to extract rRNAs[78]. tRNAs were extracted using tRNAscan-SE[79]. The nucleotide sequences of draft genomes were predicted with Prodigal v.2.6.3[80].

Draft genomes were provided an initial taxonomic placement using GTDB-Tk against the GTDB database [26,81]. A single Heimdallarchaeaceae MAG was detected as belonging to the Heimdallarchaeia order UBA460, family UBA460, genus novel, species novel (*syn*. order Heimdallarchaeales, family Heimdallarchaeaceae). Taxonomic placement was further validated by constructing a phylogenetic dataset of archaea for further phylogenetic analyses, including the target Heimdallarchaeaceae genome generated as part of this study. Briefly, a set of archaeal genomes was downloaded from GTDB on the 4th of August 2025 meeting the following criteria: CheckM2 completeness > 80%, CheckM2 contamination < 10%, and GTDB representative of species = true. The genome of highest completeness for each order was selected for DPANN archaea, Euryarchaeota and Thermoproteota, except in cases where orders contained well characterized representatives. The genome with the highest completeness for each Asgardarchaeota genus was included, and all Heimdallarchaeia above 90% completeness were included. Marker genes of the ar53 dataset were extracted with GTDB-tk identify[26] and aligned with MAFFT v.7 using the L-INS-I iterative refinement method[82]. The resulting alignment was trimmed with BMGE with entropy estimation according to the BLOSUM30 matrix and a gas-rate cutoff of 50%[83]. BMGE trimmed alignments were further manually inspected and trimmed as required. Maximum-likelihood phylogenetic trees were constructed with IQ-TREE 2 with ultrafast bootstrap approximation (option -bb 10000) and model selection with ModelFinder Plus (selected as LG+F+R10)[84-86]. The resulting phylogenetic tree was visualized and annotated in iTOL[87].

For the placement of the Heimdallarchaeaceae genome generated as part of this study in an archaeal tree including eukaryotes, the NM57 marker genes from a previously published taxonomic dataset were collected[8]. NM57 marker genes from the Heimdallarchaeaceae genome were annotated with eggNOG Mapper v.2.1.8 and aligned to reference markers with MAFFT v.7 using the L-INS-I iterative refinement method[82,88]. Trimming and phylogenetic tree calculation for the NM57 dataset was performed identically to the ar53 dataset (model LG+F+R10). The resulting phylogenetic tree was visualized and annotated with iTOL[87].

Genomic analysis

To functionally characterize the *Ca*. Y. umbracryptum genome, protein sequences were annotated against the CATH-Gene3D[89,90], CDD[91], HAMAP[92], PANTHER[93], Pfam[94], PIRSF[95], PRINTS[96], PROSITE profiles[97], PROSITE patterns[97], SFLD[98], SMART[99], and SUPERFAMILY[100] databases with InterProScan v.5.44-79.0[101]. The genome was additionally annotated against the eggNOG database with eggNOG Mapper v.2.1.8[88] and KEGG database with Ghostkoala v.3.1[102]. The resulting InterProScan and Ghostkoala annotation were used to provide the annotation of ESPs and metabolic genes, respectively. Four SPFH-family protein candidates in the *Ca*. Y. umbracryptum genome were identified by Band-7/SPFH-domain annotation (PANTHER: IPR043202; SMART and Pfam: IPR001107; SUPERFAMILY: IPR036013; Pfam and CDD: IPR033880). AlphaFold3[103] modelling of the candidates revealed two putative proteins with an elongated alpha-helix following the domains SPFH1 and 2, typical for basket-forming SPFH family members (NODE_169_length_248565_cov_475.496599_54 and NODE_955_length_102699_cov_479.414219_56, named CC1-SPFH A and CC1-SPFH B respectively). Representatives of basket-forming proteins from other organisms were chosen from the literature and by taxonomic search of the Band-7 domain containing proteins annotated in InterPro (IPR001107)[104]. Detailed domain annotation was performed based on AlphaFold3[103] structural prediction of monomeric proteins. All structures shown were rendered using ChimeraX[105].

Light microscopy

Cells from late-stage (stationary phase) cultures in duplicates were allowed to settle onto polyethyleneimine-coated glass imaging chambers in an anaerobic chamber before a brief centrifugation at 1000 rcf (relative centrifugal force). Cells were then stained with Mitotracker Red CMXRos (200 nM, Invitrogen) and Hoechst 33342 (40 μM, Invitrogen) to label the membranes and DNA, respectively. Wells were overlaid with mineral oil to minimise oxygen exposure, and cells were immediately imaged on a SoRa spinning disk confocal microscope to avoid oxygen-induced toxicity over time. Images were collected with a pixel size of 27.5 nm^2. An identical methodology was used to image cells in early-stage cultures, which had yet to reach stationary phase (see Fig. 1c).

HCR FISH microscopy

For the identification of *Ca*. Y. umbracryptum in enrichment cultures by light microscopy, cells of early-stage and late-stage cultures (at times indicated in Fig. 1c) in duplicate were allowed to settle onto 1% polyethyleneimine-coated glass imaging chambers in an anaerobic chamber prior to centrifugation at 1000 rcf. Cells were then fixed with 1.6 % formaldehyde at 4 °C for 15 minutes. After formaldehyde fixation, cells were washed three times with the anaerobic Hamelin Pool Salt (HPS) buffer. HPS buffer contained 50.7 g/L NaCl, 13.3 g/L $MgSO_4{\cdot}7H_2O$, 7.23 g/L $MgCl_2{\cdot}6H_2O$, 2.7 g/L $CaCl_2{\cdot}2H_2O$, 1.4 g/L KCl, which was sparged with 80:20 N_2:CO_2 (v/v) gas, and adjusted to pH 7.5 with 10 % (w/v) Na_2CO_3. Cells were then permeabilised with 0.01 % (w/v) SDS in HPS for 15 minutes at room temperature. To specifically tag *Ca*. Y. umbracryptum 16S rRNA fluorescently, Hybridised Chain Reaction Fluorescent in situ Hybridisation (HCR-FISH) probes were designed for the target organism by Molecular Instruments (U.S.A). Cells were pre-hybridised using HCR HiFi Probe Hybridization Buffer for 10 minutes at 37 °C. Cells were hybridized with 2 μL HCR HiFi Probe in 100 μL hybridisation buffer at 37 °C overnight. Cells were then washed four times with HCR HiFi Probe wash buffer at 37 °C for 15 minutes, and incubated with 200 μL of HCR Gold Amplification buffer at room temperature for 30 minutes. Amplified hairpins were heated to 95 °C for 90 seconds before cooling to room temperature in the dark for 30 minutes. Hairpins were then added to HCR Gold Amplifier Buffer (2 μL hairpin to 100 μL amplifier buffer), and cells incubated in the dark with the hairpin HCR Gold Amplifier Buffer solution at room temperature overnight. Cells were then washed with HCR Gold Amplifier Wash Buffer four times for 15 minutes at room temperature. Finally, cells were stained with NHS ester 488 in phosphate buffered saline (PBS) for at least one hour at room temperature prior to three washes with PBS, overlaid with 200 μL of PBS, and imaged on a SoRa spinning disk confocal microscope (pixel size 27.5 nm^2).

HCR FISH Expansion microscopy

We adapted a previously described protocol[31] for our experiments. Briefly, cells were washed three times for 15 minutes each with sodium bicarbonate (100 mM) before treatment with Glycidyl methacrylate (GMA, 0.08 %) dissolved in sodium bicarbonate (100 mM) for 3 hours at room temperature. Cells were then cross-linked overnight in freshly prepared 1 %

acrylamide/0.7 % formaldehyde solution in PBS. The following day cells were washed once with PBS before the addition of gelation solution (19 % (w/w) sodium acrylate, 10% (w/w) acrylamide, 0.1 % (w/w) bis-acrylamide in 10x PBS), each 18 µL aliquot was activated with 1 µL 10 % (w/v) ammonium persulfate (APS) and 1 µL tetramethylethylenediamine (TEMED) and mixed by vortexing to make two gels of 9 µL solution each. The gels were left on ice for 5 minutes before transfer to a humidified chamber which was placed at 37 °C for at least one hour. Following gelation, the cells were denatured at 95 °C with denaturation buffer (50 mM TRIS pH 8, 200 mM NaCl and 200 mM SDS in Millipore water). Following denaturation, gels were left to cool to room temperature before washing twice with PBS and then four successive 30-minute expansion steps with Millipore water at room temperature on a shaker. The water was exchanged once more and left at 4 °C overnight to fully expand. After overnight expansion, the full-size gels were cut into four quarters and shrunk by two 15-minute PBS washes on a shaker before storage in PBS with 0.02 % (w/v) sodium azide at 4 °C.

HCR-FISH was used to label rRNA for light microscopy. First, the HCR-HiFi Probe Hybridization Buffer was pre-heated to 37 °C. The expansion gels were then pre-hybridized in 200 µL of HiFi Probe Hybridization Buffer at 37 °C for 30 minutes. The pre-hybridisation solution was then removed, and 100 µL of probe hybridisation solution containing the HCR probes (2 pM) was added. The samples were then left to hybridise overnight at 37 °C without agitation. The samples were then washed in 500 µL of HCR HiFi Probe Wash Buffer (4 x 15 minute) at 37 °C. Following this, the samples were pre-amplified for 30 minutes at room temperature in 200 µL of HCR Gold Amplifier Buffer. 3 µL of hairpin-1 and 3 µL of hairpin-2 per sample were heated for 90 seconds at 95 °C and then cooled at room temperature in the dark. The amplifier solution was then made by combining the hairpin-1 and hairpin-2 into 100 µL (per sample) of HCR Gold Amplification Buffer at room temperature. The pre-amplification solution was then removed and replaced with the amplifier solution containing hairpins. The samples were again left to incubate overnight at room temperature in the dark for amplification. The amplifier solution was then removed, and excess hairpins were removed from the sample by washing 4 times for 15 min with 500 µL of HCR Gold Amplifier Wash Buffer at room temperature. Samples were then washed with PBS. Cells embedded in the gels were stained with NHS ester 647 or Bodipy in PBS (1:750) at room temperature for at least an hour before two washes with PBS followed by 3 × 20 min expansion steps with Millipore water. Gels were then imaged on the same day using the microscopy setup described above.

Cryo-ET sample preparation

Cryo-ET grid preparation was performed either completely anaerobically within an anaerobic glove box using a manual plunger or with a Leica EM GP2 under ambient conditions. Within the anaerobic chamber, rubber caps of serum bottles were sterilised with 70 % (v/v) ethanol and ~100 µL of sample were aseptically removed using a metal syringe. Prior to accessing cells, syringes were coated with 2 mg/mL Bovine Serum Albumin (BSA) (to minimise sample loss due sticking to the needle), formulated in 0.9% saline solution and then washed with 70 % (v/v) ethanol before being allowed to dry in an anaerobic chamber. Next, 3.5 µL of the specimen were directly applied to a freshly glow discharged Quantifoil R2/2 Au 200 mesh grid, blotted for 4-5 s and plunge-frozen into liquid ethane with a manual plunger. Vitrified grids were then transferred under liquid nitrogen, clipped and stored until cryo-ET data collection was performed. Similarly for ambient condition grid preparation, samples were aseptically removed from serum bottles and then 3.5 µL of the specimen was directly applied to a freshly glow discharged Quantifoil R2/2 Au 200 mesh grid, blotted for 4-5 s and plunge-frozen into liquid ethane with a Leica EM GP2, while the blotting chamber was maintained at 100% humidity at 18 ºC. Grids were clipped and stored under liquid nitrogen until cryo-ET data collection was performed.

Cryo-ET data collection

Cryo-ET data was collected on a either Titan Krios G3 equipped with a K3 Gatan detector behind a Gatan Imaging Filter (20 eV slit width) or a Titan Krios G4 equipped with a Falcon 4i camera behind a Selectris-X Energy filter (10 eV slit width) using the SerialEM software[106] and a dose symmetric tilting scheme[107] at 300 kV. Three tilt series datasets for late-log phase *Ca.* Y. umbracryptum cells were collected at a pixel size 3.01 Å with a defocus range of -4 to -8 µm, ±60° oscillation, 3º increments with a total dose of ~140 $e^-/Å^2$ using the Titan Krios G4 (see Supplementary Table 2). Another biological replicate for late-log phase *Ca.* Y. umbracryptum cells were collected using the Titan Krios G3 at a pixel size 2.13 Å with a defocus range of -5 to -8 µm, ±60° oscillation, 2º increments with a total dose of ~120 $e^-/Å^2$. All early-log phase *Ca.* Y. umbracryptum cells and replicates were collected using the Titan Krios G3 at a pixel size 2.13 Å with a defocus range of -5 to -8 µm, ±60° oscillation, 2º increments with a total dose of ~120-122 $e^-/Å^2$.

Cryo-ET data analysis and subtomogram averaging

Cryo-ET data processing was performed using RELION5.0[32]. After tilt series import, tilt-series movies were motion-corrected and split into odd and even images with MotionCor2[108] implemented in RELION-5.0. Tilt-series alignment and initial contrast transfer functions (CTFs) estimation were performed using ARETOMO2[109]. Tomograms were generated using RELION5.0 with options (--Fc --ctf –pre_weight --SNR 100) and denoised with Cryo-CARE[110,111]. Ribosomes were initially picked with pytom-match-pick[112] using EMDB-10223[34] as a reference and manually inspected using the ArtiaX plugin[113] implemented in UCSF ChimeraX[105]. Subtomogram averaging was performed using RELION5.0 for multiple cycles of 2D-tilt-image extraction, realignment and CTF-optimisation as described previously[33] (see Supplementary Table 2 and Supplementary Fig. 6 and 9). Figure panels containing cryo-ET images were prepared using IMOD[114], Fiji[115] and UCSF ChimeraX[105] using the ArtiaX plugin[113].

Semantic and membrane segmentation

Semantic segmentation was performed using Dragonfly (2024.1.0.1627)[116]. Briefly, Cryo-CARE denoised tomograms were converted into float32 TIFF files using the newstack function of the IMOD[114] package and then imported into Dragonfly with flip around the Y-axis option. Features (membrane, cell surface, filaments, pili, vesicles and inorganic granules) were manually annotated using a Wacom Cintiq Pro 16 pen display on three to five slices per tomogram as recently described[117]. Remaining un-labelled voxels were assigned to a background class and then used to train a semantic segmentation model with a 2.5D U-Net[118] architecture containing 5-slices and a depth level of 5. Training was performed with a patch size of 128 over 100 epochs using Categorical Cross-Entropy loss function with the Adadelta optimisation algorithm[119]. All other training parameters were kept as default within Dragonfly. After training, segmentation was performed within Dragonfly, wrongly assigned voxels were manually cleaned using the process island function and then rendered for visualisation using UCSF ChimeraX[105].

Membrane segmentation was performed using Membrain v2[120] with default parameters. For vesicle quantification, the Membrain segmentation was imported into Dragonfly, and each vesicle was manually segmented as a separate feature. Highly elliptical vesicles were excluded from further analysis and elongated Membrain segmentation in the Z-direction were not annotated due to the missing wedge. All features/vesicles were exported as multi-ROI grey-value TIFF images with each integer value corresponding to one features/vesicles. Vesicle diameter and category (single, outer-filled or inner) were quantified using a custom python script. Briefly, 1) for each vesicle a sphere was fitted to binarized density of the vesicle, 2) the radius and centroid of the fitted sphere was saved, 3) and the category was computed by comparing the centroid and radius of each vesicle to all other vesicles of the same tomogram. If the centroid of a smaller radius vesicle fell within the centroid and radius of another vesicles it was assigned as an inner vesicle, or vice-versa as an outer vesicle. If neither condition is met it was assigned as single vesicle. All fits and assignment were manually inspected using a ChimeraX marker (.cmm) file containing the centroid and radius for each vesicle. Data was plotted using seaborn[121].

Two-dimensional class averages

Two-dimensional class averages of filaments and membranes were generated by converting binarized filament, membrane or vesicles segmentations generated above, into coordinates with a 10-voxels radial cutoff. Coordinates were then imported into RELION5.0 and classical 3D-subtomograms were extracted from CTF-corrected 4x down-sampled tomograms as implemented in RELION5.0. The central three slices were then Z-projected using a custom python script and used for 2D classification implemented in RELION5.0.