Diversity, origin, and evolution of the ESCRT systems

Kira S Makarova, Victor Tobiasson, Yuri I Wolf, Zhongyi Lu, Yang Liu, Siyu Zhang, Mart Krupovic, Meng Li, Eugene V Koonin

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Check for updates

| Evolution | Research Article

Diversity, origin, and evolution of the ESCRT systems

Kira S. Makarova,[1] Victor Tobiasson,[1] Yuri I. Wolf,[1] Zhongyi Lu,[2,3] Yang Liu,[2,3] Siyu Zhang,[2,3] Mart Krupovic,[4] Meng Li,[2,3] Eugene V. Koonin[1]

AUTHOR AFFILIATIONS See affiliation list on p. 18.

ABSTRACT Endosomal sorting complexes required for transport (ESCRT) play key roles in protein sorting between membrane-bounded compartments of eukaryotic cells. Homologs of many ESCRT components are identifiable in various groups of archaea, especially in Asgardarchaeota, the archaeal phylum that is currently considered to include the closest relatives of eukaryotes, but not in bacteria. We performed a comprehensive search for ESCRT protein homologs in archaea and reconstructed ESCRT evolution using the phylogenetic tree of Vps4 ATPase (ESCRT IV) as a scaffold and using sensitive protein sequence analysis and comparison of structural models to identify previously unknown ESCRT proteins. Several distinct groups of ESCRT systems in archaea outside of Asgard were identified, including proteins structurally similar to ESCRT-I and ESCRT-II, and several other domains involved in protein sorting in eukaryotes, suggesting an early origin of these components. Additionally, distant homologs of CdvA proteins were identified in Thermoproteales which are likely components of the uncharacterized cell division system in these archaea. We propose an evolutionary scenario for the origin of eukaryotic and Asgard ESCRT complexes from ancestral building blocks, namely, the Vps4 ATPase, ESCRT-III components, wH (winged helix-turn-helix fold) and possibly also coiled-coil, and Vps28-like domains. The last archaeal common ancestor likely encompassed a complex ESCRT system that was involved in protein sorting. Subsequent evolution involved either simplification, as in the TACK superphylum, where ESCRT was co-opted for cell division, or complexification as in Asgardarchaeota. In Asgardarchaeota, the connection between ESCRT and the ubiquitin system that was previously considered a eukaryotic signature was already established.

IMPORTANCE All eukaryotic cells possess complex intracellular membrane organization. Endosomal sorting complexes required for transport (ESCRT) play a central role in membrane remodeling which is essential for cellular functionality in eukaryotes. Recently, it has been shown that Asgard archaea, the archaeal phylum that includes the closest known relatives of eukaryotes, encode homologs of many components of the ESCRT systems. We employed protein sequence and structure comparisons to reconstruct the evolution of ESCRT systems in archaea and identified several previously unknown homologs of ESCRT subunits, some of which can be predicted to participate in cell division. The results of this reconstruction indicate that the last archaeal common ancestor already encoded a complex ESCRT system that was involved in protein sorting. In Asgard archaea, ESCRT systems evolved toward greater complexity, and in particular, the connection between ESCRT and the ubiquitin system that was previously considered a eukaryotic signature was established.

KEYWORDS ESCRT, membrane remodeling, Asgard archaea, cell division

Membranes are a fundamental staple of life. All cells are membrane bounded, and membrane compartments, most likely, predated the emergence of the translation apparatus and even genetic information itself (1, 2). Genomes likely evolved

Editor Igor B. Jouline, The Ohio State University, Columbus, USA

Address correspondence to Kira S. Makarova, makarova@ncbi.nlm.nih.gov, Meng Li, limeng848@szu.edu.cn, or Eugene V. Koonin, koonin@ncbi.nlm.nih.gov.

The authors declare no conflict of interest.

See the funding table on p. 18.

Received 5 February 2024
Accepted 6 February 2024
Published 21 February 2024

within membrane-bounded protocells, and to ensure accurate segregation of replicating genomes into the daughter cells, a membrane-associated division apparatus had to evolve at the earliest stages of evolution. Accordingly, the molecular machinery involved in cell division could be expected to be conserved in all domains of life. However, this is far from being the case. Even if common structures, such as the division ring, are formed by most dividing cells, the proteins involved in this process differ (3–5). Even the most broadly conserved division machinery component, the FtsZ/tubulin GTPase involved in the division ring formation, is missing in several lineages of bacteria and archaea, and its function is taken over by other proteins (6–8). Most of the other protein components of the FtsZ-centered divisome, with a notable exception of SepF (9, 10), are not conserved but, rather, are lineage specific (3, 5, 11). Moreover, FtsZ paralogs can be repurposed to perform other functions related to membrane remodeling (12–14).

Extant living cells not only employ diverse proteins for division but also form vesicles and intracellular compartments that require specific membrane remodeling systems, which are often distinct from the cell division apparatus (15). Eukaryotic cells possess a highly complex cell division machinery and multiple endomembrane organelles, such as mitochondria, endoplasmic reticulum, Golgi complex, and others. The traffic of molecules between these compartments in eukaryotic cells is enabled by several sorting systems (16). One of such molecular machines is endosomal sorting complexes required for transport (ESCRT), which mediates a variety of membrane scission events, including a key role in cytokinesis in at least some plant and animal cells (17, 18). The composition, structure, and function of ESCRT complexes in eukaryotes have been studied in detail (17, 19–21). The core of this system consists of four interacting, multisubunit complexes that are conserved in most eukaryotes (22): ESCRT-I, ESCRT-II, ESCRT-III, ESCRT-IV, and the less deeply conserved ESCRT-0 (Fig. 1 and 2). The ESCRT complex is linked to other protein trafficking systems, such as adaptins and clathrins (Fig. 1).

For a long time after their discovery, protein components of ESCRT complexes were thought to be eukaryote specific (41). However, in 2008, Lindås et al. discovered a cell division (Cdv) machinery in Crenarchaeota that was distinct from the previously characterized FtsZ-based system present in most bacteria and archaea. CdvB and CdvC proteins were found to be homologous to eukaryotic ESCRT-III and IV (Vps4) subunits, respectively, whereas no homologs were identified for CdvA (8). Subsequently, ESCRT-III components and the Vps4 ATPase, another essential ESCRT component, were identified in a wider range of archaea (42). However, the distribution of these proteins is patchy, and they were found to be conserved only in Sulfolobales, Desulfurococcales, and Thaumarchaea (42). Nevertheless, these findings suggested deep archaeal roots for ESCRT-III and Vps4. More recently, structural and functional similarities were detected between polymers of ubiquitous bacterial membrane remodeling proteins Vipp1 and PspA and those of eukaryotic and archaeal ESCRT-III subunits, suggesting that ancestors of ESCRT-III were present in the last universal cellular ancestor (LUCA) (43). However, the origin of ESCRT-I and II remained enigmatic until the discovery of Asgardarchaeota (Asgard for short), the phylum that includes the closest known archaeal relatives of eukaryotes which, in most phylogenetic trees of conserved informational proteins, are associated with the Asgard class Heimdallarchaeia (39, 44, 45). Strikingly, homologs of the subunits of ESCRT-I, ESCRT-II, ESCRT-III, and ESCRT-IV complexes along with a complete ubiquitin machinery that plays a key role in protein fate determination in eukaryotes were identified in Asgard archaea (27, 39, 45, 46).

A closer examination of the components of the ESCRT complex revealed remarkable diversity of domain architectures in Asgard archaea (39). In view of this diversity and the relatively low sequence conservation among ESCRT components, we re-analyzed archaeal ESCRT systems, in particular, taking advantage of AlphaFold2 (47) (AF2) structure predictions for analysis beyond the sequence level. This analysis resulted in the identification of several distinct groups of ESCRT systems in archaea outside of Asgard, including proteins structurally similar to ESCRT-I and ESCRT-II components, and several other domains involved in protein sorting in eukaryotes, suggesting an early origin of

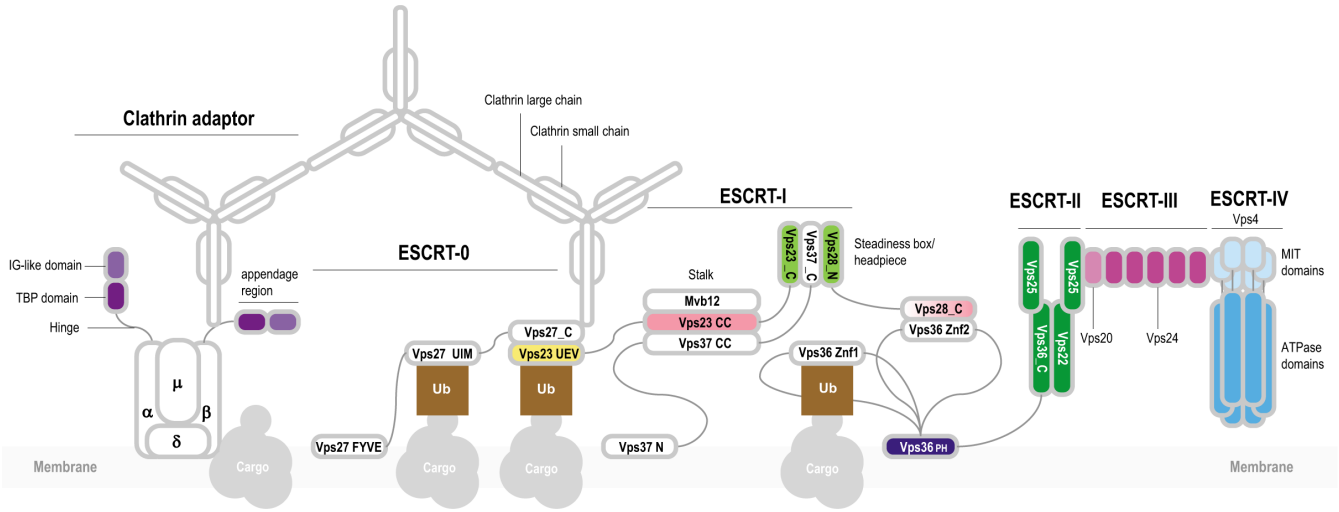

FIG 1 General organization of the eukaryotic ESCRT system and its connection with other protein sorting complexes. The ESCRT machinery was thoroughly characterized in yeast and mammals (17, 19–21). The core of the system consists of four conserved, multisubunit complexes: ESCRT-I, ESCRT-II, ESCRT-III, and ESCRT-IV. ESCRT-0 subunits are not generally conserved and apparently differ among the eukaryotic phyla. ESCRT-0 subunits interact with protein cargo and recruit ESCRT-I complex. The best characterized subunit of ESCRT-0 is Vps27 which is a hub of protein-protein interactions (19). Vps27/HRS is a multidomain protein containing a VHS (Vps27/HRS/STAM) super-helix domain, an FYVE zinc finger, several coiled-coil regions, and a C-terminal P[ST]xP motif. Vps27 interacts with the ubiquitin E2 variant (UEV) domain of Vps23 (ESCRT-I) via the P[ST]xP motif and also with clathrin, ubiquitinated cargo, and lipids (19, 23, 24). ESCRT-I consists of four subunits: Vps28, Vps23, Vps37, and Mvb12 (Fig. 2). Three subunits, Vps23, Vps37, and Mvb12, form a stalk through the interaction between three long helices forming a coiled-coil structure. Vps23, Vps28, and Vps37 form a headpiece through the interaction between homologous helical hairpins known as the steadiness box (SB) (25). The C-terminal helical bundle domain of Vps28 interacts with ESCRT-II Vps36 via a Zn finger (Znf2) inserted into the split PH domain. ESCRT-II also consists of four subunits: Vps22, Vps36, and two copies of Vps25 (26). These proteins are paralogs that share a common region consisting of a pair of wH domains. In addition to the wH domains, Vps36 is fused to a PH domain and two zinc finger domains (the latter are absent in mammalian orthologs). Vps22 and Vps36 contain an additional helical domain upstream of the proximal wH domain, which likely promotes their interaction (Fig. 2). The N-terminal domain of VPS25 binds the C-terminal domain of VPS22, and the other VPS25 subunit contacts both VPS36 and VPS22, forming an asymmetric Y-shaped structure (26). Vps25 recruits the initial Vps20 subunit of ESCRT-III complex (26). ESCRT-III consists of several paralogous subunits, which belong to two groups, Vps2-like [MIM-1 (MIT-interacting motif) containing] and Vps20-like (MIM-2 containing), respectively (16, 27) (Fig. 2). Vps2 is targeted to the membrane where it nucleates Vps20 polymerization (16). Depolymerization of ESCRT-III is regulated by the Vps4 (16). Vps4 contains a diagnostic N-terminal three-helix bundle, the microtubule interacting and trafficking (MIT) domain which interacts with the C-terminal MIM-2 in ESCRT-III subunits (28). The ESCRT machinery is required for endosomal trafficking and is connected with other complexes involved in this process including clathrin, a large complex consisting of light and heavy subunits forming a triskelion (29). In eukaryotes, clathrin forms a cage-like scaffold around a vesicle and recruits Vps27. Clathrin is associated with clathrin adaptor complexes (AP) which mediate clathrin-dependent protein trafficking to and from endosomes (30, 31). AP complexes differ minimally in subunit composition. AP2, for example, consists of four subunits: alpha, beta, gamma, and mu (30, 31). Alpha and beta subunits are paralogs that contain an N-terminal HEAT repeats domain and a C-terminal appendage domain connected through an unstructured hinge interacting with clathrin. The appendage domain consists of two distinct subdomains, the proximal immunoglobulin-like (IG) beta sandwich fold and the distal TBP (TATA-box binding protein or helix-grip) fold (32). Appendage domains of alpha and beta subunits can recruit multiple additional proteins, whereas the N-terminal domains of alpha and beta subunits and the mu subunit interact with the membrane protein cargo (33, 34).

these components. Additionally, we identified distant homologs of CdvA in Thermoproteales which are likely to be components of the elusive cell division system in this group of archaea. We propose an evolutionary scenario for the origin of eukaryotic and Asgard ESCRT complexes from ancestral building blocks, namely, the Vps4 ATPase, ESCRT-III components, wH (winged helix-turn-helix fold) and possibly also coiled-coil, and Vps28-like subdomains.

RESULTS AND DISCUSSION

Phylogeny of Vps4 ATPase and distribution of ESCRT systems across archaea

Vps4 family ATPase is present in all ESCRT systems and is conserved at the sequence level, making it suitable for phylogenetic reconstructions and for analyzing ESCRT evolution. We identified Vps4 (to the exclusion of other AAA+ ATPases) in archaea and

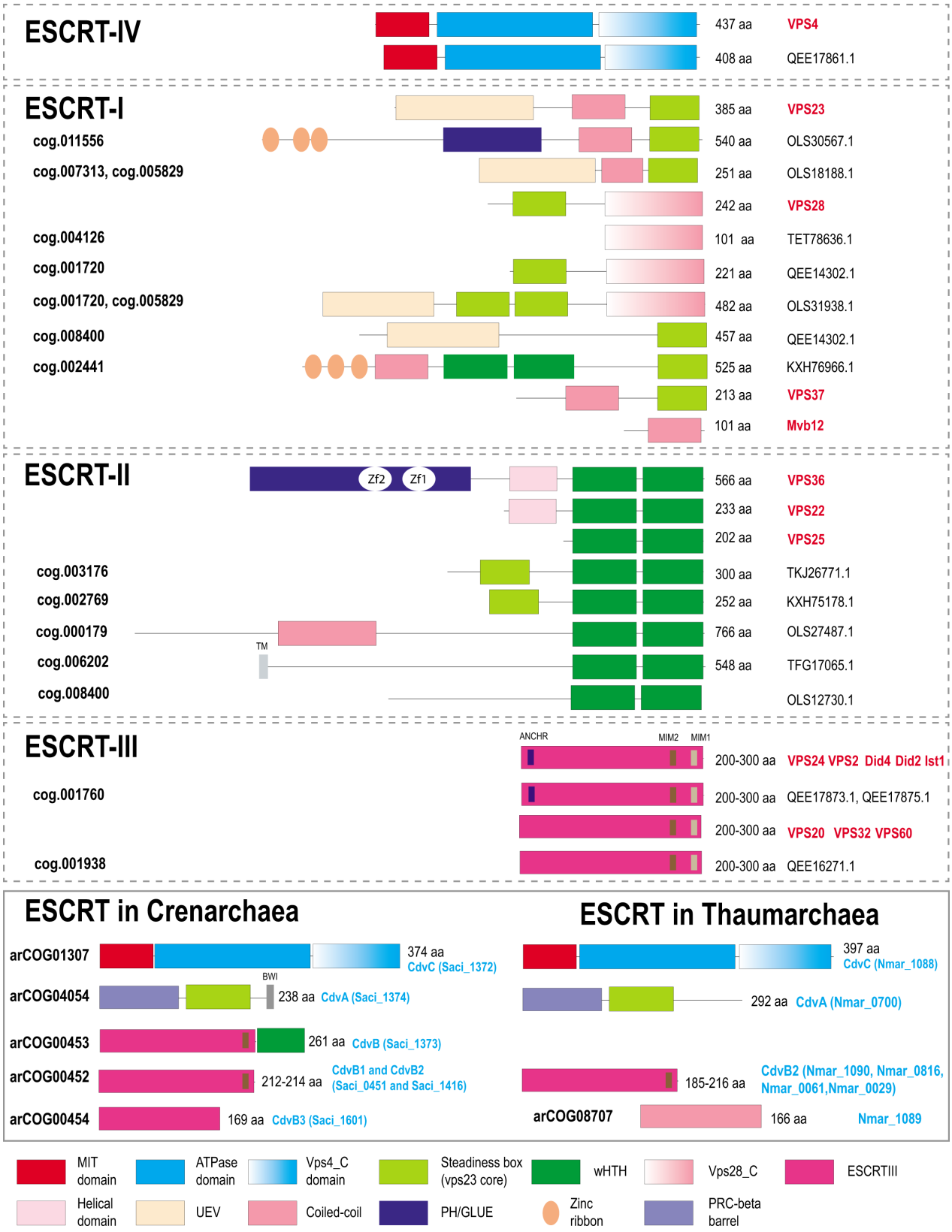

FIG 2 Domain organization and key motifs of ESCRT protein in eukaryotes, Asgard, and TACK archaea. Protein domains are shown as colored shapes according to the color code given beneath the schemes. Proteins are drawn roughly proportional to their size in amino acids which is indicated in the right. Domain boundaries for *Saccharomyces cerevisiae* are based on previously published structures and analyses (19, 24, 35). For archaeal proteins, domain boundaries are

(Continued on next page)

FIG 2 (Continued)

based either on previous analyses (22, 36–38) or on sequence analysis performed during this work as described under Materials and Methods. *S. cerevisiae* gene names are highlighted in red on the right. For Asgard and TACK archaea, protein size and accession are indicated on the right. asCOG numbers (39) are indicated for the Asgard proteins, and arCOG numbers (40) are indicated for the TACK proteins on the left. Signature sequence motifs are shown by small rectangles, and short names for the motifs are provided above the protein schematics (once for proteins with the same domain organization). Abbreviations: Zf, zinc finger; TM, transmembrane segment; UEV, ubiquitin E2 variant; MIM, MIT interaction motif; BWI, broken winged-helix interaction site. Not all proteins from cog.003176 are fused to SB, but most contain additional N-terminal subdomains distinct from those in eukaryotic homologs.

eukaryotes based on two criteria: (i) presence of the microtubule interacting and trafficking (MIT) domain, known to be essential for interaction with ESCRT-III, or (ii) presence of genes encoding ESCRT components in the respective archaeal genome neighborhoods. Using these criteria, we compiled a representative Vps4 set that included 97 proteins from Asgard archaea, 152 proteins from other archaea (arCOG01307), and 71 proteins from selected eukaryotes (Table S1; File S1). A phylogenetic tree of Vps4 was built from the multiple alignment of these sequences (Fig. 3A). The tree was rooted using the related but distinct Cdc48 ATPases as an outgroup (see Materials and Methods and Fig. S1).

Eukaryotic and Asgard Vps4s were found to be monophyletic except for three diverged paralogs from Hodarchaeota that group outside of the main Asgard clade (Fig. 3A; File S1). Considering that the same hodarchaeotal genomes encoded Vps4 variants that group within the main Asgard clade and that these three outliers are not associated with any other ESCRT system genes, they likely represent a fast-evolving, subfunctionalized subfamily. The eukaryotic-Asgard clade is a sister group to a large clade corresponding to Vps4s from the TACK [Thaumarchaeota (now Nitrososphaeria), Aigarchaeota, Crenarchaeota (both Thermoproteota), and Korarchaeota (now Korarchaeia)] superphylum. In these organisms, ESCRT systems have been shown to participate in cell division (4, 8, 48) and extracellular vesicle biogenesis (49). We denoted this clade CdvA, after the unique protein diagnostic of these systems. Most of the remaining, largely euryarchaeal, Vps4s form five clades. The largest and most taxonomically diverse clade was denoted "Main." As noted previously (42), several Thermococcales species encode a solo Vps4, but not other components of the system (Fig. S1; Table S2). These standalone Vps4 proteins belong to a separate clade on a long branch suggestive of subfunctionalization.

The tree included also a Thermoplasmata-specific clade and three clades that we denoted by their distinct features and the respective taxa: Halo_FHA [Halobacteriales, FHA (forkhead-associated) domain containing], Meth_adaptin (Methanobacteriales, adaptin appendage domain containing), and Halo_adaptin [Halobacteriales, adaptin-like IG (immunoglobulin) domain containing] (Fig. 3A).

The main clade included representatives from Thermococcales, Methanococcales, Methanomicrobiales, and Halobacteriales orders and several other archaea (File S1). Besides Vps4 and ESCRT-III, all systems in this clade encode two proteins that belong to arCOGs (40) 09748 and 08177 (Table S2). arCOG09748 proteins contain a C-terminal wH domain with sequence and structural similarity (Z-score = 6.4) to the wH domain at the C-terminus of CdvB (36) and structural similarity (Z-score = 6.7) to the C-terminal domain of Vps22 (ESCRT-II complex) (Fig. 4A, Fig. S2; Table S3). AF2 modeling for arCOG09748 protein from *Palaeococcus pacificus* revealed an N-terminal helical domain, a coiled-coil domain, a globular central domain (no structural similarity to any known domains), and a wH C-terminal domain (Fig. S2). The N-terminal domain of arCOG08177 proteins showed surprising similarity to an apparently inactivated Cas2 nuclease, the CRISPR-Cas system component that forms a complex with the Cas1 integrase and is involved in spacer acquisition (50) (Fig. S2; Table S3). Like Cas2 in the CRISPR adaptation complex, the N-terminal domain of arCOG08177 proteins could mediate dimerization of arCOG08177 proteins (50). The second domain of arCOG08177 corresponds to the archaea-specific domain DUF3568 and adopts the TBP (TATA binding protein or helix-grip) fold shared with the C-terminal domain of the adaptin appendage region (Fig. 3B and 4) (51). In many genomes from the main clade, the ESCRT locus additionally contained genes

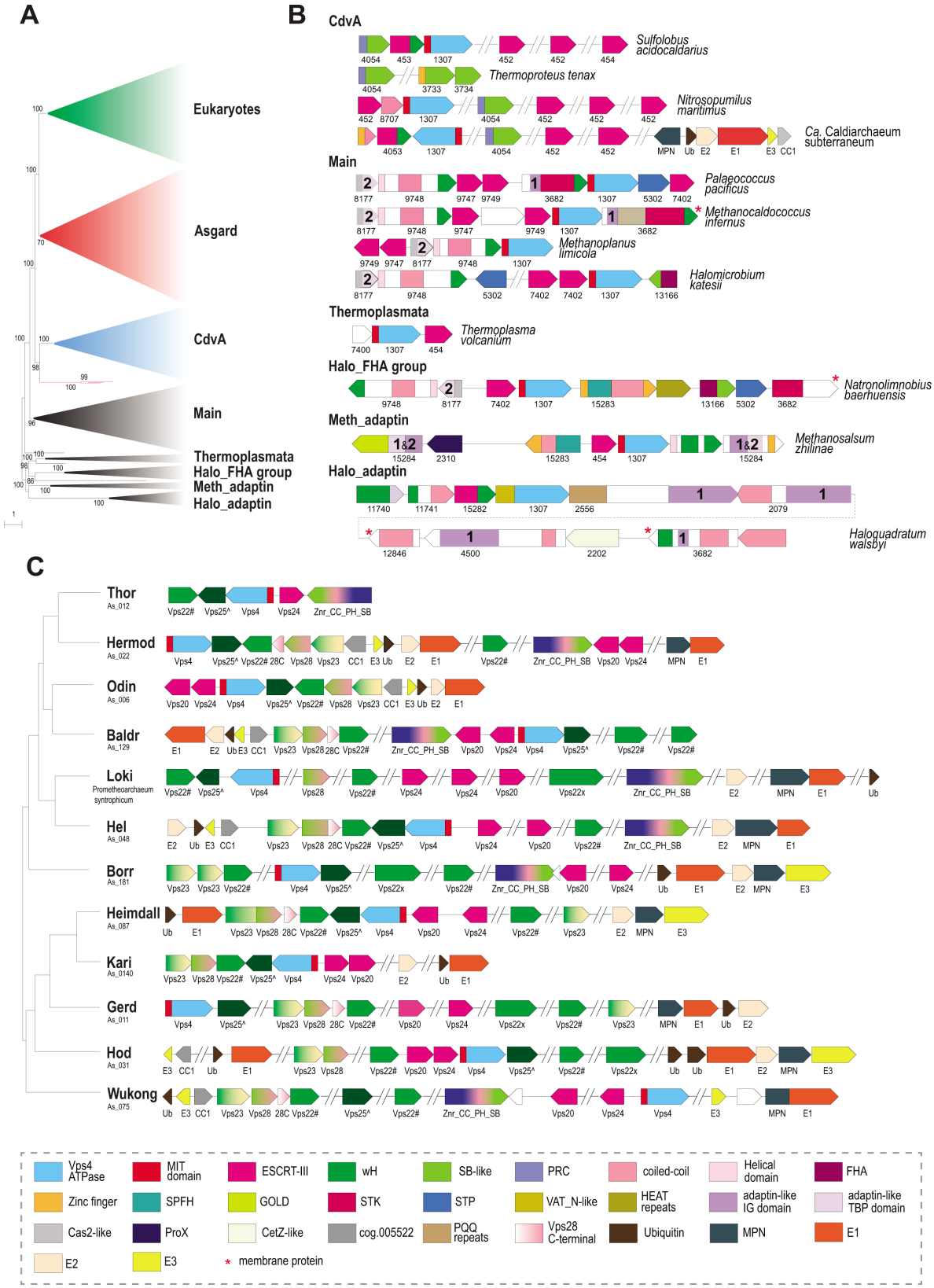

FIG 3 Phylogeny of Vps4 family ATPase and genomic neighborhoods of ESCRT systems in archaea. (A) Schematic representation of the phylogenetic tree of Vps4 protein family. The phylogenetic tree was built using IQ-Tree as described in Material and Methods. Major branches were collapsed and are designated on

(Continued on next page)

FIG 3 (Continued)

the right. Bootstrap values >70% calculated by IQ-Tree are shown. The complete tree in Newick format is available in the File S1. (B) Organization of selected ESCRT loci for each major branch of archaea (except Asgard). Genes are shown as arrows roughly proportional to gene size. Domains are indicated within the arrows for multidomain proteins, not to scale. Homologous genes and domains (or generic structural elements such as coiled-coil or helical domains) are color coded according to the code at the bottom of the figure. ArCOG numbers are indicated below the arrows. Numbers inside the arrows indicate IG domain (1) and TBP domain (2) in the adaptin appendage homology regions. (C) Components of ESCRT systems in selected genomes of 12 major Asgard lineages. The ESCRT related genes are shown for selected genomes of the 12 major lineages of Asgardarchaeota (39). The tree schematically shows the relationships among the lineages according to previously published phylogenetic analysis (39). Designations are the same as in B, except for multidomain proteins in which domains are not shown, but are explained in the color code schematics in the bottom of the figure. Protein names are indicated below the arrows. Designations: Vps22#, Asgard specific version of Vps22, cog.002769; Vps25^, Asgard specific version of Vps25, cog.002441; Znr_CC_PH_SB, Asgard specific protein family of cog.011556 (see Fig. 2). Abbreviations: wH, winged helix domain; Znr, zinc ribbon; SB, steadiness box; FHA, forkhead-associated domain; STK, serine/threonine protein kinase; STP, serine/threonine protein phosphatase; UEV, ubiquitin E2 variant; Ub, ubiquitin; CC, coiled-coil; PH, plextrin homology; SB, steadiness box; MIT, microtubule interacting and transport; HEPN, higher eukaryotes and prokaryotes nucleotide-binding domain; GOLD, Golgi dynamics domain; SPFH, stomatin, prohibitin, flotillin, and HflK; VAT_N, N-terminal domain of valosin-containing protein-like ATPase of *Thermoplasma acidophilum*; TBP, TATA-binding protein. Abbreviations for repetitive domains: HEAT, Huntington, Elongation Factor 3, PR65/A, TOR; PQQ, pyrrolo-quinoline quinone; PEGA, containing PEGA sequence motif. Abbreviations for ubiquitination pathway: E1, ubiquitin-activating enzyme; E2, ubiquitin-conjugating enzyme; E3, ubiquitin ligases; MPN, Mpr1/Pad1 N-terminal domain, deubiquitinating enzyme.

encoding serine/threonine protein kinase (STK) and protein phosphatase (STP) (Fig. 3B; Table S2). Typically, these STKs are membrane associated. Most STKs consist of several domains: a transmembrane segment at the N-terminus, immunoglobulin-like repeats (found in the appendage region of adaptins) domain, and a C-terminal small helical domain for which HHpred search revealed similarity with a distinct variant of the wH fold, PCI (for proteasome, COP9, initiation factor 3) domain (Fig. 4B, Fig. S2; Table S3). The PCI domain is responsible for oligomerization of these proteins in respective complexes (52, 53). Thus, this domain in STK could be involved in either oligomerization or interaction with wH domain of arCOG09748 proteins.

The Thermoplasmata clade included minimal ESCRT systems that consist of genes encoding Vps4, ESCRT-III, and a beta-barrel protein from arCOG07400 (Fig. 3B; Tables S1 and S2). Several Thermoplasmata species lack FtsZ but encode complete or partial ESCRT systems that are likely involved in cell division (Fig. 5).

A distinct FHA domain family (arCOG13166) is associated with Vps4 and ESCRT-III in several haloarchaeal genomes comprising the signature of the Halo_ FHA clade (Fig. 3B; Fig. S1; Table S2). The FHA domain is involved in phosphorylation-dependent protein oligomerization and/or interaction with phosphorylated proteins (55, 56). The C-terminal domain of arCOG13166 proteins showed limited similarity to CdvA (Fig. 3B; Fig. S4 and S5; Table S3). Proteins of arCOGs 09748 and 08177, STKs, and STPs are also encoded in some of these Halo_FHA loci (Fig. 3B; Table S2).

Meth_adaptin neighborhoods encode two proteins similar to the adaptin appendage region (both IG-like and TBP domains) (Fig. 3B and 4B; Fig. S6; Tables S2 and S3). One of the adaptin appendage region containing proteins (e.g., WP_013898571.1) also contains a Zn-ribbon domain and an additional uncharacterized N-terminal domain (Fig. S6). In one of these proteins (e.g., WP_013898578.1), the appendage region is additionally fused to a GOLD (Golgi dynamics) domain (57) via a transmembrane linker (Fig. S6). GOLD domains connect membrane cargo proteins with COPI (an adaptin homolog) vesicles (30, 58). Another component of the Meth_adaptin systems belongs to the SPFH (stomatin, prohibitin, flotillin, and HflK/C) family (Fig. 3B). In eukaryotes, SPFH proteins interact with the ESCRT-0 component Vps27 which presents ubiquitinated proteins to the ESCRT machinery (59). SPFH-family protein additionally contains coiled-coil and zinc ribbon domains (Fig. S6). Interestingly, bacterial SPFH family protein YdjI is encoded in the some PspA loci (ESCRT III family) and is required to localize PspA to the membrane (60), suggesting that the functional connection between ESCRT III and SPFH proteins could have been established very early in evolution. Recently, a structure of a distant bacterial SPTH homolog has been solved, and it has been shown that it forms 24-mer with coiled-coil domains interacting and forming a "lid" of an enclosed compartment on a

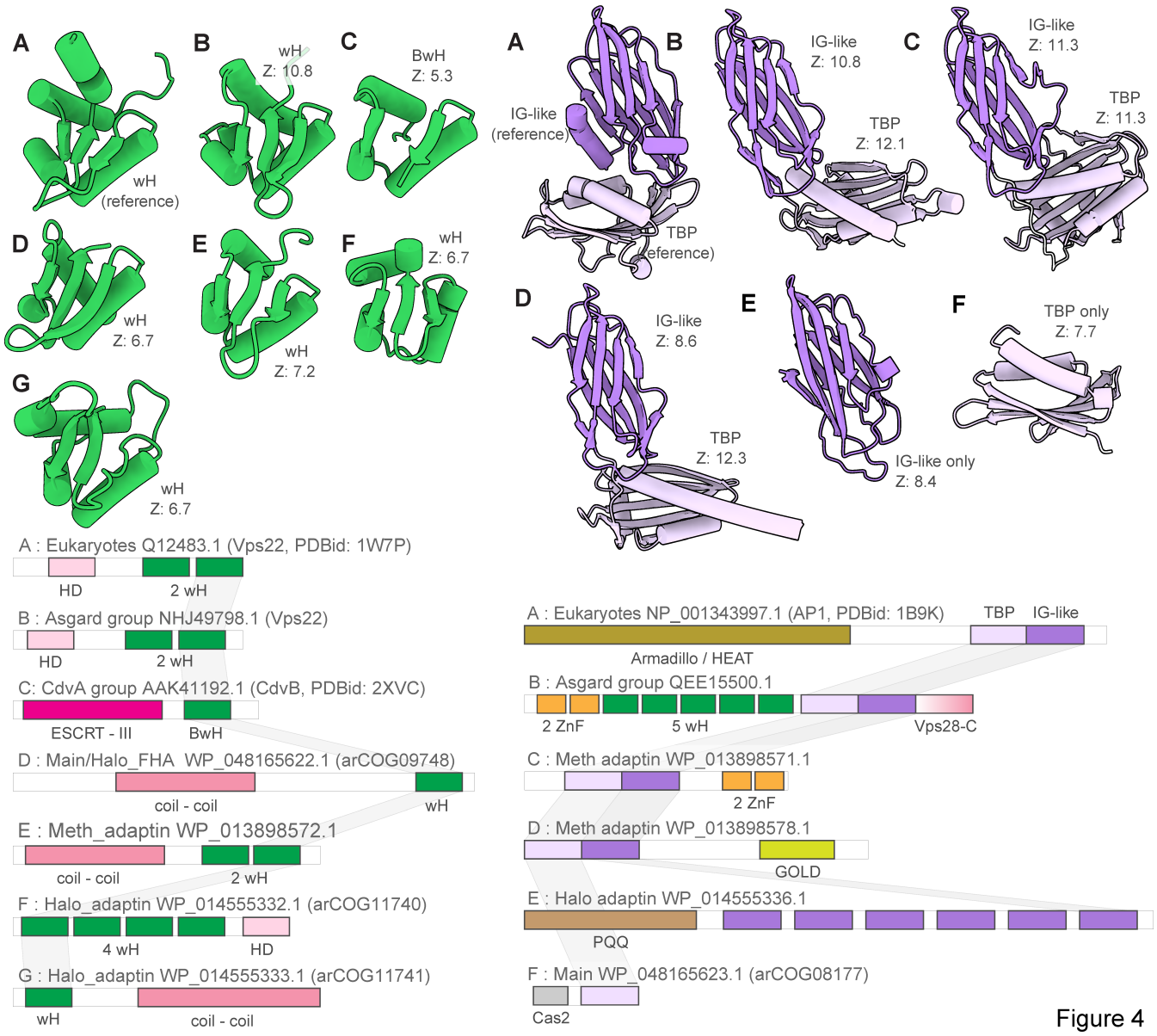

Figure 4

FIG 4 Winged helix and adaptin appendage domain homologs in eukaryotic and archaeal ESCRT systems. Left, comparison of winged helix domains (wH). Right, adaptin appendage-like regions. Multidomain proteins were trimmed to show only domains homologous to wH and adaptin appendage-like domain of Vps22 and alpha-adaptin appendage domain. Complete proteins are shown in Fig. S2, S7, and S8. Complete domain organizations of the respective proteins are shown underneath the structures. Letters denote the correspondence between structures and domain architectures. DALI Z-scores between the indicated domains and wH domain of Vps22 or IG-like and TBP domain of alpha-adaptin appendage domain are indicated. Abbreviations: IG, immunoglobulin; wH, winged helix domain; BwH, broken winged helix domain; TBP, TATA binding protein.

membrane (61). This protein might play a similar role in ESCRT systems. Finally, Meth_adaptin loci included a gene encoding ProX, a substrate-binding protein of the osmoregulatory ABC-type glycine betaine/choline/L-proline transport system (62) (Fig. 3B; Table S2). The ProX-like protein likely binds a substrate which is sorted into vesicles formed by other genes of this ESCRT system.

Despite grouping with the Meth_adaptin clade in the tree, the architecture of Halo_adaptin loci is quite different (Fig. 3B). In these systems, the Vps4-like ATPase (e.g., WP_014555335.1) lacks an identifiable MIT domain and instead contains an N-terminal domain related to the N-terminal domain of the Cdc48-like ATPase VAT (valosin-containing protein of *Thermoplasma acidophilum*) (Table S3). Upstream of the ATPase, there is a

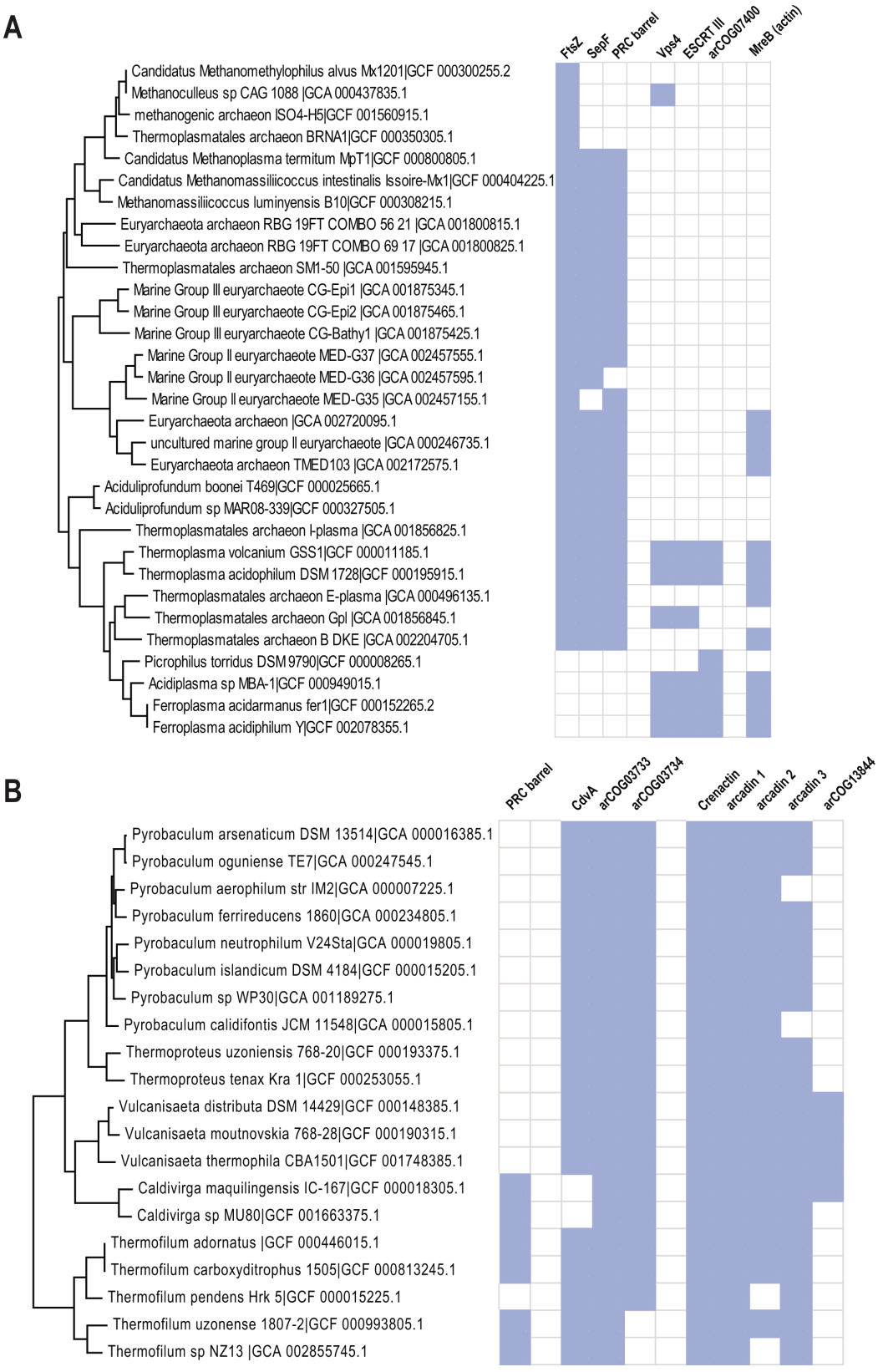

FIG 5 Phyletic patterns for putative cell division components of Thermoproteales and Thermoplasmata. The presence/absence patterns of cell division components for (A) Thermoplasmata and (B) Thermoproteales are shown. The patterns are overlayed on the respective species subtree. The tree built for all 524 arCOGs genomes using IQ-tree program (54) and concatenated alignment of ribosomal proteins. Blue indicates presence, and white indicates absence.

gene encoding a protein with weak sequence similarity to PspA, an ESCRT-III superfamily protein (Table S4; Fig. S7). Given that in most other archaeal ESCRT systems, at least one ESCRT-III gene is located upstream of Vps4, this protein is likely to be a derived version of ESCRT-III (Fig. 3B; Fig. S7). These neighborhoods also encompass genes encoding two wH containing proteins: arCOG11740 and arCOG11741. As in many other systems described above, a wH protein (arCOG11741) is encoded upstream of the ESCRT III-like protein suggesting that these proteins interact (Fig. 3B; Table S2). The wH domain in arCOG11741 proteins is fused to a coiled-coil domain. The wH containing arCOG11740 proteins are encoded by the proximal gene in the loci (Table S2). Like eukaryotic Vps36, arCOG11741 proteins contain duplicated wH domains and an N-terminal alpha-helical domain (Fig. 3B and 4A; Fig. S7 and S3). Similarly to the Meth_adaptin systems, these neighborhoods encode two distinct proteins with sequence and structural similarity to the IG-like domain of the adaptin appendage region (Fig. S7; Table S3). Both "adaptin-like" proteins in Halo_adaptin system contain several IG-like domains (Fig. S7). The N-terminal domains of both "adaptin-like" proteins adopt a WD40/PQQ-like beta propeller fold (Fig. S7). Notably, a WD40-like domain is present in two homologous COPI alpha and beta-prime subunits, which mediate interaction with specific protein cargo (63) and in the heavy subunit of clathrin (31). Halo_adaptin systems are the most complex and additionally include two more IG-like domains containing proteins, a membrane protein and a protein related to CetZ (FtsZ/tubulin family) that is involved in cell shape control in *Haloferax volcanii* (14) (Fig. 3B; Table S2). Phylogenetic analysis showed that the CetZ-like protein from *Haloquadratum walsbyi* groups outside the main CetZ branch suggesting a different function (14). Here, we link this function to the Halo_adaptin system likely involved in protein sorting.

Most members of the TACK superphylum possess a minimal ESCRT system consisting of CdvA, CdvB (ESCRT-III), and CdvC (Vps4) (Fig. 3B; Table S1). Recently, CdvA orthologs have been identified in almost all Thermoproteales (64) where they were previously thought to be missing. In addition to CdvA orthologs found in most Thermoproteales, we report here two more CdvA-related protein families specific for this lineage, arCOG03733 and arCOG03734 (Fig. 5B; Fig. S4; Table S3). These two paralogs are encoded next to each other in most of the Thermoproteales genomes (Table S2). Proteins of arCOG03733 additionally contain a Zn ribbon N-terminal domain connected to CdvA-like domain through a long coil-coiled like region, whereas proteins of arCOG03734 are closely similar but lack the Zn ribbon. The domain architectures of these proteins, along with CdvA, are similar to that of eukaryotic Vps23 and Vps37 components of ESCRT-I (Fig. 2). Furthermore, HHpred search revealed sequence similarity between the helical hairpin of the steadiness box (SB) and CdvA (Table S3; Fig. S4), supporting our previous observations (65). Recent examination of CdvB protein sequences suggested that most of them harbor a C-terminal MIT interaction motif 2 that is conserved in Vps20/32/60 subunit class (37) (Fig. 2). Most of the TACK genomes encode additional ESCRT III family proteins that, however, lack the wH extension as noticed previously (38). Furthermore, in Thaumarchaea, none of the CdvB family proteins contain fused wH domains (38). Another distinctive feature of this system in Thaumarchaea is the additional component (arCOG08707), a coiled-coil protein, encoded upstream of the *cdvC/vps4* gene (Fig. 3B). arCOG08707 proteins are not similar to any known proteins, but structural models show that they possess a helical hairpin which is present in both ESCRT III and CdvA (SB) (Fig. S4). Bathyarchaea encode additional CdvA-like proteins (e.g., KPV61682.1) which are fused to an unstructured glutamate-rich sequence and are encoded outside of the *cdv* locus.

Extensive diversification of ESCRT in Asgard archaea

The diversity of ESCRT systems in Asgardarchaeota far exceeds that in any other archaea. Most of the Asgard lineages encode multiple eukaryotic-like ESCRT components as well as proteins involved in ubiquitination, suggesting interaction between the two systems, both of which are prominent in eukaryotes (22, 39, 45). In agreement with previous observations (22, 38, 39, 43, 45), we identified orthologs of the ESCRT III Vps24 and

Vps20 families (Fig. 2 and 6A). Compared to eukaryotes (19), Asgard archaea have fewer paralogs in both families.

The ESCRT-II proteins show considerable diversity in both domain organization and number of paralogs (Fig. 2 and 5A). Unlike in eukaryotes, Vps22 family proteins (Asgard cog.002769) are fused to an SB, whereas Vps25 proteins (cog.003176) contain additional N-terminal subdomains, some of which are also similar to SB (Fig. 2; Table S3).

Another frequent domain arrangement unique for Asgards (cog.002441) also includes two wH domains along with an SB and three Zn ribbons (Fig. 2). Among the ESCRT-I components, Asgard archaea encode orthologs of both Vps23 and Vps28 (Fig. 2 and 3C). In some Asgard genomes, the C-terminal domain of Vps28 is encoded separately, in addition to a full *vps28* gene (Fig. 3C). Furthermore, we identified a protein with a distinct domain organization, consisting of three Zn ribbon domains, a PH domain, a coiled-coil region, and an SB box (Znr_PH_CC_SB), which is absent in eukaryotes but is found in several Asgard lineages and is likely to be ancestral in Asgardarchaeota (Fig. 2 and 3C). Given the presence of SB, the distinct structural module of eukaryotic ESCRT-I, Znr_PH_CC_SB is a likely ESCRT-I component in Asgard archaea. In yeast, the PH domain of Vps36 interacts with phosphatidylinositol-3-phosphate, whereas one of the two Zn finger domains binds to ubiquitinated cargo, and the other one interacts with ESCRT-I via the C-terminal domain of Vps28 (Fig. 1) (35). By analogy, the PH domain of Znr_PH_CC_SB is expected to be membrane associated.

Typically, *vps25*-like genes are codirected (and likely co-translated) with *vps4* and ESCRT III, whereas *vps22*-like genes are co-directed with *vps23*-like and *vps28*-like genes as noted previously (22) (Fig. 3C). This arrangement most likely reflects the order of assembly of these components. Unlike in eukaryotes, Asgard Vps28-like SB domain and Vps28_C terminal domain are encoded by two separate genes (Fig. 3C). In eukaryotes, the Vps28 C-terminal domain interacts with ESCRT II complex subunits (66, 67), suggesting that a similar interaction takes place in Asgard archaea.

Lokiarchaeia and Helarchaeia genomes harbor many different families and different domain organizations of ESCRT system proteins, whereas Hermodarchaeia, Hodarchaeia, and Wukongarchaeia have none (Fig. 6A). Such diversity of ESCRT systems implies their functional specialization in different groups of Asgard archaea.

In eukaryotes, ESCRT-0 components are highly diverse and differ between fungi and mammals, with only a few genes shared (19). Orthologs of the most highly conserved Vps27 or Bro1/Alix (apoptosis-linked gene-2 interacting protein X) ESCRT-0 proteins were not identifiable in Asgard archaea. Conversely, many multidomain proteins containing domains related to ESCRT core components were detected in Asgard, but not in eukaryotes (39). Most of these proteins contain either the Vps28 C-terminal domain (Vps28-C) or the SB (Fig. 6B). The Vps28 C-terminal domain is often present along with the adaptin appendage region and a longin domain. Assuming that Vps28-C interacts with ESCRT-II, these proteins could associate with ESCRT-II without the involvement of any core ESCRT-I subunits. The presence of longin suggests that these proteins can activate small GTPases, which regulate many cellular pathways related to nutrient signaling and vesicular trafficking (68). Additional domains found in these proteins include TPR repeats, IG-like domains, and HTH/PCI domains which can be involved in interactions with protein substrates for further trafficking (69–71). The SB-containing proteins might interact with SB motifs in ESCRT-I, whereas other domains are likely involved in substrate recognition (Fig. 6B). Notably, the distribution of all these families of ESCRT components across Asgard genomes is extremely patchy such that many domain combinations are present in only a few species and rarely in multiple lineages (Fig. 6A). Nevertheless, most of the Asgard lineages encompass at least one representative of each of the major ESCRT components, that is, ESCRT-I, ESCRT-II, ESCRT-III, and ESCRT-IV.

Evolution of ESCRT systems

Despite the patchy distribution of the ESCRT systems in archaea, the topology of the Vps4 tree is largely consistent with the species tree (72) (Fig. 3A and 7; Fig. S1; File S1),

FIG 6 Phyletic patterns of ESCRT components and unique multidomain proteins in Asgardarchaeota. (A) Number of proteins in ESCRT asCOGs for 78 Agard phylum genomes. Major Asgard lineages are indicated above. asCOG numbers and a short description or gene or protein name are indicated on the left; the data were retrieved from asCOGs (41) with minimal corrections. The numbers of proteins are color coded as shown in the bottom right. Potentially ancestral (most widespread) asCOG are highlighted in green. (B) Domain architectures of Asgard proteins containing one or more domain related to ESCRT system. asCOG numbers are indicated on the left. Protein sizes and accession are indicated on the right. Homologous domains are shown by the same shape and color according to the color code provided in the bottom. Abbreviations are the same as in Fig. 3. Additional abbreviations: BAR, BIN, amphiphysin and Rvs161 and Rvs167 (yeast proteins) domain; TPR, tetratricopeptide repeats.

suggesting that these systems are mostly vertically inherited, with multiple lineages-specific losses occurring during evolution (Fig. S1; File S1; Table S1). We used the

topology of the Vps4 tree as the scaffold for reconstructing an evolutionary scenario for the ESCRT systems based on the parsimony principle (minimal number of events).

ESCRT systems (containing both *vps4* and ESCRT-III genes) are confined to archaea and eukaryotes and likely evolved in early archaeal ancestors, antedating the last archaeal common ancestor (LACA). As noted above, PspA, the apparent ancestor of ESCRT-III components, can be traced to the LUCA and initially was probably involved in membrane repair (43, 73). Recent analysis has shown that PspA proteins are often encoded next to wH proteins in both bacteria and archaea (73), suggesting a functional link between these two components of future ESCRT systems. In these early archaeal ancestors, Vps4 likely originated from the Cdc48 ATPase, as can be inferred from the presence of a substrate recognition domain (double psi beta-barrel fold), a possible relic of the ancestral state in the Halo_adaptin Vps4-like ATPase, and the overall similarity between Vps4 and Cdc48 (74). One of the two ATPase domains of Cdc48 was lost during this transition. Thus, an early ancestral ESCRT system could consist of Vps4, ESCRT-III, and a wH containing protein, the ancestor of ESCRT-II (Fig. 6). The wH domain is a key building block found in most ESCRT systems, which likely interacts with ESCRT-III (Fig. 2 and 4A). Indeed, in many TACK archaea, the wH domain is fused to the C-terminus of ESCRT-III. Subsequently, other components were acquired, resulting in the specialization for different cargo proteins and increasingly sophisticated regulation of the system assembly. One of these components is a coiled-coil domain, which is found in most ESCRT systems and could form a structure analogous to the stalk of ESCRT-I (Fig. 1 and 3B). Other ancestral modules could be alpha-helical domains similar to Vps28-C (helical bundle) and the SB box (helical hairpin). These simple structural elements often comprise N- or C-terminal domains in various proteins from archaeal ESCRT systems (Fig. 5B; Fig. S4 to S7). The exact origin of these elements is difficult to pinpoint, but they might have been derived from ESCRT-III or the MIT domain. Adaptin appendage structural domains were identified in many systems and thus could be functionally linked to early archaeal ESCRT (Fig. 3B, 4B, and 7).

Thus, the LACA ESCRT system could be similar in complexity to the Meth_adaptin system and probably coordinated protein sorting (Fig. 3B). Subsequent diversification of ESCRT yielded specialized systems for phosphorylated protein sorting, in the main clade and Halo_FHA systems, whereas in TACK and Thermoplasmata, ESCRT underwent reductive evolution (Fig. 6). In the TACK ancestor, ESCRT was co-opted for cell division and likely lost most of the components involved in protein sorting. The functions of the reduced ESCRT system in Thermoplasmata remain to be elucidated, but it seems likely to be involved in cell division as well (Fig. 5; Tables S1 and S2).

The ESCRT systems of Asgard archaea are highly complex, but this complexity is achieved mostly by duplications of the basic building blocks described above rather than by acquisition of new components (Fig. 7). First, a duplication produced two paralogous wH proteins, which further duplicated giving rise to a eukaryote-like, three-component ESCRT-II complex. Second, the SB box was duplicated and fused to VPS28_C domain or coiled-coil domain, giving rise to the ortholog of eukaryotic Vps28 and Vps23-like protein, respectively. Already at an early stage of Asgard evolution, ESCRT apparently became specialized for ubiquitinated protein sorting, primarily, as a result of the emergence of the UEV domain from the E2 component of the Ub system (75). The UEV domain eventually fused to a Vps23-like protein yielding the ortholog of eukaryotic Vps23. The most complex feature of the Asgard ESCRT-I is the diversity of multidomain proteins containing either an SB or the Vps28_C domain (Fig. 6B). Like in Meth_adaptin systems, many of these proteins also contain an adaptin appendage domain. Given that the adaptin appendage domain is a known hub of protein interactions (33, 34), the diversity of these components might provide additional flexibility of interactions with different types of cargo and can be considered functional analogs of ESCRT-0.

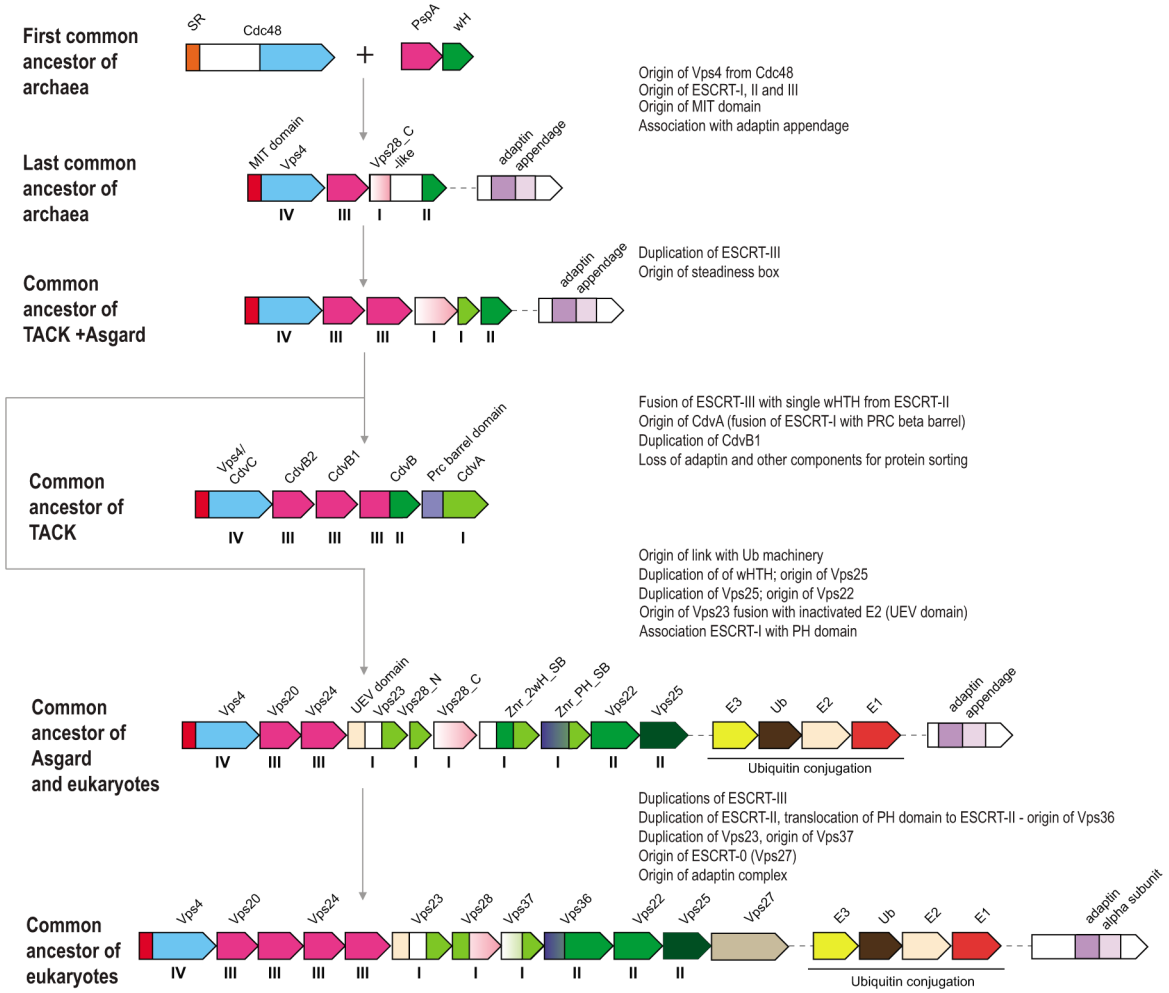

FIG 7 Origin and evolution of ESCRT systems. The inferred, hypothetical organizations of ESCRT systems in ancestral archaeal groups indicated on the left are shown. The inferred key events are described on the right. Designations of the genes and domains are the same as in Fig. 3. The Roman numerals under the genes correspond to (putative) components of the four ESCRT complexes. Abbreviations are the same as in Fig. 3, except for SR, substrate recognition domain.

The key components of the ESCRT system continued to duplicate and subfunctionalize during the evolution of the major lineages of eukaryotes; distinct ESCRT-0 components emerged, providing an interface for clathrin and clathrin adaptor complexes, the latter evolving from the adaptin appendage domain (Fig. 1 and 6).

Functions of ESCRT systems in archaea

During the recent decade, considerable progress has been reached in understanding the functionality of both FtsZ-based and ESCRT-based molecular machineries in archaea (76). Two independent recent studies reported PRC (Photosynthetic Reaction Center) barrel protein to be essential for cell division in *Haloferax volcanii* (64, 77). It has been shown that two PRC barrel proteins form a tripartite complex with SepF which recruits FtsZ to the divisosome (64, 77). It has been thought previously that Thermoproteales had a unique cell division apparatus because neither FtsZ nor ESCRT system genes were identified in these genomes (42). It has been hypothesized that these organisms used an actin homolog (crenactin hereinafter) along with three other genes (arcadin 1, 2, 3, and 4, which are often encoded in vicinity) for cell division (42, 78). However, experimental analysis of these protein led to inconclusive results (78, 79) so that the role of actin in cell division and the functions of arcadin 1, 3, and 4 remain unclear. As indicated above, PRC barrel is an N-terminal domain of CdvA, and recently, a diverged

CdvA ortholog has been identified in Thermoproteales (64) (Fig. 2 and 5). Apart from the PRC barrel domain in CdvA, several Thermoproteales genomes encode additional PRC barrel proteins (Fig. 2 and 5B). These proteins are potential additional components of the cell division machinery in this archaeal lineage.

Whereas most of euryarchaea apparently encompass the FtsZ-based division machinery, with all three currently known components (FtsZ, SepF, and PRC barrel) conserved, several Thermoplasmata genomes encode only FtsZ but not SepF and PRC, and some others lack all three components (Fig. 5A). The latter group includes genomes of three genera, *Picrophilus*, *Acidiplasma*, and *Ferroplasma*. *Acidiplasma* and *Ferroplasma* encode a minimal ESCRT system described above along with actin (Fig. 5A). Considering that this ESCRT system, similarly to the Cdv system, lacks components required for involvement in protein sorting, it might have been co-opted for cell division in these organisms. However, other functions cannot be ruled out because *Thermoplasma volcanium*, *T. acidophilum*, and Thermoplasmatales archaeon GpI (*Cuniculiplasma divulgatum*) genomes encoded this system along with the complete FtsZ machinery. The only euryarchaeal genome that lacks all genes implicated in cell division is *Picrophilus torridus* (Fig. 5A). The only candidate for involvement in cell division that we identified in this genome is a member of Thermoplasmatales-specific uncharacterized arCOG07400 family (WP_011177753.1).

All other complex ESCRT systems are present in organisms that also possess a (mostly) complete FtsZ cell division machinery suggesting that they are not involved in cell division, but rather, in vesicle formation and protein sorting (Table S1). In particular, like Thaumarchaea, most Asgard archaea encode FtsZ and SepF (and some PRC barrel proteins) along with ESCRT (Table S1). However, unlike in Thaumarchaea that use ESCRT for cell division, Asgard ESCRT systems are likely dedicated to vesicular transport of protein cargo, mostly, ubiquitinated proteins, whereas cell division in Asgard archaea likely relies solely on the FtsZ machinery. However, given that in eukaryotes, the ESCRT system is multifunctional and is involved in cytokinetic abscission (18, 19, 80, 81), the ESCRT system of Asgard archaea also might contribute to cell division.

Many archaea produce vesicles by budding (82), but only in Sulfolobales, the ESCRT system was found to be responsible for the formation of these vesicles suggesting multiple roles of the ESCRT system in these organisms (83, 84). Although many other archaea, especially Thermococcales, in which vesicles were observed (82), do harbor at least some components of the ESCRT machinery, others, such as *Halorubrum lacusprofundi* and *Aciduliprofundum boonei*, do not, making it unlikely that ESCRT is solely responsible for vesicle formation in archaea. ESCRT systems in the main and FHA group systems are associated with S/T kinases, suggesting that protein phosphorylation is key to their functions. The fusion of the FHA domain with the CdvA-like domain suggests that the ESCRT system recognizes phosphorylated proteins as cargo. In contrast, in the case of the Halo_adaptin and Meth_adaptin systems, the cargo is probably bound via protein-protein interactions involving IG-like or PQQ-like domains (Fig. 3B). Considering that the Halo_adaptin system is the most diverged one and is present in archaea that divide using the FtsZ system, this ESCRT system is likely involved in vesicle formation.

Concluding remarks

Our analysis of the distribution, domain architectures, and genomic neighborhoods of ESCRT components in archaea revealed striking evolutionary flexibility and modularity, most conspicuously, in Asgardarchaeota. Although undoubtedly subject to amendment by comparative analysis of an even greater diversity of archaeal genomes yet to be sequenced, the available data are sufficient to develop a plausible scenario of the origin and evolution of the ESCRT machinery.

The ESCRT system involved in protein sorting is an early archaeal innovation. Only ESCRT-III antedates archaea, deriving from a PspA-like protein that was probably involved in membrane repair in the LUCA (43, 73). We can now place ESCRT-II and adaptin appendage domains into the LACA and hypothesize that the LACA also possessed

analogs of Vps28-C (alpha helical bundle). The exact interactions between the components in different archaeal ESCRT systems remain to be elucidated. We identified strong contextual links between ESCRT and such domains as GOLD and SPFH, suggesting that building blocks for these eukaryotic signature systems originated early in archaeal evolution. The ESCRT machinery of the LACA was likely involved primarily in protein sorting via endocytosis. The subsequent evolution of ESCRT took divergent paths in different major lineages of archaea. In TACK and Thermoplasmata, the ancestral ESCRT system seems to have evolved toward secondary simplification, whereby only the core components of the system were retained, whereas all putative ESCRT-0 analogs were lost. In these groups of archaea, ESCRT was co-opted for cell division and, at least in Sulfolobales, for extracellular vesicle biogenesis (49). Notably, this reduction of the ESCRT system and its repurposing for cell division occurred on at least two independent occasions in the evolution of archaea. Conversely, in Asgard archaea, ESCRT evolved toward complexification, whereby ESCRT complexes subfunctionalized to mediate membrane remodeling through interaction with the actin-based cytoskeleton (85). Furthermore, the link between ESCRT and the ubiquitin network presently considered a signature of the eukaryotic cell biology, likely evolved in Asgard archaea (22, 39, 45). Asgard archaea also possess multiple ESCRT components that apparently have not been inherited by eukaryotes (Fig. 2A and 6B). The complex relationship between the archaeal and eukaryotic ESCRT machineries described here contrasts the more direct correspondence between the components of several other conserved macromolecular complexes, such as the proteasome (86) or the exosome (87). This distinction likely reflects pronounced differences between membrane remodeling and protein sorting processes in Asgard archaea and eukaryotes, despite the conservation of the core machinery.

Multiple recent phylogenetic analyses based on a variety of marker genes converge on the conclusion that eukaryotes are either the sister clade of the Heimdall supergroup, which includes Heimdallarchaeia, Kariarchaeia, Gerdarchaeia, and Hodarchaeia lineages, or the sister clade of Hodarchaeia within the Heimdall supergroup (27, 39, 45, 46). However, our phylogenetic analysis of Vps4 and previous analyses of ESCRT III (43, 73) and ESCRT II components (22, 39) suggest affinity between eukaryotes and Asgardarchaeota as a group rather than with the Heimdall supergroup or Hodarchaeia. Furthermore, we did not detect any synapomorphies (derived shared characters such as common domain architectures) between either the Heimdall supergroup or any other Asgard lineages and eukaryotes. On the contrary, there are several distinct features in the organization of the core ESCRT machinery in Asgards as discussed above. So far, the conclusion of origin of eukaryotes within the Heimdall supergroup or from a common ancestor with Hodarchaeia is solely based on phylogenetic analysis of (nearly) universal genes, but no systematic search for synapomorphies shared between eukaryotes and different Asgard lineages has been reported. Considering the essential role of ESCRT in both Asgard archaea and eukaryotes, analysis of ESCRT components suggests extra caution regarding the inference of the exact nature of the evolutionary relationship between Asgardarchaeota and eukaryotes. A comprehensive search for synapomorphies to complement the phylogenetic analyses should help clarify these relationships and shed more light on the eukaryotic ancestry.

MATERIALS AND METHODS

Orthologous gene data sets

The arCOGs and asCOGs databases were used as the comparative genomic framework (39, 40, 88). The arCOGs are annotated clusters of orthologous genes from 524 archaeal genomes covering all major archaeal lineages, whereas asCOGs are annotated clusters of orthologous genes from 78 genomes of Asgardarchaeota covering all major Asgard lineages.

Genome and protein sequence analysis

PSI-BLAST (89) was used to identify homologs of all known ESCRT proteins and homologs of all proteins found in their neighborhoods in the arCOGs database. Given that the sequences of many ESCRT proteins are compositionally biased, the searches were run with variable parameter sets and different queries. Typically, PSI-BLAST was run for five iterations or until convergence, with inclusion threshold E-value = 0.0001 and compositional adjustment statistics turn off. Low scoring hits were manually examined and verified using HHpred (90). HHpred search was run on the web against PDB (91), CDD v3.16 (92), and PFAM_A v36 (93) derived profile libraries imbedded in the HHpred toolkit with otherwise default parameters. All HHpred hits were visually examined, and the highest-scoring relevant hit is reported for newly identified domains (Table S3). The Marcoil program (94) was used to predict coiled-coil regions, and TMHMM (95) was used to predict transmembrane segments. Domain boundaries were determined based on the combination of evidence including examination of HHpred hits and comparison of output alignments with the respective structures from the PDB database (when available) and examination of AF2 models and transmembrane and coiled-coil regions predicted using TMHMM and Marcoil, respectively. Based on this analysis, 11 arCOGs and one asCOG (cog.002483) were corrected, and six new arCOGs (arCOG15281-6) were created.

In order to increase the diversity of archaeal ESCRT groups identified here (Meth_adaptin and Halo_adaptin), the NCBI NR database was searched for closely related homologs in database using Vps4 ATPase sequences from respective genomes as queries. Respective genomic loci were examined to ascertain that they encoded all or most of the identified ESCRT components from the respective groups. These loci and the respective protein accessions are listed in Tables S2 and S3.

Phylogenetic analysis

The Vps4 protein sequences were the clustered using Mmseqs2 (96) with a 90% protein identity threshold, and a representative of each cluster was included in the sequence set for phylogenetic tree reconstruction. Muscle5 (97) with default parameters was used to construct the multiple alignment of the Vps4 family. For phylogenetic analysis, several poorly aligned sequences or fragments were discarded. Columns in the multiple alignment were filtered for homogeneity value (98) of 0.05 or greater and a gap fraction less than 0.667. This filtered alignment was used as the input for the IQ-tree (54) to construct maximum likelihood phylogenetic tree with the LG + F + R10 evolutionary model (File S1; Table S1). The same program was used to calculate support values. To maximize the number of phylogenetically informative positions for the set containing only Vps4 sequences from archaea and eukaryotes, the position of the root was determined by building a second tree which contained only archaeal Vps4 sequences from the arCOGs and several selected Cdc48 family ATPases (C-terminal domain), the closest homologs of Vps4, as an outgroup (Fig. S1; File S2; Table S1). The alignment for this tree was filtered as described above, and the tree was reconstructed using IQ-tree with the LG + R6 best fitted evolutionary model. The Vps4 tree shown on the Fig. 3A was then rooted according to the root position found in the tree that included Cdc48 as an outgroup (Fig. S1).

Protein structure prediction and analysis

For protein structure predictions, all relevant sequences were extracted from the respective gene neighborhoods (Table S2) and queried against a combination of databases [UniRef30-2022/05 (99), pdb70-2022/03/01 (100), and mgnify-2022/05 (101)] using HHblits (100). The resulting multiple sequence alignemnts were used as input for Alphafold2.3.1 (47) using the monomer model and generating five model structures per alignment. For structure visualizations, the overall highest pLDDT scoring model was

used. USCF ChimeraX was used for all structural analysis and visualization (102). Structure comparisons were performed using DALI (103).

ACKNOWLEDGMENTS

This work was supported by the Shenzhen Medical Research Fund (B2301005), National Natural Science Foundation of China (grant nos. 32225003, 32393970, 32393971, 32000002, and 92251306), Guangdong Basic and Applied Basic Research Foundation (2023A1515011309), Guangdong Major Project of Basic and Applied Basic Research(2023B0303000017), and Shenzhen University 2035 Program for Excellent Research (2022B002). K.S.M., V.T., Y.I.W., and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (National Institutes of Health, National Library of Medicine).

K.S.M. and E.V.K. initiated the project; K.S.M. collected the data; K.S.M., V.T., Y.I.W., Z.L., Y.L., S.Z., M.K., M.L., and E.V.K. analyzed the data; K.S.M., V.T., and E.V.K. wrote the manuscript that was read, edited, and approved by all authors.

AUTHOR AFFILIATIONS

[1]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, USA
[2]Archaeal Biology Center, Institute for Advanced Study, Shenzhen University, Shenzhen, China
[3]Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, China
[4]Archaeal Virology Unit, Institut Pasteur, Université de Paris, Paris, France

AUTHOR ORCIDs

Kira S. Makarova http://orcid.org/0000-0002-8174-2844
Yuri I. Wolf http://orcid.org/0000-0002-0247-8708
Mart Krupovic http://orcid.org/0000-0001-5486-0098
Meng Li http://orcid.org/0000-0001-8675-0758
Eugene V. Koonin http://orcid.org/0000-0003-3943-8299

FUNDING

Funder	Grant(s)	Author(s)
National Institutes of Health	Intramural Research Program	Eugene V. Koonin

AUTHOR CONTRIBUTIONS

Kira S. Makarova, Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review and editing | Victor Tobiasson, Investigation, Writing – original draft, Writing – review and editing | Yuri I. Wolf, Investigation, Methodology, Writing – review and editing | Zhongyi Lu, Investigation, Writing – review and editing | Yang Liu, Investigation, Writing – review and editing | Siyu Zhang, Investigation, Writing – review and editing | Mart Krupovic, Investigation, Writing – review and editing | Meng Li, Investigation, Supervision, Writing – review and editing | Eugene V. Koonin, Conceptualization, Supervision, Writing – original draft, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Eugene V. Koonin, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Berend Snel, Universiteit Utrecht, and Christa Schleper, University of Vienna.

DATA AVAILABILITY

The data used in this work are available through public databases or the Supplementary Material. Additional Supplementary files (File S1 and S2 and Asgard sequences corresponding to Table S2) are available via the FTP (https://ftp.ncbi.nih.gov/pub/makarova/Supplement/ESCRT/).

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Figure S1 (mBio00335-24-s0001.pdf). Phylogenetic tree of archaeal Vps4 tree rooted using Cdc48 as an outgroup.
Figure S2 (mBio00335-24-s0002.pdf). Structural models for the main clade of ESCRT systems.
Figure S3 (mBio00335-24-s0003.pdf). Multiple alignment of arCOG08177 sequences.
Figure S4 (mBio00335-24-s0004.pdf). Alignment and structural models for CdvA helical hairpin and Steadiness box (SB).
Figure S5 (mBio00335-24-s0005.pdf). Structural models for the Halo_FHA clade.
Figure S6 (mBio00335-24-s0006.pdf). Structural models for the Meth_adaptin clade.
Figure S7 (mBio00335-24-s0007.pdf). Structural models for the Halo_adaptin clade.
Table S1 (mBio00335-24-s0008.xlsx). ESCRT and cell division system components in archaea.
Table S2 (mBio00335-24-s0009.xlsx). Annotated neighborhoods of ESCRT genes.
Table S3 (mBio00335-24-s0010.xlsx). Evidence of sequence similarity for predicted components of archaeal ESCRT systems.

REFERENCES

1. Szathmáry E. 2015. Toward major evolutionary transitions theory 2.0. Proc Natl Acad Sci U S A 112:10104–10111. https://doi.org/10.1073/pnas.1421398112
2. Koonin EV, Mulkidjanian AY. 2013. Evolution of cell division: from shear mechanics to complex molecular machineries. Cell 152:942–944. https://doi.org/10.1016/j.cell.2013.02.008
3. Adams DW, Errington J. 2009. Bacterial cell division: assembly, maintenance and disassembly of the Z ring. Nat Rev Microbiol 7:642–653. https://doi.org/10.1038/nrmicro2198
4. Caspi Y, Dekker C. 2018. Dividing the archaeal way: the ancient Cdv cell-division machinery. Front Microbiol 9:174. https://doi.org/10.3389/fmicb.2018.00174
5. den Blaauwen T, Hamoen LW, Levin PA. 2017. The divisome at 25: the road ahead. Curr Opin Microbiol 36:85–94. https://doi.org/10.1016/j.mib.2017.01.007
6. Ouellette SP, Lee J, Cox JV. 2020. Division without binary fission: cell division in the FtsZ-less chlamydia J Bacteriol 202:e00252-20. https://doi.org/10.1128/JB.00252-20
7. Vitorino IR, Lage OM. 2022. The *Planctomycetia*: an overview of the currently largest class within the phylum *Planctomycetes*. Antonie Van Leeuwenhoek 115:169–201. https://doi.org/10.1007/s10482-021-01699-0
8. Lindås A-C, Karlsson EA, Lindgren MT, Ettema TJG, Bernander R. 2008. A unique cell division machinery in the archaea. Proc Natl Acad Sci U S A 105:18942–18946. https://doi.org/10.1073/pnas.0809467105
9. Ithurbide S, Gribaldo S, Albers SV, Pende N. 2022. Spotlight on FtsZ-based cell division in archaea. Trends Microbiol 30:665–678. https://doi.org/10.1016/j.tim.2022.01.005
10. Pende N, Sogues A, Megrian D, Sartori-Rupp A, England P, Palabikyan H, Rittmann SK-MR, Graña M, Wehenkel AM, Alzari PM, Gribaldo S. 2021. SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. Nat Commun 12:3214. https://doi.org/10.1038/s41467-021-23099-8
11. Rowlett VW, Margolin W. 2015. The Min system and other nucleoid-independent regulators of Z ring positioning. Front Microbiol 6:478. https://doi.org/10.3389/fmicb.2015.00478
12. Makarova KS, Koonin EV. 2010. Two new families of the FtsZ-tubulin protein superfamily implicated in membrane remodeling in diverse bacteria and archaea. Biol Direct 5:33. https://doi.org/10.1186/1745-6150-5-33
13. Hoshino S, Hayashi I. 2012. Filament formation of the FtsZ/tubulin-like protein TubZ from the *Bacillus cereus* pXO1 plasmid. J Biol Chem 287:32103–32112. https://doi.org/10.1074/jbc.M112.373803
14. Duggin IG, Aylett CHS, Walsh JC, Michie KA, Wang Q, Turnbull L, Dawson EM, Harry EJ, Whitchurch CB, Amos LA, Löwe J. 2015. CetZ tubulin-like proteins control archaeal cell shape. Nature 519:362–365. https://doi.org/10.1038/nature13983
15. Buchanan M. 2017. The origin of cell division. Nature Phys 13:526–526. https://doi.org/10.1038/nphys4164
16. Field MC, Sali A, Rout MP. 2011. Evolution: on a bender--BARs, ESCRTs, COPs, and finally getting your coat. J Cell Biol 193:963–972. https://doi.org/10.1083/jcb.201102042
17. Vietri M, Radulovic M, Stenmark H. 2020. The many functions of ESCRTs. Nat Rev Mol Cell Biol 21:25–42. https://doi.org/10.1038/s41580-019-0177-4
18. Carlton JG, Martin-Serrano J. 2007. Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. Science 316:1908–1912. https://doi.org/10.1126/science.1143422
19. Christ L, Raiborg C, Wenzel EM, Campsteijn C, Stenmark H. 2017. Cellular functions and molecular mechanisms of the ESCRT membrane-scission machinery. Trends Biochem Sci 42:42–56. https://doi.org/10.1016/j.tibs.2016.08.016
20. Schöneberg J, Lee I-H, Iwasa JH, Hurley JH. 2017. Reverse-topology membrane scission by the ESCRT proteins. Nat Rev Mol Cell Biol 18:5–17. https://doi.org/10.1038/nrm.2016.121

21. Migliano SM, Wenzel EM, Stenmark H. 2022. Biophysical and molecular mechanisms of ESCRT functions, and their implications for disease. Curr Opin Cell Biol 75:102062. https://doi.org/10.1016/j.ceb.2022.01.007
22. Hatano T, Palani S, Papatziamou D, Salzer R, Souza DP, Tamarit D, Makwana M, Potter A, Haig A, Xu W, Townsend D, Rochester D, Bellini D, Hussain HMA, Ettema TJG, Löwe J, Baum B, Robinson NP, Balasubramanian M. 2022. Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. Nat Commun 13:3398. https://doi.org/10.1038/s41467-022-30656-2
23. Katzmann DJ, Stefan CJ, Babst M, Emr SD. 2003. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. J Cell Biol 162:413–423. https://doi.org/10.1083/jcb.200302136
24. Kostelansky MS, Schluter C, Tam YYC, Lee S, Ghirlando R, Beach B, Conibear E, Hurley JH. 2007. Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. Cell 129:485–498. https://doi.org/10.1016/j.cell.2007.03.016
25. Feng GH, Lih CJ, Cohen SN. 2000. TSG101 protein steady-state level is regulated posttranslationally by an evolutionarily conserved COOH-terminal sequence. Cancer Res 60:1736–1741.
26. Im YJ, Wollert T, Boura E, Hurley JH. 2009. Structure and function of the ESCRT-II-III interface in multivesicular body biogenesis. Dev Cell 17:234–243. https://doi.org/10.1016/j.devcel.2009.07.008
27. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173–179. https://doi.org/10.1038/nature14447
28. Scott A, Gaspar J, Stuchell-Brereton MD, Alam SL, Skalicky JJ, Sundquist WI. 2005. Structure and ESCRT-III protein interactions of the MIT domain of human Vps4A. Proc Natl Acad Sci U S A 102:13813–13818. https://doi.org/10.1073/pnas.0502165102
29. ter Haar E, Musacchio A, Harrison SC, Kirchhausen T. 1998. Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. Cell 95:563–573. https://doi.org/10.1016/s0092-8674(00)81623-2
30. Gomez-Navarro N, Miller E. 2016. Protein sorting at the ER-golgi interface. J Cell Biol 215:769–778. https://doi.org/10.1083/jcb.201610031
31. Kovtun O, Dickson VK, Kelly BT, Owen DJ, Briggs JAG. 2020. Architecture of the AP2/clathrin coat on the membranes of clathrin-coated vesicles. Sci Adv 6:eaba8381. https://doi.org/10.1126/sciadv.aba8381
32. Traub LM, Downs MA, Westrich JL, Fremont DH. 1999. Crystal structure of the α appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. Proc Natl Acad Sci U S A 96:8907–8912. https://doi.org/10.1073/pnas.96.16.8907
33. Praefcke GJK, Ford MGJ, Schmid EM, Olesen LE, Gallop JL, Peak-Chew S-Y, Vallis Y, Babu MM, Mills IG, McMahon HT. 2004. Evolving nature of the AP2 α-appendage hub during clathrin-coated vesicle endocytosis. EMBO J 23:4371–4383. https://doi.org/10.1038/sj.emboj.7600445
34. Chen Z, Schmid SL. 2020. Evolving models for assembling and shaping clathrin-coated pits. J Cell Biol 219:e202005126. https://doi.org/10.1083/jcb.202005126
35. Teo H, Gill DJ, Sun J, Perisic O, Veprintsev DB, Vallis Y, Emr SD, Williams RL. 2006. ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. Cell 125:99–111. https://doi.org/10.1016/j.cell.2006.01.047
36. Samson RY, Obita T, Hodgson B, Shaw MK, Chong PL-G, Williams RL, Bell SD. 2011. Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. Mol Cell 41:186–196. https://doi.org/10.1016/j.molcel.2010.12.018
37. Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, Dai J, Koonin EV, Li G, Chu H, Li M. 2020. Coevolution of eukaryote-like Vps4 and ESCRT-III subunits in the Asgard archaea. mBio 11:e00417-20. https://doi.org/10.1128/mBio.00417-20
38. Frohn BP, Härtel T, Cox J, Schwille P. 2022. Tracing back variations in archaeal ESCRT-based cell division to protein domain architectures. PLoS One 17:e0266395. https://doi.org/10.1371/journal.pone.0266395
39. Liu Y, Makarova KS, Huang WC, Wolf YI, Nikolskaya AN, Zhang X, Cai M, Zhang CJ, Xu W, Luo Z, Cheng L, Koonin EV, Li M. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature 593:553–557. https://doi.org/10.1038/s41586-021-03494-3
40. Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. Life (Basel) 5:818–840. https://doi.org/10.3390/life5010818
41. Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci U S A 99:1420–1425. https://doi.org/10.1073/pnas.032658599
42. Makarova KS, Yutin N, Bell SD, Koonin EV. 2010. Evolution of diverse cell division and vesicle formation systems in archaea. Nat Rev Microbiol 8:731–741. https://doi.org/10.1038/nrmicro2406
43. Liu J, Tassinari M, Souza DP, Naskar S, Noel JK, Bohuszewicz O, Buck M, Williams TA, Baum B, Low HH. 2021. Bacterial Vipp1 and PspA are members of the ancient ESCRT-III membrane-remodeling superfamily. Cell 184:3660–3673. https://doi.org/10.1016/j.cell.2021.05.041
44. Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. Nat Ecol Evol 4:138–147. https://doi.org/10.1038/s41559-019-1040-x
45. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJG. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541:353–358. https://doi.org/10.1038/nature21031
46. Eme L, Tamarit D, Caceres EF, Stairs CW, De Anda V, Schön ME, Seitz KW, Dombrowski N, Lewis WH, Homa F, Saw JH, Lombard J, Nunoura T, Li W-J, Hua Z-S, Chen L-X, Banfield JF, John ES, Reysenbach A-L, Stott MB, Schramm A, Kjeldsen KU, Teske AP, Baker BJ, Ettema TJG. 2023. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature 618:992–999. https://doi.org/10.1038/s41586-023-06186-2
47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
48. Pelve EA, Lindås A-C, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. 2011. Cdv-based cell division and cell cycle organization in the thaumarchaeon *Nitrosopumilus maritimus*. Mol Microbiol 82:555–566. https://doi.org/10.1111/j.1365-2958.2011.07834.x
49. Liu J, Cvirkaite-Krupovic V, Commere PH, Yang Y, Zhou F, Forterre P, Shen Y, Krupovic M. 2021. Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. ISME J 15:2892–2905. https://doi.org/10.1038/s41396-021-00984-0
50. Nuñez JK, Lee ASY, Engelman A, Doudna JA. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature 519:193–198. https://doi.org/10.1038/nature14237
51. Iyer LM, Koonin EV, Aravind L. 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43:134–144. https://doi.org/10.1002/1097-0134(20010501)43:2<134::aid-prot1025>3.0.co;2-i
52. Mao Y. 2021. Structure, dynamics and function of the 26S proteasome. Subcell Biochem 96:1–151. https://doi.org/10.1007/978-3-030-58971-4_1
53. Dubiel W, Chaithongyot S, Dubiel D, Naumann M. 2020. The COP9 signalosome: a multi-DUB complex. Biomolecules 10:1082. https://doi.org/10.3390/biom10071082
54. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
55. Almawi AW, Matthews LA, Guarné A. 2017. FHA domains: phosphopeptide binding and beyond. Prog Biophys Mol Biol 127:105–110. https://doi.org/10.1016/j.pbiomolbio.2016.12.003
56. Durocher D, Jackson SP. 2002. The FHA domain. FEBS Lett 513:58–66. https://doi.org/10.1016/s0014-5793(01)03294-x
57. Anantharaman V, Aravind L. 2002. The GOLD domain, a novel protein module involved in golgi function and secretion. Genome Biol 3:research0023. https://doi.org/10.1186/gb-2002-3-5-research0023
58. Nagae M, Hirata T, Morita-Matsumoto K, Theiler R, Fujita M, Kinoshita T, Yamaguchi Y. 2016. 3D structure and interaction of p24β and p24δ golgi dynamics domains: implication for p24 complex formation and cargo transport. J Mol Biol 428:4087–4099. https://doi.org/10.1016/j.jmb.2016.08.023

59. Meister M, Bänfer S, Gärtner U, Koskimies J, Amaddii M, Jacob R, Tikkanen R. 2017. Regulation of cargo transfer between ESCRT-0 and ESCRT-I complexes by flotillin-1 during endosomal sorting of ubiquitinated cargo. Oncogenesis 6:e344. https://doi.org/10.1038/oncsis.2017.47
60. Scholz AS, Baur SSM, Wolf D, Bramkamp M. 2021. An stomatin, prohibitin, flotillin, and HflK/C-domain protein required to link the phage-shock protein to the membrane in *Bacillus subtilis*. Front Microbiol 12:754924. https://doi.org/10.3389/fmicb.2021.754924
61. Ma C, Wang C, Luo D, Yan L, Yang W, Li N, Gao N. 2022. Structural insights into the membrane microdomain organization by SPFH family proteins. Cell Res 32:176–189. https://doi.org/10.1038/s41422-021-00598-3
62. Haardt M, Kempf B, Faatz E, Bremer E. 1995. The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system prou of *Escherichia coli* K-12. Mol Gen Genet 246:783–786. https://doi.org/10.1007/BF00290728
63. Eugster A, Frigerio G, Dale M, Duden R. 2004. The α- and β'-COP WD40 domains mediate cargo-selective interactions with distinct di-lysine motifs. Mol Biol Cell 15:1011–1023. https://doi.org/10.1091/mbc.e03-10-0724
64. Zhao S, Makarova KS, Zheng W, Liu Y, Zhan L, Wan Q, Gong H, Krupovic M, Lutkenhaus J, Chen X, Koonin EV, Du S. 2023. Widespread PRC barrel proteins play critical roles in archaeal cell division. bioRxiv. https://doi.org/10.1101/2023.03.28.534520
65. Lu Z, Zhang S, Liu Y, Xia R, Li M. 2023. Origin of eukaryotic-like Vps23 shapes an ancient functional interplay between ESCRT and ubiquitin system in Asgard archaea. Cell Rep. https://doi.org/10.1016/j.celrep.2024.113781
66. Gill DJ, Teo H, Sun J, Perisic O, Veprintsev DB, Emr SD, Williams RL. 2007. Structural insight into the ESCRT-I/-II link and its role in MVB trafficking. EMBO J 26:600–612. https://doi.org/10.1038/sj.emboj.7601501
67. Im YJ, Hurley JH. 2008. Integrated structural model and membrane targeting mechanism of the human ESCRT-II complex. Dev Cell 14:902–913. https://doi.org/10.1016/j.devcel.2008.04.004
68. Jansen RM, Hurley JH. 2023. Longin domain GAP complexes in nutrient signalling, membrane traffic and neurodegeneration. FEBS Lett 597:750–761. https://doi.org/10.1002/1873-3468.14538
69. Hofmann K, Bucher P. 1998. The PCI domain: a common theme in three multiprotein complexes. Trends Biochem Sci 23:204–205. https://doi.org/10.1016/s0968-0004(98)01217-1
70. Perez-Riba A, Itzhaki LS. 2019. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. Curr Opin Struct Biol 54:43–49. https://doi.org/10.1016/j.sbi.2018.12.004
71. Bork P, Holm L, Sander C. 1994. The immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol 242:309–320. https://doi.org/10.1006/jmbi.1994.1582
72. Rinke C, Chuvochina M, Mussig AJ, Chaumeil P-A, Davín AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. 2021. A standardized archaeal taxonomy for the genome taxonomy database. Nat Microbiol 6:946–959. https://doi.org/10.1038/s41564-021-00918-8
73. Popp PF, Gumerov VM, Andrianova EP, Bewersdorf L, Mascher T, Zhulin IB, Wolf D. 2022. Phyletic distribution and diversification of the phage shock protein stress response system in bacteria and archaea. mSystems 7:e0134821. https://doi.org/10.1128/msystems.01348-21
74. Iyer LM, Leipe DD, Koonin EV, Aravind L. 2004. Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146:11–31. https://doi.org/10.1016/j.jsb.2003.10.010
75. Sancho E, Vilá MR, Sánchez-Pulido L, Lozano JJ, Paciucci R, Nadal M, Fox M, Harvey C, Bercovich B, Loukili N, Ciechanover A, Lin SL, Sanz F, Estivill X, Valencia A, Thomson TM. 1998. Role of UEV-1, an inactive variant of the E2 ubiquitin-conjugating enzymes, in *in vitro* differentiation and cell cycle behavior of HT-29-M6 intestinal mucosecretory cells. Mol Cell Biol 18:576–589. https://doi.org/10.1128/MCB.18.1.576
76. van Wolferen M, Pulschen AA, Baum B, Gribaldo S, Albers S-V. 2022. The cell biology of archaea. Nat Microbiol 7:1744–1755. https://doi.org/10.1038/s41564-022-01215-8
77. Nußbaum P, Kureisaite-Ciziene D, Bellini D, van der Does C, Kojic M, Taib N, Gribaldo S, Loose M, Löwe J, Albers S-V. 2023. PRC domain-containing proteins modulate FtsZ-based archaeal cell division. bioRxiv. https://doi.org/10.1101/2023.03.28.534543
78. Ettema TJG, Lindås A-C, Bernander R. 2011. An actin-based cytoskeleton in archaea. Mol Microbiol 80:1052–1061. https://doi.org/10.1111/j.1365-2958.2011.07635.x
79. Izoré T, Kureisaite-Ciziene D, McLaughlin SH, Löwe J. 2016. Crenactin forms actin-like double helical filaments regulated by arcadin-2. Elife 5:e21600. https://doi.org/10.7554/eLife.21600
80. Gao C, Zhuang X, Shen J, Jiang L. 2017. Plant ESCRT complexes: moving beyond endosomal sorting. Trends Plant Sci 22:986–998. https://doi.org/10.1016/j.tplants.2017.08.003
81. Yagisawa F, Fujiwara T, Takemura T, Kobayashi Y, Sumiya N, Miyagishima SY, Nakamura S, Imoto Y, Misumi O, Tanaka K, Kuroiwa H, Kuroiwa T. 2020. ESCRT machinery mediates cytokinetic abscission in the unicellular red alga *Cyanidioschyzon merolae*. Front Cell Dev Biol 8:169. https://doi.org/10.3389/fcell.2020.00169
82. Gill S, Catchpole R, Forterre P. 2019. Extracellular membrane vesicles in the three domains of life and beyond. FEMS Microbiol Rev 43:273–303. https://doi.org/10.1093/femsre/fuy042
83. Soler N, Marguet E, Verbavatz JM, Forterre P. 2008. Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order Thermococcales. Res Microbiol 159:390–399. https://doi.org/10.1016/j.resmic.2008.04.015
84. Liu J, Cvirkaite-Krupovic V, Baquero DP, Yang Y, Zhang Q, Shen Y, Krupovic M. 2021. Virus-induced cell gigantism and asymmetric cell division in archaea. Proc Natl Acad Sci U S A 118:e2022578118. https://doi.org/10.1073/pnas.2022578118
85. Rodrigues-Oliveira T, Wollweber F, Ponce-Toledo RI, Xu J, Rittmann SK-MR, Klingl A, Pilhofer M, Schleper C. 2023. Actin cytoskeleton and complex cell architecture in an Asgard archaeon. Nature 613:332–339. https://doi.org/10.1038/s41586-022-05550-y
86. Fuchs ACD, Alva V, Lupas AN. 2021. An astonishing wealth of new proteasome homologs. Bioinformatics 37:4694–4703. https://doi.org/10.1093/bioinformatics/btab558
87. Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E. 2005. The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat Struct Mol Biol 12:575–581. https://doi.org/10.1038/nsmb952
88. Makarova KS, Wolf YI, Koonin EV. 2019. Towards functional characterization of archaeal genomic dark matter. Biochem Soc Trans 47:389–398. https://doi.org/10.1042/BST20180560
89. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389
90. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
91. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, Craig PA, Crichlow GV, Dalenberg K, Duarte JM, et al. 2023. RCSB protein data bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Res 51:D488–D508. https://doi.org/10.1093/nar/gkac1077
92. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48:D265–D268. https://doi.org/10.1093/nar/gkz991
93. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
94. Delorenzi M, Speed T. 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625. https://doi.org/10.1093/bioinformatics/18.4.617
95. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model:

application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315
96. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988
97. Edgar RC. 2022. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun 13:6968. https://doi.org/10.1038/s41467-022-34630-w
98. Esterman ES, Wolf YI, Kogay R, Koonin EV, Zhaxybayeva O. 2021. Evolution of DNA packaging in gene transfer agents. Virus Evol 7:veab015. https://doi.org/10.1093/ve/veab015
99. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932. https://doi.org/10.1093/bioinformatics/btu739
100. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20:473. https://doi.org/10.1186/s12859-019-3019-7
101. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, Burgin J, Caballero-Pérez J, Cochrane G, Colwell LJ, Curtis T, Escobar-Zepeda A, Gurbich TA, Kale V, Korobeynikov A, Raj S, Rogers AB, Sakharova E, Sanchez S, Wilkinson DJ, Finn RD. 2023. MGnify: the microbiome sequence data analysis resource in 2023. Nucleic Acids Res 51:D753–D759. https://doi.org/10.1093/nar/gkac1080
102. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 30:70–82. https://doi.org/10.1002/pro.3943
103. Holm L, Laiho A, Törönen P, Salgado M. 2023. DALI shines a light on remote homologs: one hundred discoveries. Protein Sci 32:e4519. https://doi.org/10.1002/pro.4519