A hot spring Asgard archaeon sheds light on the origin of eukaryotic endosomal system.

Weili Lin
Tongji University
Lu Fan (fanl@sustech.edu.cn)
Southern University of Sicence and Technology https://orcid.org/0000-0002-4184-7211
Jing Xiao
Tongji University
Sa Fang
Tongji University
Yanbing Xu
Tongji University
Rui Zhao
Tongji University
Rushi Qin
Tongji University
Ruixin Zhu
Tongji University

Original article

Keywords: Odinarchaeceae Tengchong, Asgard archaea, ESCRT machinery, endosomal system, autotrophic lifestyle

Posted Date: March 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16808/v1

Abstract

Recent metagenomics studies have identified a novel archaeal superphylum namely Asgard, which are characterized by enriched eukaryotic-specific proteins. In this study, we screened unclassified archaeal genomes in public databases and obtained a high-qualified metagenome-assembled genome that can be assigned as a novel family-level Asgard member namely Odinarchaeceae Tengchong. Metabolic analysis indicates an autotrophic lifestyle of this hot spring archaeon with a complete tetrahydromethanopterin Wood-Ljungdahl pathway for carbon dioxide reduction and an arsenic efflux detoxification. Examination of public databases found that thus far Odinarchaeceae Tengchong may be the only prokaryote that encodes a C-terminal domain of Vps28 in the endosomal sorting complex required for transport (ESCRT), a critical connector of multiple ESCRT components. Therefore, the identification of this archaeon provides valuable evidence of the archaeal origin of eukaryotic ESCRT. We posit that all the key components of the eukaryotic endosomal system might have evolved from a common ancestor of Asgard archaea and eukaryotes.

Key Points

- A family-level new autotroph Asgard archaeon with complete Wood-Ljungdahl pathway.
- This archaeon possesses all the conserved components of eukaryotic ESCRT system.
- The ESCRT system might have evolved from common ancestor of Asgard and eukaryotes.

Introduction

Asgard archaea are comprised of at least five phyla including Heimdallarchaeota, Lokiarchaeota, Odinarchaeota, Thorarchaeota and Helarchaeota (Seitz et al. 2019; Zaremba-Niedzwiedzka et al. 2017). They are found from a variety of environments including lake sediments, mangrove sediments, estuarine sediments and mud volcanos (Cai et al. 2018). Recently, a Lokiarchaeota-related archaeon (Candidatus Prometheoarchaeum syntrophicum strain MK-D1) has been reported to grow in co-culture with a methanogen (Imachi et al. 2019). Metagenomic analysis indicated that Asgard archaea are of divergent metabolic capabilities and potentially live a mixotrophic life style (Cai et al. 2018). For example, Lokiarchaeota and Thorarchaeota use both the tetrahydrofolate Wood–Ljungdahl (THF/H_4F-WL) and tetrahydromethanopterin Wood–Ljungdahl (THMPT/H_4MPT-WL) pathway for fixing CO_2 and performing acetogenesis. They also have the capability to degrade organic matter (Cai et al. 2018; Spang et al. 2019). Heimdallarchaeota and Odinarchaeota can only use THF-WL or THMPT-WL pathway, respectively (Cai et al. 2018). In particular, Heimdallarchaeota is the only group of Asgard archaea that possesses the complete set of genes for the forward and reverse tricarboxylic acid (TCA) cycle (Cai et al. 2018).

Asgard archaea are also noted by enrichment in eukaryotic signature proteins (ESPs) (Zaremba-Niedzwiedzka et al. 2017), which are considered to be ubiquitous in eukaryotes but have few homologs in bacteria and other archaea. For instance, metagenome-assembled genomes (MAGs) of Thorarchaeota encode several eukaryotic membrane-trafficking components and a odinarchaeal MAG possesses a bona fide eukaryotic tubulin (Zaremba-Niedzwiedzka et al. 2017). These ESPs shed light on the origin of eukaryotic cellular complexity. However, only partial components of some key eukaryotic-specific processes are encoded by Asgard archaea. Archaeal homologues of some other essential membrane-trafficking components including coat protein and adaptor protein complexes are yet to be discovered (Dacks and Robinson 2017).

The endosomal sorting complex required for transport (ESCRT) is an important part of the eukaryotic endomembrane system and a substantial feature that distinguishes eukaryotes from prokaryotes. ESCRT is involved in cytokinetic abscission and phagophore formation. Moreover, it mediates the sorting of ubiquitylated membrane proteins into multivesicular bodies and plays an important role in vesicular trafficking processes (Henne et al. 2011). Since the development of vesicle trafficking is pivotal in the formation of a more complex endomembrane system, the discovery of ESCRT components in archaeal genomes is of great significance. However, up to now, only part of the ESCRT components have been found in archaea. Among all ESCRT components, three are conserved across the eukaryotic lineages: ESCRT-I, -II and -III (Leung et al. 2008). ESCRT-III has been found in many archaeal lineages (Obita et al. 2007; Samson et al. 2008), while ESCRT-I and ESCRT-II are only enriched in Asgard group (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017). ESCRT-II is encoded by all Asgard archaea but ESCRT-I is only reported in Heimdallarchaeota LC_3 and Lokiarchaeum sp. GC14_75 (Zaremba-Niedzwiedzka et al. 2017). The incompleteness of ESCRT in Asgard archaea (Spang et al. 2015) limits our understanding of the origin of this system.

The lack of high-quality genomes in each phylum of Asgard has created confusion and contradiction in Asgard phylogeny and evolution (Betts et al. 2018; Da Cunha et al. 2017; Rokas et al. 2018; Spang et al. 2018; Xiao et al. 2019; Zaremba-Niedzwiedzka et al. 2017), especially for the phylum Odinarchaeota. Here we conducted a screening of unclassified archaeal MAGs in public databases and identified a novel Asgard archaeon, which was originally sampled from hot spring sediments. This MAG belongs to the phylum Odinarchaeota and is 98.13% complete

and has 2.8% contamination. It is to date the only discovered Asgard archaeon that contains a complete and conserved ESCRT system, which may shed light on the origination of eukaryotic endosomal system.

Materials And Methods

Phylogenetic identification of novel archaeal MAGs. The unclassified archaeal MAGs were downloaded from IMG and NCBI databases. Their information is provided in Table S1. The methods of assembly and binning of Odinarchaeceae Tengchong were described in the analysis project in JGI database (https://gold.jgi.doe.gov/analysis_projects?id=Ga0181714). The JGI study project Gs0127627 was recently published in the study of Hua et al (Hua et al. 2019). Specifically, metagenomic assembly and binning were performed with SPAdes v.3.9.0 (Bankevich et al. 2012) and Metabat (Kang et al. 2015) & ESOM (Ultsch and Mörchen 2005), respectively. Protein sequences of the studied MAGs were downloaded from the IMG database. Specifically, protein sequences of Odinarchaeceae Tengchong were predicted with Prodigal (v2.6.3) (Hyatt et al. 2010) with the "-p single" option, as described in the study of Hua et al (Hua et al. 2019).

A set of 15 syntenic ribosomal proteins: L2, L3, L4, L5, L6P, L14, L15, L18, L22, L24, S3, S8, S10, S17 and S19 (Castelle et al. 2015; Zaremba-Niedzwiedzka et al. 2017) were searched in the unclassified archaeal MAGs by using PSI-BLAST. 51 unclassified archaeal MAGs that contain at least six of the 15 ribosomal proteins were identified. The ribosomal proteins of these 51 archaea were aligned with those of known Asgard and TACK archaea by using MAFFT-L-INS-i (Katoh and Standley 2013), trimmed by using trimAl (Capella-Gutierrez et al. 2009) and concatenated. Maximum-likelihood phylogeny of the 15 ribosomal proteins was inferred with the PROTCATLG model and fast bootstrapped in RAxML (Stamatakis 2014) (version 8.2.10). Further maximum-likelihood analysis with LG + C60 + F + G + PMSF model (Wang et al. 2018) was performed on concatenated 55 ribosomal proteins (Table S2) in more qualified Asgard and TACK genomes (> 70% completeness and < 5% contamination in CheckM assessment). The support values were calculated with 100 bootstraps replicated under the LG + C60 + F + G model. 16S ribosomal RNA gene of Odinarchaeceae Tengchong was predicted using RNAmmer (Lagesen et al. 2007). 16S rRNA gene identity of Odinarchaeceae Tengchong and Odinarchaeceae LCB4 was calculated using needle (http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/needle.html). The completeness, contamination and strain heterogeneity of the novel archaeal MAG were evaluated using the "checkm lineage_wf" command in CheckM (Parks et al. 2015) and A'nvio (Eren et al. 2015).

Identification of ESPs. Protein domains were identified with IPRscan with default parameters (Jones et al. 2014). As to the definition of ESP (Hartman and Fedorov 2002), they are present in most eukaryotes but absent in almost all archaea and bacteria. Accordingly, we selected eukaryotic-specific IPR domains based on the standard that each was detected in more than 500 eukaryotes but less than 20 archaea and 20 bacteria. Proteins with eukaryotic-specific IPR domains were identified as ESPs. It is needed to be specified here that both COG5491 of the CDD database and PF03357 of the Pfam database are annotated as Snf7 family. However, contents of COG5491 and PF03357 are different in two ways. First, the seed alignment of PF03357 contains 32 eukaryotic sequences (https://pfam.xfam.org/family/PF03357#tabview=tab5), while that of COG5491 contain six archaeal sequences and six yeast sequences (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=227778). Second, PF03357 contains sequences annotated as eukaryotic-specific endosome-mediated trafficking in function, while those of COG5491 are annotated as general cellular functions such as cell cycle control, cell division and chromosome partitioning. Therefore, COG5491 and PF03357 represent two functionally distinguishable groups of proteins although these proteins all belong to the Snf7 family. In this study, we think PF03357 is more suitable to annotate eukaryotic specific proteins in archaeal genomes. Furthermore, ESPs of ESCRT components in Asgard were also searched using HHpred (Soding et al. 2005) and PSI-BLAST in HHpred Server (Zimmermann et al. 2018) with default E-value cutoff (< 1-e3). Standard databases used in HHpred and PSI-BLAST analyses were PDB_mmCIF70_4_Feb (default) and uniprot_tembl_6_Dec, respectively. ESPs of ESCRT were strictly identified as positive protein domains only when they were detected by all of the IPRscan, HHpred and PSI-BLAST methods. Phylogenetic analyses were performed on selective ESPs. Sequences of the selected ESPs were aligned by using MAFFT-L-INS-I (Katoh and Standley 2013) and trimmed by using trimAl (Capella-Gutierrez et al. 2009) with gappyout option. Maximum likelihood inference was performed on the trimmed alignments by using RAxML (Stamatakis 2014) with PROTGAMMALG model (10 independent trees were generated for optimization) and 100 non-parametric bootstrap replicates. Protein structure of ESPs were predicted with homology modelling by using Phyre2 (Kelley et al. 2015) tool. Chimera (Pettersen et al. 2004) was used to view the predicted structures of ESPs.

Metabolic reconstruction. The orthologous genes were retrieved to KEGG Orthology (KO) IDs and archaeal clusters of orthologous genes (arCOGs) by using the eggnog-mapper tool (version:1.0.3) with archaea database in emapper DB (version: 4.5.1) (Huerta-Cepas et al., 2017). Genes of KO were mapped to KEGG metabolic pathways with KEGG mapper tool (Kanehisa and Goto, 2000). In addition, genes were queried against the NCBI non-redundant (nr) database with Diamond (Buchfink et al., 2015) blastp option 0.9.22, and the top hit of each gene was extracted.

Results

Taxonomic classification

Phylogenetic analysis of 51 unclassified archaeal MAGs based on 15 concatenated ribosomal proteins revealed a novel Asgard archaeon (IMG Taxon ID 2721755898) that is closely related to *Odinarchaeceae* LCB4 (Fig. 1). The identity of 16S rRNA gene sequences between the novel archaeon and *Odinarchaeceae LCB4* is 85.8%, which is below the family threshold of 86.5% but above the order threshold of 83.6% (Yarza et al. 2014). Therefore, this novel archaeon is a family-level novel Asgard member. We named it *Odinarchaeceae* Tengchong, referring to the place where its discovered –Tengchong, Yunnan, China (Hua et al. 2019). The taxonomic classification of *Odinarchaeceae* Tengchong was futher confirmed with in-depth phylogenetic analyses of 55 concatenated ribosomal proteins in more qualified Asgard genomes (Fig. 2). The MAG of *Odinarchaeceae* Tengchong is 2.21M bp in size with a G+C content of 47.6%. It is 98.13% complete and contains 2.8% contaminated sequences as assessed by using CheckM (Parks et al. 2015), while is 91.36% complete as predicted by using A'nvio (Eren et al. 2015).

Metabolism of *Odinarchaeceae* Tengchong

Metabolic reconstruction indicated that *Odinarchaeceae* Tengchong encodes the complete reductive acetyl-CoA WL pathway suggesting its ability of assimilating CO_2 (Fig. 3). Two subtypes of the WL pathway uses different C1 carriers: the THF-WL pathway uses tetrahydrofolate and the THMPT-WL pathway uses tetrahydromethanopterin. Our results show that in Asgard archaea, *Odinarchaeceae* Tengchong, *Lokiarchaeum* sp. GC14_75 and all *Thorarchaeota* archaea encode all enzymes of a complete THMPT-WL pathway (Fig. S1a-S1j, Table S3a-S3j). In contrast, *Heimdallarchaeota* archaea use THF as C1 carrier. Both *Heimdallarchaeota* LC_2 and *Heimdallarchaeota* LC_3 encode a complete THF-WL pathway (Fig. S1k-S1m, Table S3k-S3m). In Archaea, the WL pathway has been linked to methanogenesis (Borrel et al. 2016). However, *Odinarchaeceae* Tengchong lacks the gene encoding methyl-CoM reductase, a key enzyme required for methane production.

Odinarchaeceae Tengchong has genetic potential to convert acetyl-CoA to β-D-Frutose-6P by enzymes involved in gluconeogenesis. In the ribulose-5-phosphate pathway, β-D-Frutose-6P can be further converted to D-Ribulose-5P by 6-phospho-3-hexuloisomerase (5.3.1.27 in Fig. 3) and 3-hexulose-6-phosphate synthase (4.1.2.43). Ribose 5-phosphate isomerase A then converts D-Ribulose-5P to D-Ribose-5P, which is an important precursor in nucleotide biosynthesis.

The MAG of *Odinarchaeceae* Tengchong contained a nearly complete glycolytic pathway except two key enzymes, glucose-6-phosphate isomerase (5.3.1.9) and pyruvate kinase (2.7.1.40) (Fig. 3). This result is consistent with previous studies for *Odinarchaeceae* LCB4 (Cai et al. 2018; Spang et al. 2019), indicating that archaea in the *Odinarchaeota* phylum may not have a functional glycolysis pathway. In addition, most of the key enzymes in TCA cycle were missing in *Odinarchaeceae* Tengchong, suggesting that this archaeon does not rely on this pathway for energy generation.

Arsenic transport and metabolism pathways were found in *Odinarchaeceae* Tengchong, including phosphate ABC transporter, arsenic reductase (1.20.4.4) and arsenic exporter (ArsA) (Fig. 3). In addition, arsenite methyltransferase was found, indicating that *Odinarchaeceae* Tengchong may have an arsenic methylation pathway.

Distribution of ESPs in Asgard phyla

Over 1,860 eukaryotic-specific InterPro (IPR) domains (Jones et al. 2014) were identified in this study (Table S4). Among Asgard archaea, two types of ESP distribution patterns are noticeable. In *Odinarchaeota*, *Lokiarchaeota* and *Heimdalarchaeota*, the ESCRT and ubiquitin modifier systems are enriched but the trafficking machinery is lacking, whereas in *Thorarchaeota*, the opposite is true (Fig. 4). We named them 'ESCRT-ubiquitin modifier-enriched pattern' and 'trafficking machinery-enriched pattern', respectively.

Previous studies have questioned the purity of the MAGs of Asgard archea (Da Cunha et al., 2017; (Garg et al. 2019). While the MAG of *Odinarchaeceae* Tengchong appears to be high quality based on single-copy gene justification, we cannot rule out the possibility that some of the ESPs predicted in this MAG may be exogenous. To further ensure the ESPs discovered are encoded by *Odinarchaeceae* Tengchong, we manually examinated the G+C content distribution in the MAG of *Odinarchaeceae* Tengchong especially around the regions encoding the ESPs (Table S5) and detected no evidence of outstanding differences with the rest regions indicating that the *Odinarchaeceae* Tengchong MAG suffers little in assembly or binning errors.

Similarity of the ESCRT proteins between *Odinarchaeceae* Tengchong, other archaea and eukaryotes

All conserved components of the ESCRT system were identified in the MAG of *Odinarchaeceae* Tengchong including an accessory component — vacuolar fusion protein Mon1 (Table 1, Table S6 and Fig. 4). Notably, *Odinarchaeceae* Tengchong uniquely encodes the C-terminal domain

of Vps28 (Vps28^{CTD}) in ESCRT-I compared to other prokaryotes including other Asgard archaea (Fig. 5). The Vps28^{CTD} of *Odinarchaeceae* Tengchong locates precisely at the C-terminal of Vps28, which is similar to most eukaryotic counterparts (Fig. 5). Homology modelling analysis shows the higher similiarity of the C-terminal squences of Vps28 (with 26% identity in the aligned 119th-128th amino acid) between *Odinarchaeceae* Tengchong and its eukaryotic templete (c2j9wB_ in *Xenopus laevis*), compared to other Asgard archaea.

The structural models of ESCRT components between archaea and eukaryotic counterparts were generated and compared based on the PDB database. (Fig. 6 and Table 2). In both *Odinarchaeceae* Tengchong and *Caenorhabditis briggsae*, Vps28-like proteins of ESCRT-I have alpha helixes of C-terminal domains (Fig. 6a-6b). Their steadiness box proteins of ESCRT-I have both alpha helixes and beta strands (Fig. 6e-6f), while the homologues of *Lokiarchaeum* sp. GC14_75 and *Thorarchaeota* AB25 only have one secondary structure, respectively (Fig. 6g-6h). Furthermer, Snf7 structure of *C. briggsae* resembles those of *Odinarchaeceae* Tengchong and *Heimdallarchaeota* LC3, but is vastly different from the structure of remote homology in *Nitrososphaera viennensis* EN76, a species of *Thaumarchaeota* (Fig. 6l-6o). Compared with other Asgard and TACK archaea (TACK superphylum comprises the *Thaumarchaeota*, *Crenarchaeota* and *Korarchaeota* as decrisebed in the study of Guy *et al.* (Guy and Ettema 2011)), all ESCRT components of *Odinarchaeceae* Tengchong showed high model confidence, alignment coverage and identity to the eukaryotic template proteins (Table 2).

To further exclude the possibility that the ESCRT proteins of *Odinarchaeceae* Tengchong are from contamination of eukaryotic sequences, we conducted a genomic G+C content-based examination. G+C contents of the DNA sequences of the ESPs range from 39.0% to 53.4%, which is similar to that of the rest of the MAG (i.e. 47.6%).

Discussion

Odinarchaeceae Tengchong may be a thermophilic autotroph

Odinarchaeceae Tengchong has a complete WL pathway, indicating it has the potential for autotrophic metabolism. Meanwhile, the WL pathway is usually linked to methanogenesis and acetogenesis in Archaea (Ljungdahl 2009). Methanogenesis is considered to be one of the oldest metabolic pathways in Archaea (Borrel et al. 2016), and plays an important role in the global carbon cycle (Evans et al. 2019; Lyu et al. 2018). Genomic analysis showed that *Odinarchaeceae* Tengchong had the potential to produce acetate but lacked some genes involved in methane production. The results are consistent with the metabolic characteristics of Asgard archaea and some *Bathyarchaeota* (Cai et al. 2018; Zhou et al. 2018). On the other hand, the missing of most of the enzymes in the TCA cycle is consistent with previous studies for *Odinarchaeceae* LCB4 (Cai et al. 2018; Spang et al. 2019), indicating that *Odinarchaeota* may be unable to conduct heterotrophic metabolism. Thus, we posit that *Odinarchaeceae* Tengchong is an anaerobic thermophilic autotroph living in the sediments of hot springs in which the oxygen content is relatively low and organic matter may be limited.

It is intriguing to find the complete pathway for arsenic efflux detoxification in *Odinarchaeceae* Tengchong. Arsenic is a toxic element and, in response to its toxicity, prokaryotes have evolved a variety of coping strategies (Paez-Espino et al. 2009). Previous studies have shown that arsenic can be abundant in geothermal waters (Paez-Espino et al. 2009). Thus, *Odinarchaeceae* Tengchong may have evolved to adapt to geothermal fluids of Tengchong hot springs containing high arsenic (Guo et al. 2017). Other studies also reported genes related to arsenic metabolism in *Odinarchaeceae* LCB4 (Cai et al. 2018; Spang et al. 2019) and *Thorarchaeota* (Liu et al. 2018). Thus, the arsenic efflux detoxification pathway may reflect adaption towards arsenic toxicity by diverse species of archaea.

Odinarchaeceae Tengchong shares highly similar characteristics of ESCRT proteins with eukaryotes.

With an independently folded four-helical bundle, Vps28^{CTD} function as a critical connector of multiple ESCRT components (Pineda-Molina et al., 2006). The interactions between Vps28^{CTD} and other ESCRT components are essentially required for the sorting function (Bowers et al. 2004; Gill et al. 2007; Teo et al. 2006). The deletion of Vps28^{CTD} leads to accumulation of cargoes and large aberrant late multivesicular endosomes (MVE), and finally blocks the sorting function of endosome, which is termed “class E compartment” (Kostelansky et al. 2006). According to our analysis, all conserved ESCRT components (ESCRT-I, -II and -III) and vesicle trafficking protein related with endosomal sorting (vacuolar fusion protein Mon1) are encoded by *Odinarchaeceae* Tengchong. Specifically, *Odinarchaeceae* Tengchong to date is the only known prokaryote encodes Vps28^{CTD} (Fig. 4). Vps28^{CTD} of *Odinarchaeceae* Tengchong locates precisely at the C-terminal of Vps28 as that of eukaryotes do (Fig. 5). Therefore, the MAG of *Odinarchaeceae* Tengchong may encode the currently most complete archaeal set of ESCRT proteins. On the other hand, homology modelling shows that ESCRT proteins of *Odinarchaeceae* Tengchong have high similarity evaluations with the eukaryotic counterparts, including model confidence, alignment coverage and identity (Table 2). ESCRT-I proteins of *Odinarchaeceae* Tengchong and eukaryotes have more common secondary structures, including alpha helixes of Vps28^{CTD} and alpha helixes and beta strands

of steadiness box proteins (Fig. 6a-6h). These results suggest ESCRT proteins of *Odinarchaeceae* Tengchong are among those mostly closely related to the eukaryotic counterparts.

Heterogeneous evolutionary histories of eukaryotic ESCRT components and membrane trafficking proteins

The respective origins of eukaryotic ESCRT components are still in mystery. Endosomal systems are comprised of four ESCRT components: ESCRT-0, -I, -II and -III, plus accessory components (such as vesicle trafficking proteins). Among them, ESCRT-I, -II and -III are conserved across the eukaryotic lineages (Leung et al. 2008). Phylogenetic analysis in this study revealed different evolutionary histories of ESCRT components: ESCRT-III might emerge earlier than ESCRT-I and ESCRT-II. Specifically, ESCRT-III likely originated from the common ancestor of TACK, Asgard archaea and eukaryotes (Fig. S3e, Fig. S4e). A possible functional divergence of ESCRT-III occurred in the common ancestor of *Thaumarchaeota* archaea, Asgard archaea and eukaryotes supported by Snf7 protein structure divergence between Asgard archaea and *N. viennensis* EN76 (Fig. 6l-6o).

In contrast, ESCRT-I and ESCRT-II components seemed to originate later in the common ancestor of Asgard and eukaryotes (Fig. S3a-S3d, Fig. S4a-S4d and Fig. 7). During the divergence of Asgard lineages, *Odinarchaeceae* Tengchong preserved all the ESCRT components while other Asgard species in may have gradually lost certain ESCRT-I components. Furthermore, the obvious 'ESCRT-ubiquitin modifier-enriched pattern' and 'trafficking machinery-enriched pattern' of ESPs in Asgard archaea suggests functional divergence of these proteins in evolutionary history of this superphylum.

In this study, we performed analysis of unclassified archaeal MAGs across public databases. A novel family-level Asgard archaeon *Odinarchaeceae* Tengchong was identified and it is distinguishable from other prokaryotes in that it encodes all conserved components of ESCRT machinery including $Vps28^{CTD}$. These ESCRT components shares highly similar characteristics with eukaryotic counterparts. The complete and conserved ESCRT machinery in *Odinarchaeceae* Tengchong indicates that ESCRT-I and -II systems possibly originated from the common ancestor of Asgard archaea and eukaryotes. Future work is encouraged to explore the details in the archaeal origin of the eukaryotic ESCRT machinery.

Declarations

Ethics approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Competing interests

The authors have declared that no competing interests exist.

Funding

This work was supported by the National Natural Science Foundation of China No. 81774152 (RZ) and No. 91951120 (LF), the Shanghai Committee of Science and Technology 16ZR1449800 (RZ), and the Shenzhen Science and Technology Innovation Commission JCYJ20180305123458107 (LF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

RZ and LF conceived and designed the project. Each author has contributed significantly to data analysis. WL and LF drafted the manuscript. JX, SF, YX, RZ, RQ and RZ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr Wenjun Li and Dr Zhengshuang Hua from the Sun Yat-Sen University for permitting us to use the genomic sequencing of *Odinarchaeceae* Tengchong for this study. Discussion with Dr Wenjun Li helped improving the quality of this paper. Computation in this study was supported by the Centre for Computational Science and Engineering at the Southern University of Science and Technology.

Abbreviations

ESCRT, endosomal sorting complexes required for transport; **pH**, potential of hydrogen; **MAG**, metagenome-assembled genome; **ESP**, eukaryotic signature protein; **IPR**, InterPro**; Vps28**, Vacuolar protein sorting-associated protein 28; **Vps28^{CTD}**, C-terminal domain of vacuolar protein sorting-associated protein 28; **THF/H_4F-WL**, tetrahydrofolate Wood–Ljungdahl; **THMPT/H_4MPT-WL**, tetrahydromethanopterin Wood–Ljungdahl; **TCA**, tricarboxylic acid.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455-77 doi:10.1089/cmb.2012.0021

Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol 2(10):1556-1562 doi:10.1038/s41559-018-0644-x

Borrel G, Adam PS, Gribaldo S (2016) Methanogenesis and the Wood-Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. Genome Biol Evol 8(6):1706-11 doi:10.1093/gbe/evw114

Bowers K, Lottridge J, Helliwell SB, Goldthwaite LM, Luzio JP, Stevens TH (2004) Protein-protein interactions of ESCRT complexes in the yeast *Saccharomyces cerevisiae*. Traffic 5(3):194-210 doi:10.1111/j.1600-0854.2004.00169.x

Cai M, Liu Y, Zhou Z, Yang Y, Pan J, Gu J-D, Li M (2018) Asgard archaea are diverse, ubiquitous, and transcriptionally active microbes. bioRxiv doi:10.1101/374165

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-3 doi:10.1093/bioinformatics/btp348

Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC, Williams KH, Banfield JF (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25(6):690-701 doi:10.1016/j.cub.2015.01.014

Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P (2017) *Lokiarchaea* are close relatives of *Euryarchaeota*, not bridging the gap between prokaryotes and eukaryotes. PLoS Genet 13(6):e1006810 doi:10.1371/journal.pgen.1006810

Dacks JB, Robinson MS (2017) Outerwear through the ages: evolutionary cell biology of vesicle coats. Current opinion in cell biology 47:108-116 doi:10.1016/j.ceb.2017.04.001

Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319 doi:10.7717/peerj.1319

Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, Tyson GW (2019) An evolving view of methane metabolism in the Archaea. Nature reviews Microbiology 17(4):219-232 doi:10.1038/s41579-018-0136-7

Garg SG, Kapust N, Lin W, Nelson-Sathi S, Tria FD, Gould SB, Fan L, Zhu R, Zhang C, Martin WF (2019) Anomalous phylogenetic behavior of ribosomal proteins in metagenome assembled genomes. bioRxiv:731091 doi:https://doi.org/10.1101/731091

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. The EMBO journal 26(2):600-12 doi:10.1038/sj.emboj.7601501

Guo Q, Planer-Friedrich B, Liu M, Li J, Zhou C, Wang YJCG (2017) Arsenic and thioarsenic species in the hot springs of the Rehai magmatic geothermal system, Tengchong volcanic region, China. Chemical Geology 453:12-20

Guy L, Ettema TJ (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol 19(12):580-7 doi:10.1016/j.tim.2011.09.002

Hartman H, Fedorov A (2002) The origin of the eukaryotic cell: a genomic investigation. Proceedings of the National Academy of Sciences of the United States of America 99(3):1420-5 doi:10.1073/pnas.032658599

Henne WM, Buchkovich NJ, Emr SD (2011) The ESCRT pathway. Dev Cell 21(1):77-91 doi:10.1016/j.devcel.2011.05.015

Hua ZS, Wang YL, Evans PN, Qu YN, Goh KM, Rao YZ, Qi YL, Li YX, Huang MJ, Jiao JY, Chen YT, Mao YP, Shu WS, Hozzein W, Hedlund BP, Tyson GW, Zhang T, Li WJ (2019) Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea. Nat Commun 10(1):4574 doi:10.1038/s41467-019-12574-y

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics 11:119 doi:10.1186/1471-2105-11-119

Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, Matsui Y, Miyazaki M, Murata K, Saito Y, Sakai S, Song C, Tasumi E, Yamanaka Y, Yamaguchi T, Kamagata Y, Tamaki H, Takai K (2019) Isolation of an archaeon at the prokaryote-eukaryote interface. bioRxiv:726976 doi:10.1101/726976

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236-40 doi:10.1093/bioinformatics/btu031

Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165 doi:10.7717/peerj.1165

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30(4):772-80 doi:10.1093/molbev/mst010

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols 10(6):845-58 doi:10.1038/nprot.2015.053

Kostelansky MS, Sun J, Lee S, Kim J, Ghirlando R, Hierro A, Emr SD, Hurley JH (2006) Structural and functional organization of the ESCRT-I trafficking complex. Cell 125(1):113-126

Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic acids research 35(9):3100-8 doi:10.1093/nar/gkm160

Leung KF, Dacks JB, Field MC (2008) Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. Traffic 9(10):1698-716 doi:10.1111/j.1600-0854.2008.00797.x

Liu Y, Zhou Z, Pan J, Baker BJ, Gu JD, Li M (2018) Comparative genomic inference suggests mixotrophic lifestyle for *Thorarchaeota*. ISME J 12(4):1021-1031 doi:10.1038/s41396-018-0060-x

Ljungdahl LG (2009) A life with acetogens, thermophiles, and cellulolytic anaerobes. Annual review of microbiology 63:1-25 doi:10.1146/annurev.micro.091208.073617

Lyu Z, Shao N, Akinyemi T, Whitman WB (2018) Methanogenesis. Current biology : CB 28(13):R727-R732 doi:10.1016/j.cub.2018.05.021

Obita T, Saksena S, Ghazi-Tabatabai S, Gill DJ, Perisic O, Emr SD, Williams RL (2007) Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4. Nature 449(7163):735-9 doi:10.1038/nature06171

Paez-Espino D, Tamames J, de Lorenzo V, Canovas D (2009) Microbial responses to environmental arsenic. Biometals 22(1):117-30 doi:10.1007/s10534-008-9195-y

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7):1043-55 doi:10.1101/gr.186072.114

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. Journal of computational chemistry 25(13):1605-12 doi:10.1002/jcc.20084

Rokas A, Da Cunha V, Gaia M, Nasir A, Forterre P (2018) Asgard archaea do not close the debate about the universal tree of life topology. PLOS Genetics 14(3) doi:10.1371/journal.pgen.1007215

Samson RY, Obita T, Freund SM, Williams RL, Bell SD (2008) A role for the ESCRT system in cell division in archaea. Science 322(5908):1710-3 doi:10.1126/science.1165322

Seitz KW, Dombrowski N, Eme L, Spang A, Lombard J, Sieber JR, Teske AP, Ettema TJG, Baker BJ (2019) Asgard archaea capable of anaerobic hydrocarbon cycling. Nat Commun 10(1):1822 doi:10.1038/s41467-019-09364-x

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244-8 doi:10.1093/nar/gki408

Spang A, Eme L, Saw JH, Caceres EF, Zaremba-Niedzwiedzka K, Lombard J, Guy L, Ettema TJG (2018) Asgard archaea are the closest prokaryotic relatives of eukaryotes. PLoS Genet 14(3):e1007080 doi:10.1371/journal.pgen.1007080

Spang A, Saw JH, Jorgensen 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(7551):173-179 doi:10.1038/nature14447

Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, Greening C, Baker BJ, Ettema TJG (2019) Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat Microbiol 4(7):1138-1148 doi:10.1038/s41564-019-0406-9

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312-3 doi:10.1093/bioinformatics/btu033

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(1):99-111 doi:10.1016/j.cell.2006.01.047

Ultsch A, Mörchen F (2005) ESOM-Maps: tools for clustering, visualization, and classification with emergent SOM. Department of Mathematics and Computer Science, University of Marburg, Marburg. Germany. Research Report 46

Wang HC, Minh BQ, Susko E, Roger AJ (2018) Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst Biol 67(2):216-235 doi:10.1093/sysbio/syx068

Xiao J, Fan L, Wu D, Xu Y, Lai D, Martin WF, Zhu R, Zhang C (2019) Archaea, the tree of life, and cellular evolution in eukaryotes. Science China Earth Sciences 62(3):489-506

Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W, Schleifer KH, Whitman WB, Euzeby J, Amann R, Rossello-Mora R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature reviews Microbiology 12(9):635-45 doi:10.1038/nrmicro3330

Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Backstrom 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 TJ (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541(7637):353-358 doi:10.1038/nature21031

Zhou Z, Pan J, Wang F, Gu JD, Li M (2018) *Bathyarchaeota*: globally distributed metabolic generalists in anoxic environments. FEMS microbiology reviews 42(5):639-655 doi:10.1093/femsre/fuy023

Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. Journal of molecular biology 430(15):2237-2243 doi:10.1016/j.jmb.2017.12.007

Tables

Table 1. Overview of ESCRT components in *Odinarchaeceae* Tengchong.

ESP	Protein ID	IPR-domains (e-value)	HHpred (probability/e-value)	PSI-BLAST (e-value)	Summarized annotations
ESCRT-I: Vps28-like	2724281155	IPR007143: Vacuolar protein sorting-associated, VPS28 (2.1E-15)	2J9W_A: VPS28-prov protein (100/1E-33)	A0A524CD77: VPS28 C-terminal domain-containing protein OS=*Candidatus* Lokiarchaeota archaeon OX=2 (5.92E-47)	VPS28
ESCRT-I: C-terminal domain of Vps28	2724281155	IPR017899: Vacuolar protein sorting-associated, VPS28, C-terminal (14.527 [a])	2J9W_A: VPS28-prov protein (100/1E-33)	A0A524CD77: VPS28 C-terminal domain-containing protein OS=*Candidatus* Lokiarchaeota archaeon OX=2 (5.92E-47)	VPS28
ESCRT-I: steadiness box domain	2724281154	IPR017916: Steadiness box (SB) domain (9.201 [a])	3BZH_A: Ubiquitin-conjugating enzyme E2 E1 (E.C.6.3.2.19) (99.84/2.2E-22)	A0A523TPU1: Ubiquitin-conjugating enzyme E2 OS=*Candidatus* Heimdallarchaeota archaeon OX=2026747 (3.88E-15)	Steadiness box (SB) domain-containing ubiquitin-conjugating enzyme E2 [c]
	2724283317	IPR017916: Steadiness box (SB) domain (9.201 [a])	2CAY_B: Vacuolar protein sorting protein 36 (97.3/3E-05)	A0A524E902: RING-type domain-containing protein OS=*Thorarchaeota* archaeon (strain OWC) OX=205349 (8.52E-45)	Steadiness box (SB) domain-containing protein
ESCRT-II: EAP30 domain	2724281156	IPR007286: EAP30 (1.7E-19)	3CUQ_A: Vacuolar-sorting protein SNF8, Vacuolar protein-sorting-associated (100/6.2E-52)	A0A0F8XJ03: EAP30 domain protein (ESCRT-II) (Vps22/36-like) OS=*Lokiarchaeum* sp. (strain GC14_75) (4.61E-55)	EAP30 domain protein
ESCRT-II: Vps25-like	2724281931	IPR014041: ESCRT-II complex, Vps25 subunit, N-terminal winged helix (2.9E-8)	3HTU_A: Vacuolar protein-sorting-associated protein (Vps) 25, Vacuolar (99.68/8.6E-20)	A0A0F8W1S5: Vps25-like protein (ESCRT-II) OS=*Lokiarchaeum* sp. (strain GC14_75) OX=1538547 GN=Lok (1.57E-15)	Vps25 protein
	2724281931	IPR008570: ESCRT-II complex, vps25 subunit (6.9E-11)	3HTU_A: Vacuolar protein-sorting-associated protein (Vps) 25, Vacuolar (99.68/8.6E-20)	A0A0F8W1S5: Vps25-like protein (ESCRT-II) OS=*Lokiarchaeum* sp. (strain GC14_75) OX=1538547 GN=Lok (1.57E-15)	Vps25 protein
ESCRT-III: Snf7 family	2724283318	IPR005024: Snf7 family (4.9E-19)[b]	5FD9_A: Endosomal sorting complex required for transpot; ESCRT, Snf7, Active, Core, Protein (99.78/1.3E-19)	A0A5B9DB92: Snf7 OS=anaerobic archaeon MK-D1 OX=2594042 GN=DSAG12_02101 PE=4 SV=1 (1.87E-15)	Snf7 family

[a] Profilescan-location score in Profilescan analysis method.

[b] See Materials and Methods for protein module selection.

[c] Ubiquitin-conjugating enzyme E2 and the steadiness box domains of ESCRT-I are located in the same protein in *Odinarchaeceae* Tengchong, they are both involved in ESCRT-mediated protein degradation. Of note, the ubiquitin-conjugating enzyme E2 and steadiness box domains are also both lie in the protein P25604 of *Saccharomyces cerevisiae*.

Table 2. Homology information of protein domains and their templates in homology modelling.

Protein domains	Species	Protein ID	Template	Template Information (PDB Molecule)	Confidence	Coverage (Start-End/Protein length)	Identity
ESCRT-I: Vps28-like	*Odinarchaeceae* Tengchong	2724281155	c2j9wB_	vps28-prov protein	100%	44% (119-218/222)	26%
	Odinarchaeceae LCB4	OLS18189.1	c2j9wB_	vps28-prov protein	100%	42% (129-225/229)	22%
	Heimdallarchaeota LC 3	OLS24758.1	c2j9wB_	vps28-prov protein	93.6%	26% (253-344/347)	27%
	Heimdallarchaeota AB 125	OLS31938.1	c2j9wB_	vps28-prov protein	100%	21% (370-467/469)	21%
	Lokiarchaeum sp. GC14_75	KKK45392.1	c2j9wB_	vps28-prov protein	100%	39% (128-215/218)	23%
		KKK45107.1	c2j9wB_	vps28-prov protein	81.4%	81% (20-107/109)	20%
	Lokiarchaeota CR 4	OLS13437.1	c2j9wB_	vps28-prov protein	94.4%	33% (122-207/261)	15%
ESCRT-I: Steadiness box domain	*Odinarchaeceae* Tengchong	2724281154	c5a4pA_	ubiquitin-conjugating enzyme e2 z [b]	100%	79% (8-215/253)	19%
		2724283317	d2f6ma1	VPS23 C-terminal domain [a]	86.9%	11% (429-481/484)	28%
	Lokiarchaeum sp. GC14_75	KKK44603.1	d2f6mb1	VPS28 N-terminal domain [a]	89.4%	10% (568-632/650)	28%
	Thorarchaeota AB 25	OLS30567.1	d2caya1	VPS36 N-terminal domain-like [a]	81.5%	15% (230-310/505)	18%
	Thorarchaeota WOR 45	KXH75182.1	d2caza1	VPS23 C-terminal domain [a]	89.7%	13% (357-410/418)	28%
	Thorarchaeota MP11T 1	PJET01000114.1_23	d2caza1	VPS23 C-terminal domain [a]	90.6%	11% (444-497/505)	28%
ESCRT-II: EAP30 domain	*Odinarchaeceae* Tengchong	2724281156	c2zmeA_	vacuolar-sorting protein snf8	100%	84% (30-244/244)	27%
	Odinarchaeceae LCB 4	OLS18190.1	c2zmeA_	vacuolar-sorting protein snf8	100%	84% (30-242/243)	25%
	Heimdallarchaeota LC 3	OLS27539.1	c2zmeA_	vacuolar-sorting protein snf8	100%	85% (30-235/235)	20%
	Heimdallarchaeota LC 2	OLS22747.1	c2zmeA_	vacuolar-sorting protein snf8	100%	87% (30-227/227)	19%
	Heimdallarchaeota AB 125	OLS31936.1	c2zmeA_	vacuolar-sorting protein snf8	100%	85% (30-240/241)	27%
		OLS32704.1	c2zmeA_	vacuolar-sorting protein snf8	100%	82% (42-246/247)	20%
	Lokiarchaeum sp. GC14_75	KKK42119.1	c2zmeA_	vacuolar-sorting protein snf8	100%	86% (30-237/237)	24%
	Thorarchaeota WOR 83	KXH77685.1	c1u5tA_	appears to be functionally related to snf7	100%	59% (87-242/252)	20%
	Thorarchaeota MP9T 1	PJES01000055.1_6	c3cuqA_	vacuolar-sorting protein snf8	100%	82% (25-245/252)	26%
	Thorarchaeota MP8T 1	PJER01000002.1_56	c3cuqA_	vacuolar-sorting protein snf8	100%	81% (18-244/253)	27%
	Thorarchaeota AB 25	OLS30571.1	c1u5tA_	appears to be functionally related to snf7	100%	83% (25-241/252)	18%
		OLS27487.1	c3cuqA_	vacuolar-sorting protein snf8	100%	27% (539-747/752)	27%
	Thorarchaeota WOR 45	KXH75178.1	c3cuqA_	vacuolar-sorting protein snf8	100%	76% (44-241/255)	27%
	Thorarchaeota MP11T 1	PJET01000114.1_27	c2zmeA_	vacuolar-sorting protein snf8	100%	80% (36-240/252)	26%
ESCRT-II: Vps25-like	*Odinarchaeceae* Tengchong	2724281931	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	68% (71-226/227)	24%
	Odinarchaeceae LCB 4	OLS18191.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	71% (62-223/228)	20%
	Heimdallarchaeota LC 3	OLS27543.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	58% (105-256/260)	22%
	Heimdallarchaeota AB 125	OLS31935.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	65% (80-245/251)	24%
	Lokiarchaeum sp. GC14_75	KKK42120.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	45% (159-293/297)	22%
	Thorarchaeota WOR 83	KXH77686.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	64% (87-242/243)	27%
	Thorarchaeota MP9T 1	PJES01000055.1_5	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	61% (90-240/246)	24%
	Thorarchaeota MP8T 1	PJER01000002.1_57	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	62% (90-238/241)	22%
	Thorarchaeota AB 25	OLS30570.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	62% (89-240/245)	26%
	Thorarchaeota WOR 45	KXH75179.1	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	62% (89-233/235)	23%

	Thorarchaeota MP11T 1	PJET01000114.1_26	c3cuqC_	vacuolar protein-sorting-associated protein 25	100%	62% (89-240/245)	26%
ESCRT-III: Snf7 family	*Odinarchaeceae* Tengchong	2724283318	c5fd7A_	vacuolar-sorting protein snf7	99.9%	58% (14-135/211)	24%
	Odinarchaeceae LCB4	OLS18194.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	57% (14-134/213)	24%
		OLS18193.1	c3frvA_	charged multivesicular body protein 3	92.7%	59% (15-134/203)	14%
	Heimdallarchaeota LC 3	OLS27541.1	c2gd5B_	charged multivesicular body protein 3	99.9%	70% (3-138/194)	18%
		OLS27540.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	53% (16-135/227)	21%
	Heimdallarchaeota LC 2	OLS27395.1	c2gd5B_	charged multivesicular body protein 3	100%	73% (8-156/204)	16%
		OLS27394.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	54% (18-138/223)	23%
	Heimdallarchaeota AB 125	OLS31932.1	c2gd5B_	charged multivesicular body protein 3	99.9%	74% (2-149/199)	14%
		OLS31933.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	55.5% (14-133/218)	28%
	Lokiarchaeum sp. GC14_75	KKK44605.1	c5fd7A_	vacuolar-sorting protein snf7	82.5%	42% (39-129/218)	23%
		KKK42122.1	c2gd5B_	charged multivesicular body protein 3	100%	68% (11-153/209)	13%
	Lokiarchaeota CR 4	OLS16333.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	56% (14-133/215)	22%
		OLS15111.1	c2gd5B_	charged multivesicular body protein 3	100%	67% (11-159/223)	16%
	Thorarchaeota WOR 83	KXH77688.1	c2gd5B_	charged multivesicular body protein 3	100%	70% (8-149/202)	18%
		KXH71213.1	c5fd7A_	vacuolar-sorting protein snf7	99.9%	55% (14-134/221)	18%
	Thorarchaeota MP9T 1	PJES01000065.1_2	c5fd7A_	vacuolar-sorting protein snf7	99.9%	55% (11-131/221)	21%
		PJES01000154.1_42	c5fd7A_	vacuolar-sorting protein snf7	99.8%	51% (1-96/187)	20%
		PJES01000170.1_19	c2gd5B_	charged multivesicular body protein 3	100%	69% (11-152/205)	18%
	Thorarchaeota MP8T 1	PJER01000078.1_3	c5fd7A_	vacuolar-sorting protein snf7	99.9%	55% (11-131/221)	21%
		PJER01000003.1_50	c5fd7A_	vacuolar-sorting protein snf7	99.9%	54% (12-131/222)	20%
		PJER01000002.1_60	c2gd5B_	charged multivesicular body protein 3	100%	71% (7-149/202)	18%
	Thorarchaeota WOR 45	KXH70284.1	c5fd7A_	vacuolar-sorting protein snf7	99.7%	55% (11-131/221)	25%
		KXH75181.1	c2gd5B_	charged multivesicular body protein 3	100%	71% (7-149/202)	17%
	Thorarchaeota MP11T 1	PJET01000114.1_24	c2gd5B_	charged multivesicular body protein 3	100%	76% (10-179/205)	18%
		PJET01000200.1_3	c5fd7A_	vacuolar-sorting protein snf7	99.9%	54% (12-131/221)	23%
	Nitrosopumilus maritimus SCM1	ABX12712.1	c2gd5B_	charged multivesicular body protein 3	100%	67% (20-166/216)	12%
	Nitrosoarchaeum limnia SFB1	EGG41565.1	c2gd5B_	charged multivesicular body protein 3	100%	69% (22-185/213)	15%
		EGG42458.1	c2gd5B_	charged multivesicular body protein 3	100%	70% (18-163/206)	19%
		EGG42703.1	c2gd5B_	charged multivesicular body protein 3	100%	78% (2-173/199)	17%
	Geoarchaeon NAG1	2554139482	c2gd5B_	charged multivesicular body protein 3	99.9%	68% (23-168/214)	16%
	Metallosphaera sedula	WP_012021637.1	c2gd5B_	charged multivesicular body protein 3	100%	67% (26-189/221)	20%
		WP_012021613.1	c2gd5B_	charged multivesicular body protein 3	99.8%	53% (15-158/270)	11%
	Metallosphaera cuprina	WP_013737137.1	c2gd5B_	charged multivesicular body protein 3	100%	69% (26-189/221)	19%
	Sulfolobus tokodaii	WP_010979233.1	c2gd5B_	charged multivesicular body protein 3	99.8%	59% (12-189/186)	17%
	Aeropyrum pernix K1	BAA79054.2	c2gd5B_	charged multivesicular body protein 3	99.9%	65% (26-172/223)	12%

[a] Family information of the template.

[b] Ubiquitin-conjugating enzyme E2 and the steadiness box domains of ESCRT-I are located in the same protein in *Odinarchaeceae Tengchong*.

Figures

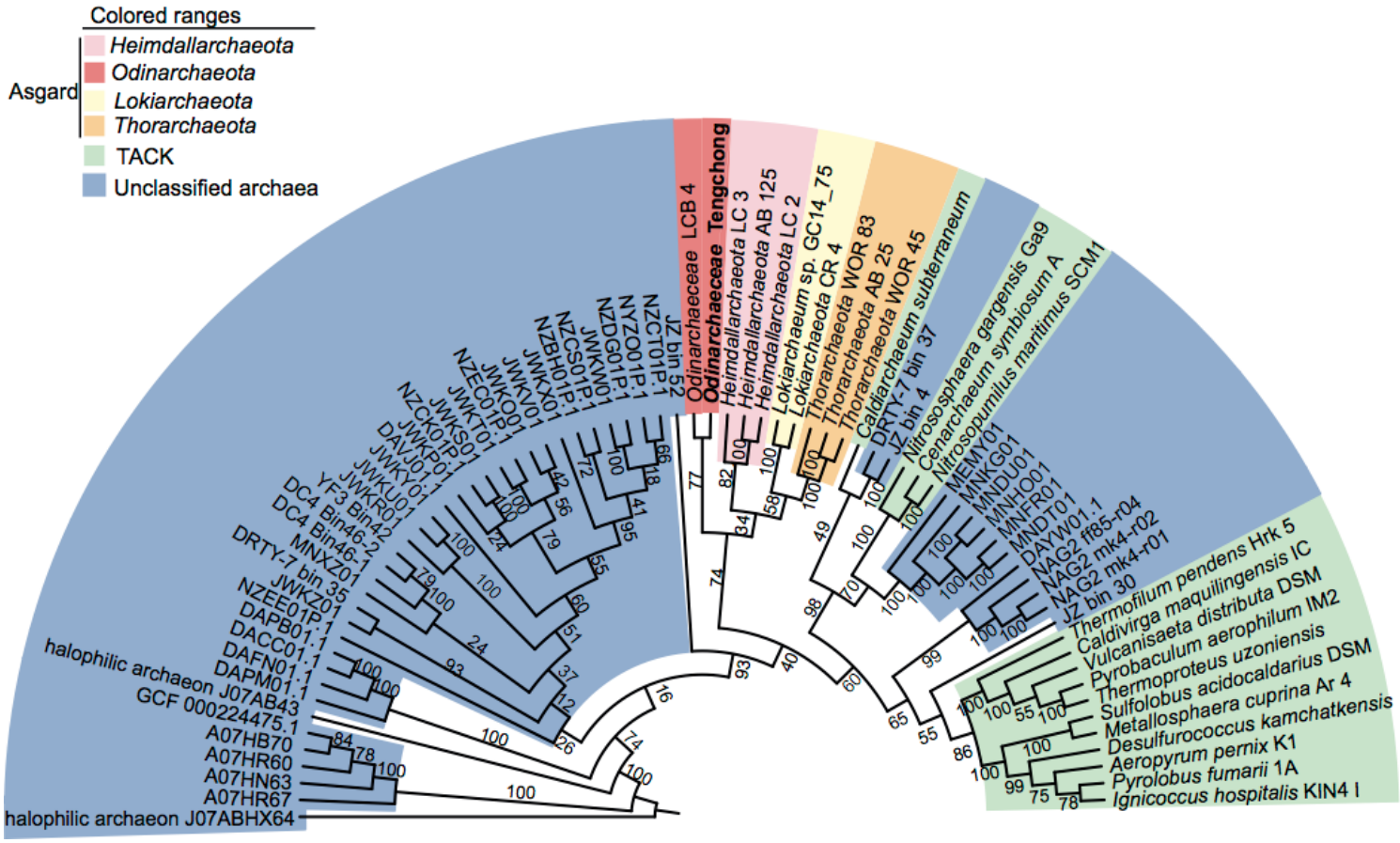

Figure 1

Maximum-likelihood tree of concatenated 15 conserved ribosomal proteins in unclassified archaeal MAGs. The tree is inferred with PROTCATLG model in RAxML. Values at nodes represent fast bootstrap supports.

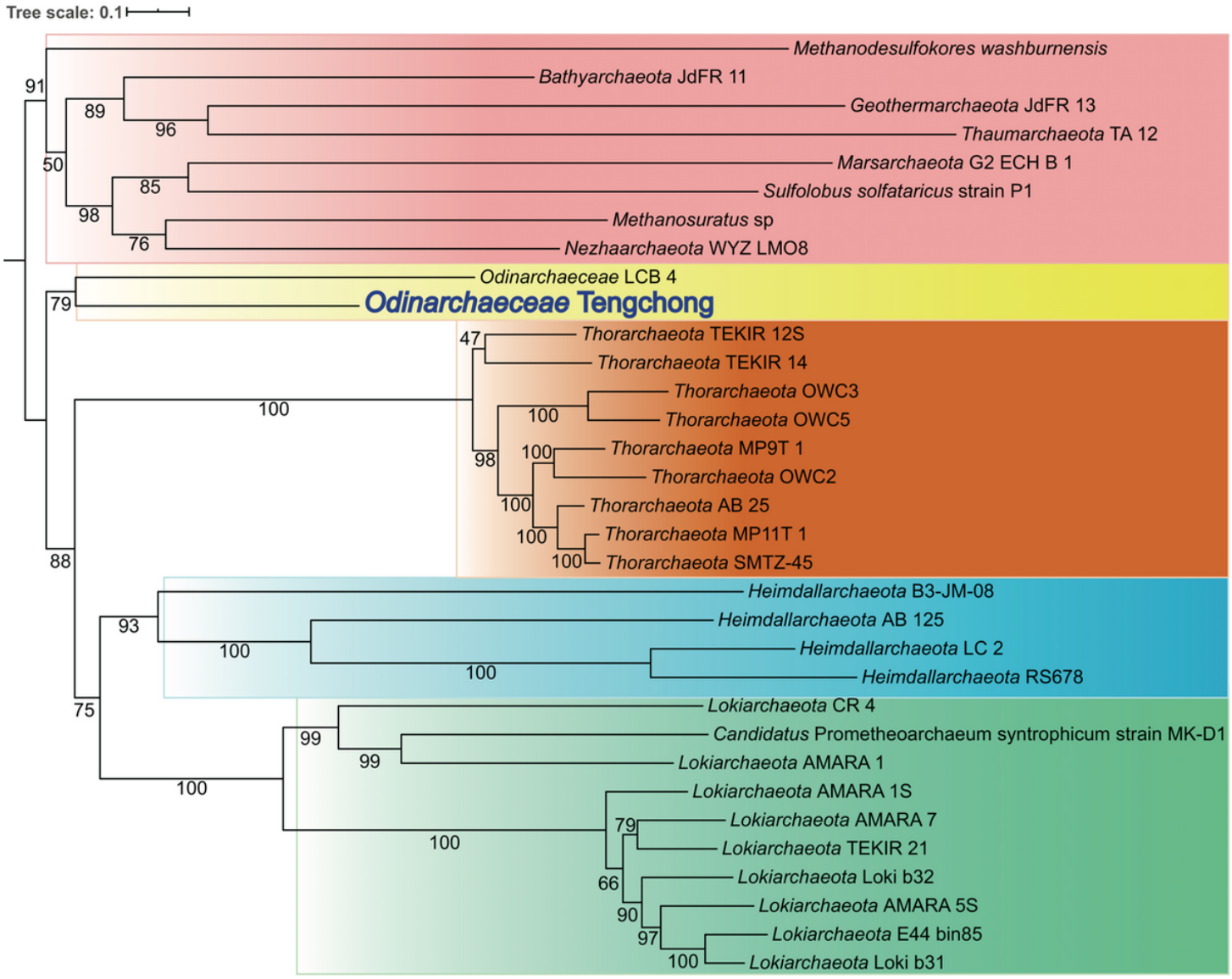

Figure 2

Maximum-likelihood tree of concatenated 55 conserved ribosomal proteins in Odinarchaeceae Tengchong and other qualified Asgard genomes. The tree is inferred with LG+C60+F+G+PMSF model in IQ-TREE. Values at nodes are bootstrap supports. TACK genomes were used as outgroup.

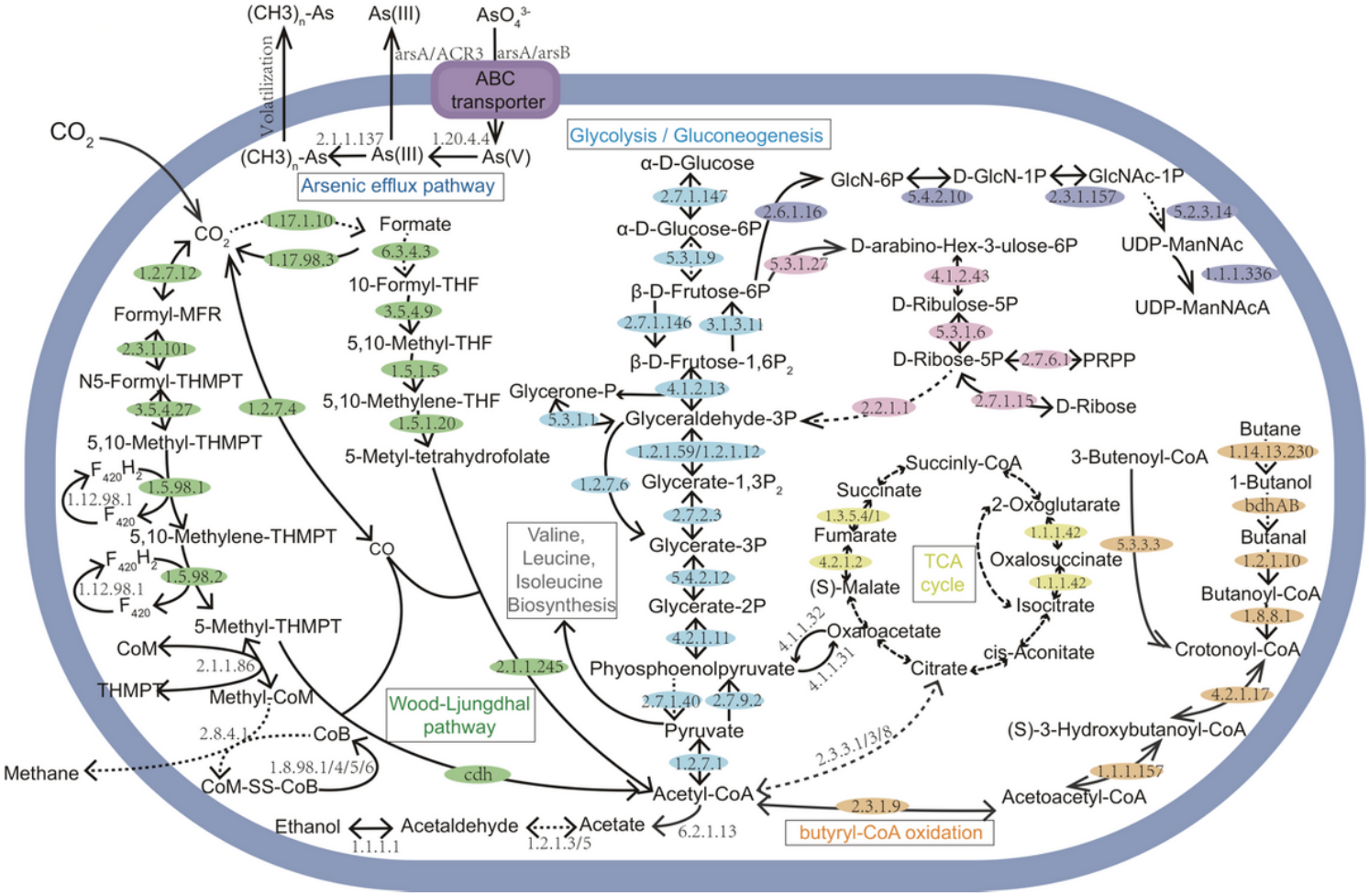

Figure 3

Overview of potential metabolic capabilities of Odinarchaeceae Tengchong. Dash lines refer to genes not identified. Gene annotations are shown in Table S3a.

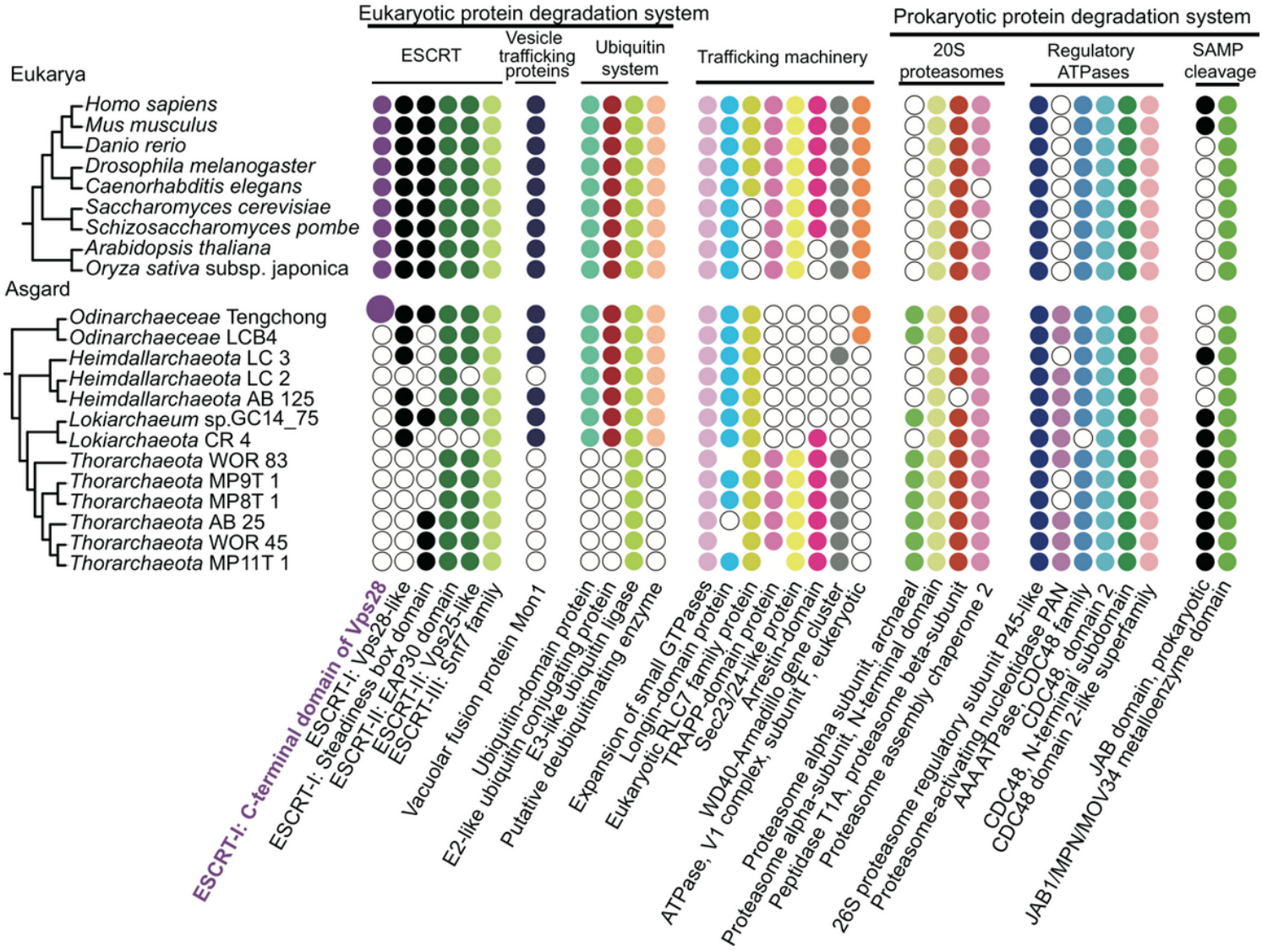

Figure 4

Schematic trees of Asgard archaea and representative eukaryotes (Left panel) and distribution of certain ESPs (Right panel). Filled circles show presence of IPR domains and empty circles show absence. Additional taxa are shown in Fig. S2.

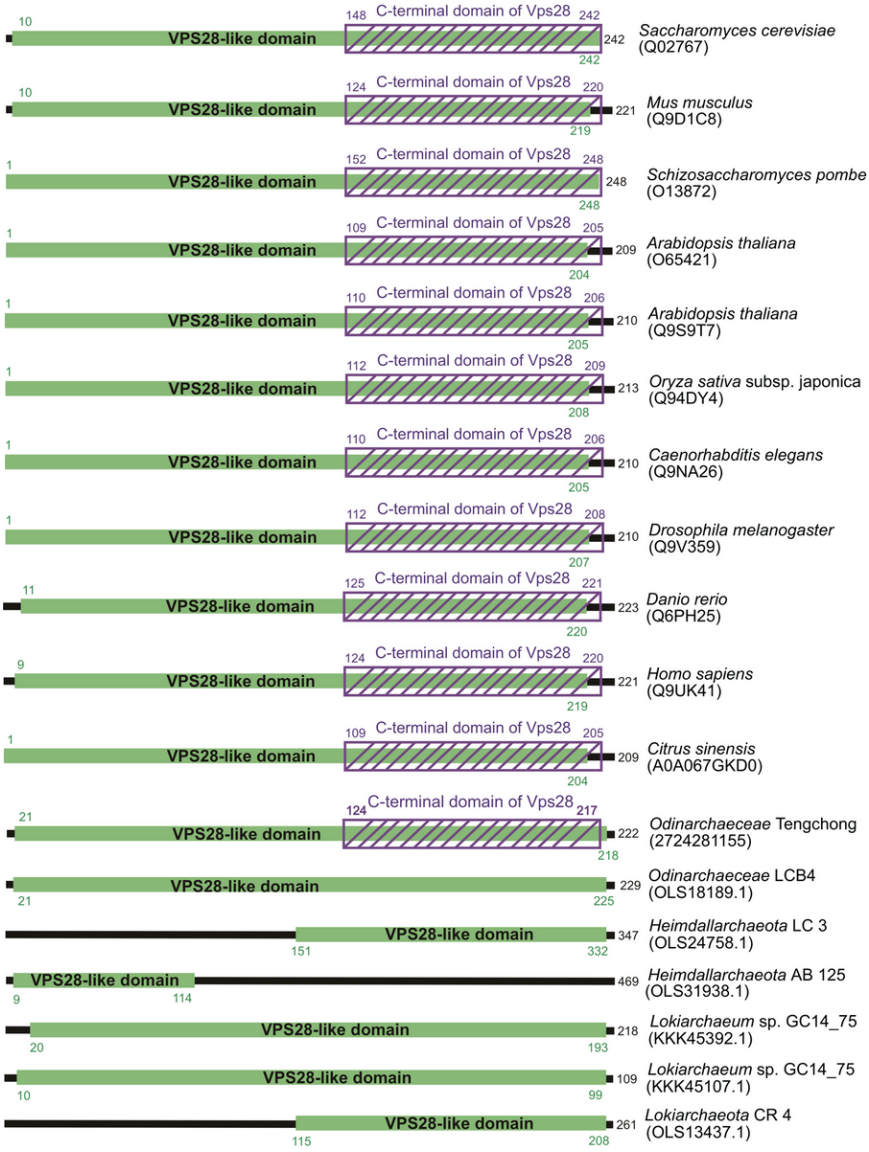

Figure 5

Domain positions of Vps28 and Vps28CTD in genomes of eukaryotes and Asgard archaea.

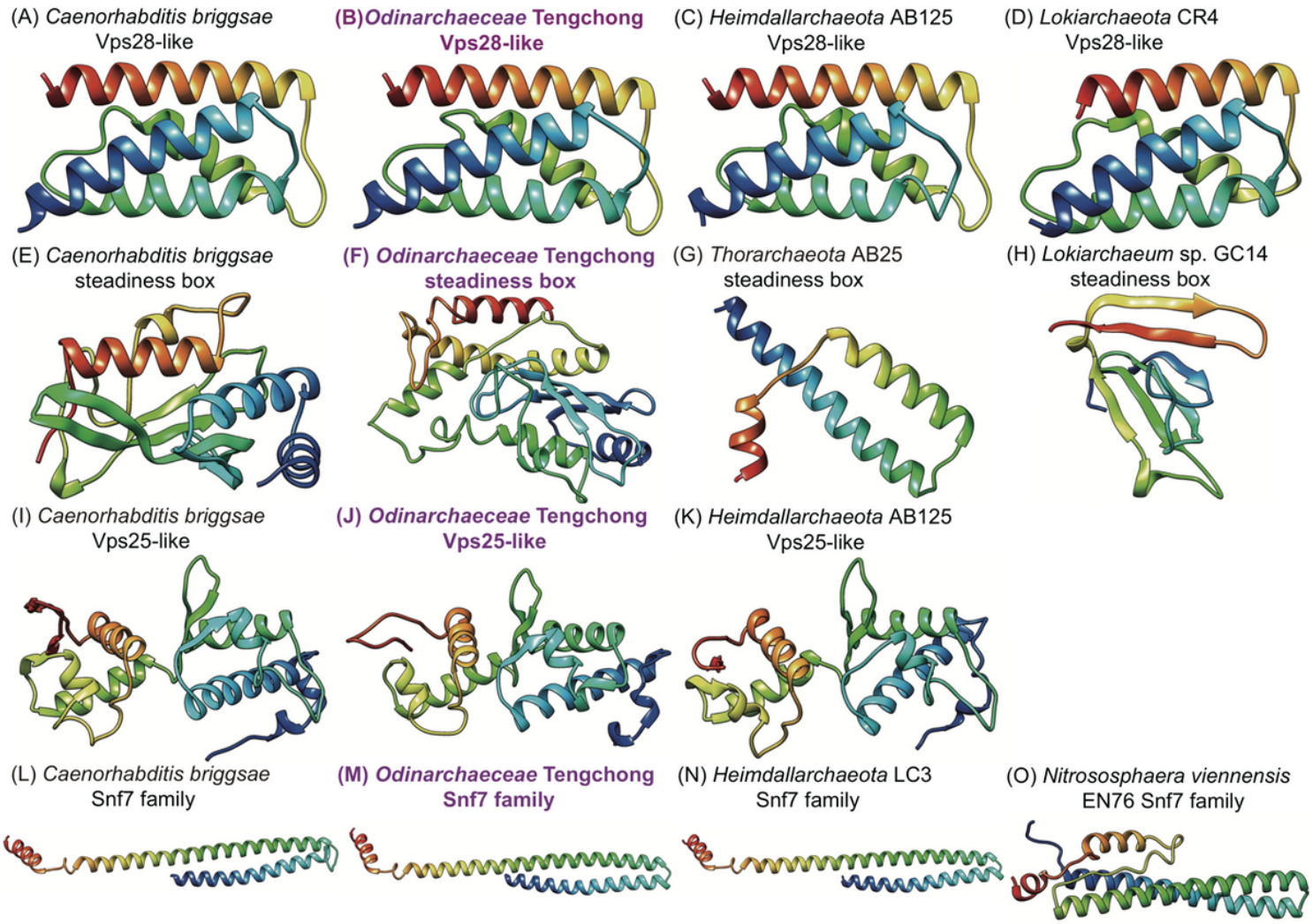

Figure 6

Homology modelling for components of ESCRT machinery. (a)-(d) Vps28-like protein of ESCRT-I; (e)-(h) Steadiness box domain of ESCRT-I; (i)-(k) Vps25-like of ESCRT-II; (l)-(o) Snf7 family protein of ESCRT-III. In each predicted protein structure, N-terminus to C-terminus are indicated by rainbow colours (red to purple). Helix shows alpha helix structure and arrow refers to beta strand structure.

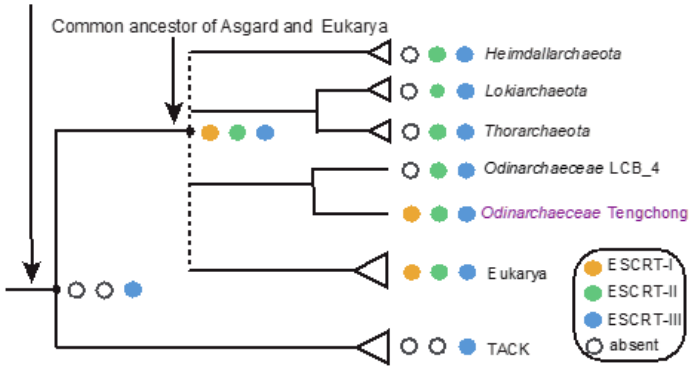

Figure 7

Hypothetical origination of eukaryotic ESCRT components. Empty circles indicate the ESCRT components are incomplete or entirely absent. Dash lines show unresolved phylogenetic relationships.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.xlsx
- Supplementarymaterials.pdf
- Graphicalabstract.eps