Evolutionary Diversification of Methanotrophic *Ca.* Methanophagales (ANME-1) and Their Expansive Virome

Rafael Laso-Pérez[1,*,#,@], Fabai Wu[2,#,@], Antoine Crémière[3], Daan R. Speth[4,$], John S. Magyar[3], Mart Krupovic[5,@], Victoria J. Orphan[1,3,4,@]

1. MARUM, Center for Marine Environmental Science, and Department of Geosciences, University of Bremen, 28359 Bremen, Germany
2. ZJU-Hangzhou Global Scientific and Technological Innovation Center, No. 733 Jianshe San Road, Xiaoshan District, Hangzhou, Zhejiang, China 311200
3. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA 91125
4. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA 91125
5. Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, F-75015 Paris, France

* Current address: Systems Biology Department, Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain

$ Current address: Max-Planck Institute for Marine Microbiology, 28359 Bremen, Germany

# These authors contributed equally to this work

@ For correspondence please address to: Rafael Laso-Pérez (rafael.laso@cnb.csic.es), Fabai Wu (fabaiwu@zju.edu.cn), Mart Krupovic (mart.krupovic@pasteur.fr), Victoria J. Orphan (vorphan@gps.caltech.edu)

Abstract

'Candidatus Methanophagales' (ANME-1) is a major order-level clade of archaea responsible for methane removal in deep-sea sediments through anaerobic oxidation of methane. Yet the extent of their diversity and factors which drive their dynamics and evolution remain poorly understood. Here, by sampling hydrothermal rocks and sediments, we expand their phylogenetic diversity and characterize a new deep-branching, thermophilic ANME-1 family, '*Candidatus* Methanoxibalbaceae' (ANME-1c). They are phylogenetically closest to the short-chain-alkane oxidizers '*Candidatus* Syntrophoarchaeales' and '*Candidatus* Alkanophagales', and encode ancestral features including a methyl coenzyme M reductase chaperone McrD and a hydrogenase complex. Global phylogeny and near-complete genomes clarified that the debated hydrogen metabolism within ANME-1 is an ancient trait that was vertically inherited but differentially lost during lineage diversification. Our expanded genomic and metagenomic sampling allowed the discovery of viruses constituting 3 new orders and 16 new families that so far are exclusive to

ANME-1 hosts. These viruses represent 4 major archaeal virus assemblages, characterized by tailless icosahedral, head-tailed, rod-shaped, and spindle-shaped virions, but display unique structural and replicative signatures. Exemplified by the analyses of thymidylate synthases that unveiled a virus-mediated ancestral process of host gene displacement, this expansive ANME-1 virome carries a large gene repertoire that can influence their hosts across different timescales. Our study thus puts forth an emerging evolutionary continuum between anaerobic methane and short-chain-alkane oxidizers and opens doors for exploring the impacts of viruses on the dynamics and evolution of the anaerobic methane-driven ecosystems.

Introduction

Anaerobic methanotrophic archaea (ANME) is an assemblage of several archaeal clades capable of anaerobic oxidation of methane (AOM), a process that is estimated to globally remove more than 80% of the methane produced in deep-sea sediments[1-4]. Whereas the ANME-2 and ANME-3 lineages share common ancestors with the present-day methanogens of the *Methanosarcinales* order, members of ANME-1 form their own order ´*Candidatus* Methanophagales´ and are a sister group to the non-methane alkane degraders ´*Candidatus* Syntrophoarchaeales´ and ´*Candidatus* Alkanophagales´[5]. Phylogenomic analysis of the ANME clades indicates that they have independently evolved the ability to oxidize methane by reversing the methanogenesis pathway, including the key reaction catalyzed by methyl coenzyme M reductase (MCR), responsible for the activation of methane in ANME[6]. Distinctly, the majority of ANME-1 cells exhibit a segmented rod-shape morphology that is different from that of other ANMEs; they can grow beyond the cold and temperate deep-sea habitats that they often share with other ANMEs, uniquely thriving at higher temperatures within hydrothermal environments[6-9]. However, it remains largely unclear what factors have contributed to the physiological and ecological diversification of ANME-1 from their short-chain-alkane relatives and other ANME lineages.

In marine sediments, ANME archaea appear mostly in syntrophic association with sulfate-reducing bacteria (SRB)[10]. In these consortia, ANME cells oxidize methane and then transfer the reduced equivalents produced during the oxidation to the sulfate-reducing bacteria using direct interspecies electron transfer (DIET)[11,12]. Additionally, ANME-1 cells have been observed as single cells or as monospecific consortia without partner bacteria[9,13-18]. Some of these single cells have been detected in methanogenic horizons and have been interpreted as evidence that some ANME-1 might be able to perform methanogenesis or switch between methanotrophic/methanogenic lifestyles thanks to their (reverse) methanogenesis pathway[18,19]. Hydrogen was suggested as the electron donor for this proposed methanogenic metabolism[19], but cultivation and microcosm experiments have failed to support this hypothesis thus far[12,20], and ANME-1 genomes largely lack genes for hydrogenases[12,21-23], except in a few instances[6,24,25].

Despite the fact that ANME archaea dominate many methane-rich ecosystems and are frequently the dominant microorganisms in methane seeps[10], viruses targeting ANME lineages are largely unexplored[26-28]. By exploiting and spilling host cellular resources through their replication and lytic cycles, viruses play a major role in the ecological dynamics and nutrient cycling in diverse microbial systems[29]. In deep-sea ecosystems, viral lysis has been estimated to cause annual archaeal mortality that releases up to ~0.3 to 0.5 gigatons of carbon globally[30]. It can thus be inferred that discovering viruses targeting ANME is a critical step in constructing a comprehensive view of elemental and energy flows in deep-sea methane-driven ecosystems. Viruses affect host evolution in many ways: their threat to host survival and replicative health forces the hosts to develop strategies for viral evasion, triggering an evolutionary arms race[31]; certain viruses are known to encode diverse metabolic functions that augment host capabilities and

increase their fitness compared to uninfected populations[32]; viral genomes can recombine with host genomes and shuffle genes across hosts, promoting horizontal gene transfer[33]. Currently, we are lacking the knowledge not only about the viruses infecting ANME-1 archaea, but also on all other types of mobile genetic elements (MGEs), including plasmids, transposons, defense islands and synthetic gene clusters potentially present in ANME genomes. These MGEs often encode unique functions that can confer a competitive advantage to their hosts, as has been documented with antibiotic resistance and antiviral defense[34,35]. Characterizing the distributions and functions of viral and non-viral MGEs of ANMEs is thus one of the most important tasks for linking ANME physiology to their environmental impact.

Results

Expanded ANME-1 diversity reveals a new deep-branching clade in hydrothermal vents

In this study, we substantially expanded the ANME-1 genome diversity recovered from metagenomes of native and laboratory-incubated hydrothermal mineral samples from the South Pescadero Basin in the Gulf of California, Mexico, a recently discovered hydrothermal system with vent fluids enriched in methane (up to 81 mmol kg^{-1})[36]. Thirteen metagenome-assembled genomes (MAGs) affiliated with ANME-1 were recovered in our samples (Supplementary Tables 1 and 2). Phylogenomic analysis indicated that these MAGs not only expanded the known diversity within the previously described ANME-1a clade, in particular the group represented by ANME-1 G60[10], but also contained several representatives of a deep branching clade phylogenetically positioned at the base of the ANME-1 order (Fig.1, Fig. S1, Supplementary Tables 2 and 3). This clade corresponds to the group of 16S rRNA gene sequences previously reported as "ANME-1 Guaymas"[18]. Our phylogenomic analysis affiliated 8 genomes to this clade: a circular scaffold and 5 MAGs from South Pescadero Basin from this study and two recently reported MAGs, B22_G9 from Guaymas Basin[37], and PB_MBMC_218 from Pescadero Basin[38] (Supplementary Table 2). These draft genomes form a new ANME-1 family, ANME-1c, which we now name ´*Candidatus* Methanoxibalbaceae´. Notably, the ANME-1c clade is the closest ANME-1 group to the sister archaeal orders of '*Ca.* Alkanophagales' and '*Ca.* Syntrophoarchaeales', which anaerobically degrade alkanes larger than methane in syntrophy with sulfate-reducing bacteria[12].

Our ANME-1c MAGs represent two different genera within the same family with an average nucleotide identity (ANI) of 76%, which we call ´*Candidatus* Methanoxibalbensis´ and ´*Candidatus* Methanospirare´, respectively, represented by species ´*Ca.* Methanoxibalbensis ujae´ (species 1) and ´*Ca.* Methanospirare jalkutatii´ (species 2, see methods for species description). According to genome coverage, the ANME-1c represented by these two species were the most abundant organisms in two of the rock samples (12019 and NA091.008), while they are close to the detection limit in other analyzed rocks (11868 and 11719) and the sediment cores (Fig.2a). ´*Ca.* M. jalkutatii´ was dominant in the 12019 sample (17% of the total prokaryotic community), while in the NA091.008 sample, ´*Ca.* M. jalkutatii´ had a similar genomic relative abundance to ´*Ca.* M. ujae´ (both 17%; Fig.2a).

The environmental distribution of ANME-1c from prior 16S rRNA surveys and our metagenomic analysis and laboratory incubations described herein, is suggestive of a thermophilic lifestyle. All ANME-1c MAGs and 16S rRNA gene sequences from the NCBI and SILVA databases have originated from hydrothermal environments, specifically the sediments of Guaymas and South Pescadero Basins. These hydrothermal vent systems are separated by approximately 400 km along the same fault system in the Gulf of California and exhibit 20% species-level overlap in the microbial community[38]. This distribution suggests a strong physiological specialization of ANME-1c to such environments. Indeed, genome-based prediction[39] suggested a high theoretical optimal growth temperature (OGT, Supplementary Table 4) for both ANME-1c species (>70 °C) that was higher than the OGT for both ANME-1a (62 °C) and ANME-1b (52 °C). Interestingly, the ANME-1c species had reduced genome sizes (´*Ca.* M. ujae´: 1.81 Mb; ´*Ca.* M. jalkutatii´: 1.62 Mb) compared to the estimated genome size of other ANME-1 groups (ANME-1a: 2.35

Mb; ANME-1b: 2.46 Mb; Supplementary Table 4 and Fig. S2). This trend is in line with previously observed negative correlation between genome size and growth temperature in thermophilic bacteria and archaea[40].

Using fluorescence in situ hybridization (FISH) with an ANME-1-targeted 16S rRNA probe, we detected ANME-1 cells in rock NA091.008 (Fig.2b), where ANME-1c are the dominant group according to genome coverage (Fig. 2a). These putative ANME-1c cells exhibit the typical cylindrical shape previously reported for other ANME-1 populations[10] and were loosely associated with bacterial cells in an uncharacterized EPS matrix. We also observed cells of ANME-1c outside of the biofilm which were found as single cells consistent with previous reports of other ANME-1 organisms[9,13-18].

Physiological differentiation of diverse ANME-1 archaea

The existence of the deep-branching ANME-1c, which is phylogenetically positioned closest to the sister archaeal orders of '*Ca.* Alkanophagales' and '*Ca.* Syntrophoarchaeales', allowed us to examine the genomic patterns reflective of the emergence and differentiation of ANME-1. Like all ANME organisms, ANME-1c encode for a complete reverse methanogenesis pathway including a single operon for the methyl coenzyme M reductase enzyme (MCR) and the replacement of F_{420}-dependent methylene-H_4MPT reductase (Mer) by 5,10-methylenetetrahydrofolate reductase (Met) characteristic for the ANME-1 order [21,23,41]. Similar to other ANME clades, ANME-1c contain several genes encoding for multiheme cytochromes (MHC), which are proposed to mediate the transfer of electrons during AOM to their syntrophic sulfate-reducing partner[11,23]. Each ANME-1c MAG encodes between 4 to 11 MHC with 4 to 9 heme motifs.

Strikingly, ANME-1c exhibit distinct features in the reverse methanogenesis pathway compared to their sister clades ANME-1a and ANME-1b. The MCR enzyme consists of six subunits with the structure $\alpha_2\beta_2\gamma_2$; and its activity depends on a unique cofactor known as coenzyme F_{430}, which contains a nickel atom[42]. In most methanogens, the *mcr* genes appear in an operon with two additional genes: *mcrC* and *mcrD*[43]. McrC has been identified as a component that activates the MCR complex by reducing the nickel atom to the Ni^{1+} form[44], while McrD was suggested to act as a chaperone to deliver the F_{430} into the MCR [45]. Genes encoding both subunits were present in ANME-1c; but only *mcrD* forms an operon with *mcrABG*. Surprisingly, *mcrD* genes are not present in any other ANME-1 group (Fig.1). Previous phylogenetic analysis of the *mcrA* of ANME-1 have shown that ANME-1 likely acquired the genes encoding the MCR enzyme from distant H_2-dependent methylotrophic methanogens of the class *Methanofastidiosa* [6], while they lost the divergent MCRs present in *Syntrophoarchaeales* and *Alkanophagales*, which seem to use larger alkanes. Likewise, phylogenetic analysis of the ANME-1c McrD (Fig. S3) shows that the ANME-1c McrD is closely related to the McrD of *Methanofastidiosa* and only distantly related to the McrD of *Syntrophoarchaeales* and *Alkanophagales* that form a different cluster with the McrD of other organisms with divergent MCRs. This analysis suggests that during the emergence of ANME-1, all alkanotrophic *mcr* genes of *Syntrophoarchaeales* and *Alkanophagales* were lost (including the *mcrD*), while a whole operon of methane-cycling *mcr* (including *mcrD*) was acquired by horizontal gene transfer from a methylotrophic methanogen, presumably related to the *Methanofastidiosa*. However, the *mcrD* was later lost in both ANME-1a and ANME-1b clades. Interestingly, the MCR protein of an ANME-1a population isolated from Black Sea mats was previously reported to harbor a methylthionated version of the coenzyme F_{430} compared to other MCRs in ANMEs[46,47]. Future studies should investigate if this modification is related to the loss of the *mcrD* gene.

Additionally, ANME-1c genomes exhibit other distinct features in relation to the amino acid metabolism (Fig. 1, Supplementary Text and Supplementary Table 5). The ANME-1c appear to lack the *proA* and *proB* genes involved in proline synthesis in other ANME-1, along with the *csdA* gene, an L-cysteine desulfurase, that is proposed to be involved in alanine synthesis from L-cysteine[6] and in the biosynthesis of Fe-S centers[48]. In ANME-1c, *csdA* might be replaced by a cysteine desulfidase (*cdsB*), which

was suggested to be involved in the biosynthesis of Fe-S centers in the archaeon *Methanocaldococcus jannaschii*[49]. Various physiological and structural features have an unexpected mosaic distribution across the diverse ANME-1 lineages (Fig.1). They include 1) the biosynthesis pathway of the coenzyme M (CoM), essential for the reverse methanogenesis pathway, 2) chemotaxis systems, including type IV pili, and the archaeal flagellum – archaellum[50-52], and 3) the nitrogenase genes *nifI* and *nifK*[53] (Fig. 1, see supplementary text for detailed descriptions). It was particularly intriguing to observe the presence of genes coding for an archaellum and an elaborate chemotaxis system in some of the MAGs, including most of the ANME-1c. In endolithic habitats inhabited by ANME-1c, the archaellum might give a selective advantage in different ways: conferring cellular motility[51], as attachment element[54,55] or acting as a conductive nano-wire as has been shown in other archaea[56,57]. In porous or fractured rock matrices, these components might allow ANME-1c to search for the optimal environmental conditions within their heterogeneous and fluctuating habitats. Notably, Type IV pili, which function as receptors for diverse archaeal viruses[58,59], could mediate adhesion of ANME-1c to different surfaces.

Shared origin and differential loss of hydrogenases

Hydrogen was one of the first proposed intermediates for syntrophic AOM, but this hypothesis was disregarded as the majority of ANME genomes, including ANME-1, do not encode for hydrogenases. However, recent studies have reported hydrogenases in genomes from an ANME-1b subclade, '*Candidatus* Methanoalium' as well as from some ANME-1a genomes (Fig.1)[6,24,25], where it is possible that they are mediating methane formation instead of AOM. Interestingly, the genomes of the sister orders '*Ca.* Syntrophoarchaeales' and '*Ca.* Alkanophagales' contain genes for a NiFe hydrogenase (Fig. 1)[12]. In '*Ca.* Syntrophoarchaeales', these genes are highly expressed during butane oxidation coupled to sulfate reduction in syntrophic bacteria, but physiological experiments showed that mere hydrogen production was not sufficient to consider hydrogen as a syntrophic intermediate[12]. The significant expansion of ANME-1 diversity of our study allowed us to resolve the evolutionary trajectory of hydrogenases across ANME-1 lineages. Our analyses revealed three different subclades of ANME-1 genomes with an operon encoding a NiFe hydrogenase and the corresponding maturation factors (Fig.1). They correspond to the ANME-1c, a subclade of ANME-1a, and a subclade of ANME-1b. A phylogenetic analysis of the large subunit of these NiFe hydrogenases showed that most ANME-1 hydrogenases, including those of ANME-1c, form a monophyletic group with the hydrogenases of *Syntrophoarchaeales* and *Alkanophagales* (Fig.3a, Supplementary Table 6). Hence, hydrogenase is a common ancient trait of class *Syntrophoarchaeia* that was vertically inherited by the common ancestor of ANME-1 and later differentially lost during ANME-1 clade diversification.

Given the apparent mosaic distribution of hydrogenases across ANME-1 lineages, we further detailed patterns of hydrogenase occurrence within the currently available genomes of ANME-1c. All MAGs of ´*Ca.* M. ujae´ and two out of five in ´*Ca.* M. jalkutatii´ (FW4382_bin126 and NA091.008_bin1) contained the hydrogenase operon, whereas the ´*Ca.* M. jalkutatii´ MAG FWG175, the most contiguous genome that was assembled into a single scaffold, does not contain hydrogenases. To confirm that the presence of hydrogenase genes in ´*Ca.* M. jalkutatii´ is different between MAGs, we mapped the metagenomic reads from our full South Pescadero Basin sample set to the MAGs. This analysis revealed that samples where ANME-1c MAGs did not have hydrogenase genes indeed did not have reads mapping the hydrogenase genes of MAGs FW4382_bin126 and NA091.008_bin1 (Fig. 3b). Additionally, the local absence of hydrogenase genes in FWG175 was confirmed in a genome-to-genome alignment (Fig. S4). Hydrogenase genes thus appear to be a part of the pangenomic repertoire of ´*Ca.* M. jalkutatii´. Since the presence of the hydrogenase operon varies even between subspecies (as demonstrated with ´*Ca.* M. jalkutatii´), hydrogenases might have been preserved in the ANME-1 pangenome as an environmental adaptation rather than as an absolute requirement for their methanotrophic core energy metabolism.

The presence of hydrogenases in ANME-1 genomes raises the question about their metabolic role. Most ANME-1 hydrogenases are closely related to the hydrogenase groups NiFe 1g and 1h (Fig. 3a; only a few affiliated to NiFe Group 3 and 4, Fig. S5). The NiFe 1g group includes several crenarchaeal hydrogenases that consume hydrogen under anaerobic conditions, presumably to channel electrons for sulfur respiration, while the 1h group represents actinobacterial enzymes that use hydrogen for aerobic respiration[60,61,62,63]. Therefore, the role of these hydrogenases could be to mediate hydrogenotrophic methanogenesis in ANME-1, as previously proposed based on biochemical[64], environmental[18,19], and metagenomic data[24] even though attempts to culture ANME under methanogenic conditions have been unsuccessful[12,20]. A methanogenic metabolism was recently proposed for the hydrogenase-encoding ANME-1b group '*Ca.* Methanoalium' based on additional unique features including genes encoding an Rnf complex and a cytochrome b, while genes for multiheme cytochromes, necessary for the electron cycling in AOM, were missing[6]. By contrast, hydrogenase-encoding ANME-1c does not appear to have these features and does contain genes encoding multiheme cytochromes. Alternatively, hydrogen could be produced as a metabolic intermediate during AOM, but themodynamic models are inconsistent with hydrogen as the sole intermediate[11,20,65]. Instead, hydrogen could be produced in the context of a mixed model involving direct electron transfer and metabolite exchange as proposed recently for syntrophic AOM[6].

CRISPR-based discovery of an expansive ANME-1 mobilome

ANME-1 genomes recovered in this study contained various CRISPR-Cas loci, enabling the analysis of ANME-1-hosted MGEs through CRISPR spacer-based sequence mapping[66-69] with additional stringent filters (see Methods). These CRISPR repeats were frequently found to be directly associated with Type IB and Type III *cas* gene operons (see Fig. S6a for examples), typical in archaea. A 95% sequence identity cutoff indicates that the CRISPR repeats are shared by different ANME-1 lineages, yet different from the CRISPR repeats found in the '*Ca.* Alkanophagales', a sister group to the ANME-1 archaea. Surveying previously published and our newly assembled metagenomes from two hydrothermal vent systems in the Gulf of California, South Pescadero Basin[38,69] and this study (22 assemblies) and Guaymas Basin ([37], 13 assemblies, Supplementary Table 7), led to the extraction of 20649 unique ANME-1 CRISPR spacers. Due to the apparent overlap of CRISPR repeats across diverse ANME-1 lineages, these spacers, and thus the host-MGE interactions, were not further assigned taxonomically to specific ANME-1 subclades.

Mapping these spacers to metagenomic assemblies from South Pescadero and Guaymas Basins, as well as the metagenome-derived virus database IMG/VR v.3[70] captured 79, 70, and 86 MGE contigs larger than 10 kb, respectively, totaling 235 ANME-1 MGEs (Fig. S6b). These contigs were up to 80 kb in size and contained up to 532 unique protospacers (Fig.S6c). As shown in Fig. 4a, the ANME-1 MGEs from South Pescadero Basin were primarily targeted by spacers found locally (n=1912), but also showed a high number of matches to Guaymas Basin spacers (n=894). As previously found for the Asgard archaeal mobilome[69], the apparent frequency of cross-site spacer-mobilome mapping indicates a significant fraction of the ANME-1 mobilome has migrated across these sediment-hosted hydrothermal vent ecosystems, along with their hosts[38].

To examine the relationship between these ANME-1 MGEs and currently described viruses, we conducted gene similarity network analyses using vCONTACT2[71] that included the above dataset, the RefSeq202 database, and the recently reported head-tailed viruses infecting haloarchaea and methanogens[72]. The resulting network indicated that all MGEs identified in this study are distant from all other known viruses, without a single gene-sharing signature under the vCONTACT2 criteria (Fig. S7, Supplementary Table 8). While lacking representation in known viral databases, a large fraction of these 228 ANME-1 MGEs were found to be interconnected, forming one large complex network of 185 nodes and a medium-sized network of 28 nodes (Fig.4b). The remaining 22 MGEs fell into 7 small groups of 2-3 nodes, and 7 singletons. The singletons were removed from further analyses.

Based on the conservation of signature genes encoding viral structural proteins, we concluded that ANME-1 MGEs encompass double-stranded DNA viruses belonging to at least 4 widely different virus assemblages characterized by different evolutionary histories and distinct virion morphologies. In particular, head-tailed viruses of the class *Caudoviricetes* (realm *Duplodnaviria*) encode characteristic HK97-fold major capsid proteins (MCP) as well as the large subunit of the terminase and portal proteins[73,74]; tailless icosahedral viruses of the realm *Varidnaviria* are characterized by double jelly-roll MCPs [73,75]; viruses of the realm *Adnaviria* encode unique α-helical MCPs which form claw-like dimers that wrap around the viral DNA forming a helical, rod-shaped capsid[76-78]; and all spindle-shaped viruses (realm yet unassigned) encode unique, highly hydrophobic α-helical MCPs[79,80] (Supplementary Table 9-10). With the exception of adnaviruses, all other viral types associated with ANME-1 appear to be highly diverse, each comprising several new families. In total, 16 new candidate viral families were discovered in this study, including five families with representative complete genomes (Fig. 4c). We named these candidate virus families after Mayan gods, owing to their discovery in the Gulf of California hydrothermal vents off the coast of Mexico.

Tailless icosahedral ANME-1 viruses with previously undescribed major capsid proteins

Tailless icosahedral viruses (*Varidnaviria*) infecting ANME-1 are well distinguished from known viruses, with all 32 representatives unique to this study. They form three disconnected modules, and based on gene similarity analysis, represent three new viral families (Fig. 4c), which lack overlap in their proteomes (amino acid identity cutoff of 30% (Fig. 4d-f)). Members of the 'Huracanviridae' encode single jelly-roll (SJR) MCPs related to those conserved in the kingdom *Helvetiavirae*, whereas 'Chaacviridae' (after Chaac, the god of death in the Mayan mythology) and 'Ixchelviridae' do not encode MCPs recognizably similar at the sequence level to the MCPs of other known viruses. However, structural modeling of the candidate proteins conserved in 'Chaacviridae' and 'Ixchelviridae' using AlphaFold2[81] and RoseTTAFold[5] revealed the identity of the MCPs with a double jelly-roll (DJR) fold (Fig.4g).

Phylogenetic analysis revealed that these DJR MCPs form three distinct families, MCP-1-3 (Fig.4g,h). Some structural variations in the C-terminal jelly-roll domain of the ANME-1 virus DJR MCPs are apparent when compared with the minimal DJR fold present in the previously described MCP of the bacteriophage PM2 (family *Corticoviridae*)[82]. In MCP-1, the beta-strands in the C-terminal jelly-roll domain are considerably longer than in the N-terminal domain, whereas in MCP-2 and MCP-3, an additional small beta-barrel is inserted after the alpha-helix between jelly-roll beta-strands F' and G'. The location of these additional structural elements suggests that they will be pointing outwards from the capsid surface and are likely to be at the interface of virus-host interaction. The three DJR MCP clades have significant sequence divergence, which are reflected by the long branches in the phylogenetic tree (Fig. 4h).

'Chaacviridae' have linear dsDNA genomes with inverted terminal repeats (ITR) and, accordingly, encode protein-primed family B DNA polymerases (pPolB). Chaaviruses display a remarkable genome plasticity – not only do these viruses encode two different variants of the DJR MCPs, MCP-1 and MCP-2, but also their pPolBs belong to two widely distinct clades. Notably, the two MCP and pPolB variants do not strictly coincide, so that viruses with MCP-1 can encode either pPolB-1 or pPolB-2, suggesting multiple cases of recombination and gene replacement within the replicative and morphogenetic modules. Maximum likelihood analysis of these divergent groups of pPolB sequences revealed relatedness to two separate clades of pPolBs encoded by Wyrdviruses, spindle-shaped viruses that target Asgard archaea[83,84] (Fig. 4i). pPolB is not found in 'Ixchelviridae', 'Huracanviridae', or any other of the ANME-1-associated viruses described in this study. Notably, pPolB is not the only gene shared between chaaviruses and Asgard archaeal viruses. Upstream of the MCP gene, all chaaviruses encode a functionally uncharacterized protein with homologs in Asgard archaeal viruses of the Huginnvirus group, where the latter gene occupies an equivalent position with respect to the DJR MCP gene[85,86]. This observation suggests a remarkable evolutionary entanglement between these ANME-1 and Asgard archaeal viruses, potentially

facilitated by the ecological (i.e., deep-sea ecosystems) rather than evolutionary proximity of the respective hosts.

Based on the gene complement, chaacviruses resemble members of the class *Tectiliviricetes*, but are not closely related to any of the existing virus families. Thus, we propose placing 'Chaacviridae' into a new monotypic order 'Coyopavirales' (after Coyopa, the god of thunder in Mayan mythology) within the existing class *Tectiliviricetes*.

Complex ANME-1 viruses with unique structural and replicative features

The head-tailed viruses targeting ANME-1 encode the typical morphogenetic toolkit shared between all bacterial and archaeal members of the *Caudoviricetes*, including the HK97-fold MCP, portal protein, large subunit of the terminase as well as various tail proteins[72,78]. In agreement with previous analyses[72], our MCP phylogenetic tree is generally consistent with the family-level taxonomy of *Caudoviricetes,* with a few exceptions in the families *Graaviviridae* and *Vertoviridae* (Fig. S8). Notably, in the MCP phylogeny, family-level clades of ANME-1 viruses are interspersed with the established families of haloarchaeal viruses, suggesting deeper evolutionary relationships between these virus groups, possibly reflecting the shared ancestry of ANME-1 and Haloarchaea, which connect at the phylum level[87]. To further assess the relationships between ANME-1 and haloarchaeal head-tailed viruses, we carried out a global proteome-based phylogenetic analysis using ViPTree[88], which recapitulated the existing haloarchaeal virus taxonomy. Interestingly, unlike in the MCP phylogeny, global proteomic analysis revealed a clear division between ANME-1 and haloarchaeal head-tailed viruses (Fig. 5a). This result suggests that although these viruses encode related core proteins for virion formation, the overall gene (and protein) contents of ANME-1 and haloarchaeal viruses differ considerably, likely reflecting the adaptation to their respective hosts and ecological contexts. Based on the minimum genetic distances between halovirus families and cross-genome comparisons (Fig. S9), we propose nine new candidate *Caudoviricetes* families. 'Ekchuahviridae' (after Ek Chuah, the patron god of warriors and merchants in the Mayan mythology) and 'Ahpuchviridae' (after Ah Puch, the god of death in the Mayan mythology) are each represented by ANME-1 viruses with complete genomes. Viruses in these new families exhibit little proteome overlap with each other (Fig. S9), further illustrating the vast genetic diversity of ANME-1 head-tailed viruses.

In the proteomic tree, 'Ekchuahviridae' and 'Ahpuchviridae' form sister clades which are most distant from the haloviruses that include 3 orders (Fig. 5a). We thus propose to create a new order 'Nakonvirales' (after Nakon, the most powerful god of war in Mayan mythology) for unification of these two families. Their representative complete genomes were assembled as circular contigs, sized around 70-80 kb, and encode all structural proteins typical of *Caudoviricetes*. The South Pescadero Basin ahpuchviruses PBV299 (70.9 kb, complete, Fig. 5b) and IMGVR0573778 (74.8 kb, likely near complete) each encode one copy of MCP, while the two ekchuahviruses, IMGVR0083622 (80.6 kb, complete, Fig. 5c) and IMGVR0540589 (71.8 kb, complete), each encode two MCP copies. This is unique among other known *Caudoviricetes* targeting haloarchaea and ANME-1. We can exclude an assembly artifact, as the initial assemblies of the two ekchuahviruses were found to have a circular alignment with each other (Fig. 5d). Both MCP genes are accompanied by cognate capsid maturation protease genes, whereas all other virion morphogenetic proteins are encoded as single copy genes (Fig. 5c). Maximum likelihood analyses indicate that the two MCPs of ekchuahviruses could have distinct evolutionary histories (Fig. S8). One of the copies, MCP-1, forms a sister clade to the MCPs of 'Ahpuchviridae', mirroring the relationship between the two families based on the global proteomic analysis (Fig. 5a), and hence likely represents the ancestral copy of the 'Ekchuahviridae' MCP. However, the second copy, MCP-2, forms a sister clade to the MCPs of haloarchaeal viruses in the *Haloferuviridae* family, with the two clades collectively branching next to the clade including ekchuahvirus MCP-1 and ahpuchvirus MCP. It is currently unclear whether these two clades of MCPs originated from an ancient duplication within 'Ekchuahviridae' or represent an ancient incoming horizontal MCP gene transfer from haloferuviruses. Nevertheless, the large phylogenetic

distances between these MCP clades suggest a long co-existence and co-evolution of the two MCPs in ekchuahviruses, likely conferring a selective advantage.

The coexistence of two divergent MCP genes is also found in members of putative rod-shaped viruses within the new family 'Ahmunviridae' (after Ah Mun, the god of agriculture in Mayan mythology), which we propose including into the class *Tokiviricetes* (realm *Adnaviria*) within a new monotypic order 'Maximonvirales' (after Maximon, a god of travelers, merchants, mecidine men/women, mischief and fertility in Mayan mythology, Fig. 5e), and viruses with predicted spindle-shaped morphology, the 'Itzamnaviridae' (after Itzamna, lord of the heavens as well as night and day in the Mayan mythology, Fig. 5f, 5g). These two new clades of viruses are respectively represented by complete linear genomes with inverted terminal repeats and circular genomes. This is in contrast to another spindle shaped ANME-1 virus, the tepeuvirus PBV144 which has the largest genome (72.6 kb, not yet circularized) but only one MCP.

The coexistence of divergent MCPs is rare among *Caudoviricetes*, but has been previously documented. For example, the head-tailed T4 phage encode two structurally similar homologs, with one forming hexameric capsomers and the other pentameric capsomers which occupy the 5-fold icosahedral vertices[89]. Rod-shaped *Lipothrixviridae*, *Tristromaviridae* and *Ungulaviridae* all encode two MCPs forming a functional MCP heterodimer[59,76,77], whereas some rod-shaped viruses of the *Rudiviridae* encode two MCP homologs, but only one copy is used for virion formation[77,78]. At the moment, it is unclear whether both MCP genes have a structural role in ANME-1 viruses.

Pan-virus auxiliary functions and virus-driven ANME-1 evolution

Besides the unique structural features described above, the large genomes of head-tailed and spindle-shaped viruses of ANME-1 exhibit strong clustering of functionally related genes. In particular, one half of the viral genome contains all structural genes, while the other half encodes diverse enzymes involved in DNA synthesis and modification as well as various metabolic pathways and defense (Fig. 5b-d,f,g). Notably, the entire ~20-kb replicative/metabolism module is missing from the circular genomes of demiitzamnaviruses. Cross-genome alignments revealed a larger variation in gene content for the enzymatic arms in both head-tailed and spindle-shaped viruses, frequently in the form of multi-gene cluster insertions (Fig. 5f, Fig. S9). Head-tailed 'Ekchuahviridae' and 'Ahpuchviridae' and spindle-shaped 'Itzamnaviridae' (but not members of the proposed genus 'Demiitzamnavirus') and 'Tepeuviridae' encode RNA-primed family B DNA polymerases, replicative enzymes commonly encoded by dsDNA viruses with larger genomes[90]. The structural-enzymatic arm split thus resembles the core- and pan-genomes of microbes, allowing versatile interactions between these viruses and their ANME-1 hosts (Supplementary Table 10). For example, head-tailed and spindle-shaped viruses encode various proteins involved in nucleotide and amino acid metabolisms, including ribonucleoside triphosphate reductase (NrdD), Queuosine biosynthesis enzymes (QueCDEF), and asparagine synthase, which can respectively boost nucleotide, preQ, and amino acid synthesis (Fig. 5b, 5c). Tepeuvirus PBV144 encodes a phosphoenolpyruvate carboxykinase (PEPCK), a central component of the gluconeogenesis pathway; ahpuchvirus PBV299 encodes a H+/gluconate symporter (GntT); various unclassified ANME-1 MGEs also encode rubisco activase CbbQ-like proteins. These enzymes may facilitate the carbon assimmilation of ANME-1 hosts (Supplementary Table 11). PhoU, involved in phosphate transport[91], and 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) reductase, which activates sulfate for assimilation, were also detected in ekchuahviruses, suggesting a potential viral boost of cellular P and S intake in ANME-1 cellular hosts during infection. PAPS reductase has been previously found in bacterial and archaeal viruses in hypersaline and marine environments[72,92,93], while *phoU* was reported from pelagic metaviromic surveys in the Pacific Ocean, but not assigned to archaeal viruses[94]. The detection of these virus-encoded auxiliary metabolic genes (AMGs) from hydrothermal vent systems reflects a broader trend in phosphate and nutrient manipulation by viruses in diverse environmental settings.

Our analysis of viral AMGs also suggested the involvement of viruses in the ancestral metabolic diversification of ANME-1. Specifically, the detection of *thyX,* a gene encoding thymidylate synthase, an essential enzyme involved in the synthesis of thymidine, in head-tailed ahpuchviruses (Fig. 5b) and ekchuahviruses (outside of the clade shown in Fig. 5c and d), and in spindle-shaped itzamnaviruses (but not demiitzamnaviruses, Fig. 5f) is consistent with the presence of *thyX* in the ANME-1 host. ThyX was recently reported as a unique feature which differentiated ANME-1 (with the exception of one sublineage) from other ANME lineages which encode the non-homologous thymidylate synthase gene, *thyA*[6]. ThyX catalyzes dUMP methylation into dTMP, likely boosting host thymidine synthesis during viral production. Notably, in itzamnavirus PBV082, *thyX* has been apparently replaced by a gene cassette including *thyA* as well as a phosphatase/kinase pair (Fig. 5f, 5g). The dichotomous distribution of the functional analogs *thyA/thyX* is prevalent across microbes, and the occurrence of *thyX* among the majority of ANME-1 compared with the use of *thyA* by their short-chain-alkane-degrading relatives (Fig. 1) and other ANMEs, highlights their unique evolutionary history and possible role of virus-mediated horizontal gene transfer. We carried out phylogenetic analyses to investigate the provenance of the ThyX encoded by ANME-1 and their viruses and found that these sequences form a distinct clade, distant from canonical ThyX encoded by bacteria, archaea as well as other *Caudoviricetes* (Fig. 6, Fig. S10a, S10b). Strikingly, ThyX encoded by itzamnaviruses form a well-supported monophyletic group located at the base of this divergent clade, and the deep-branching ANME-1c encode ThyX that belong to the second deepest branch. Notably, the Guaymas Basin-derived ANME-1c bin B22_G9 contains both a genomic *thyX*, as well as *thyX* encoded by a partial itzamnavirus-derived provirus (Fig. 6, Fig. S10a, S10c).

The above analyses thus suggest that *thyX* was first acquired by spindle-shaped ANME-1 viruses and then transmitted into the common ancestors of ANME-1, displacing *thyA*. Due to higher promiscuity of viral DNA polymerases and the intense arms race, viral genes are known to evolve rapidly[95], which is in line with the extreme divergence of the ANME-1/viral *thyX* from the canonical clade. Notably, our phylogenetic analysis indicates that ekchuahviruses and ahpuchvirues likely acquired *thyX* independently at a later stage.

Discussion

In this study, metagenomic characterization of a new hydrothermal vent environment in the South Pescadero Basin led to the expansion of the known '*Ca.* Methanophagales' (ANME-1) diversity to include '*Ca.* Methanoxibalbaceae' (ANME-1c) and their viruses. '*Ca.* Methanoxibalbaceae' is a previously undescribed deep-branching family that so far has only been detected in high temperature hydrothermal environments. Comparative genomics of this deep-branching ANME-1c clade provides a valuable perspective of the evolutionary continuum within the class '*Ca.* Syntrophoarchaeia', as this new group shares unique features with both '*Ca.* Syntrophoarchaeales' and '*Ca.* Alkanophagales', including hydrogenases, which are a rare feature within ANME-1. The phylogeny of these hydrogenases is congruent with the genome phylogeny indicating an apparent vertical inheritance and differential loss of these genes in ANME-1. While the specific physiological role of these hydrogenases, such as potential energy generation through methanotrophy or methanogenesis[6,18,19,24,64], is yet unknown, the differential inheritance as elucidated in this study suggest a nonobligatory role which appears to confer a long-standing, and likely unique selective advantage.

Our study also uncovered a putative viral source of the ANME-1-specific thymidylate synthase gene *thyX* that replaced the functional analog *thyA* gene that is otherwise maintained by other members of *Ca.* Syntrophoarchaeia. ThyX differs from ThyA in that it uses NADPH as an electron donor when transferring the methyl group from the C_1 intermediate $H_4MPT{=}CH_2$ to dUMP to yield dTMP, without oxidizing the H_4MPT moiety[6]. H_4MPT is a core co-factor constantly recycled through the Wood-Ljungdahl pathway that fuels ANME-1 anabolism[6,96]; NADPH abundance is highly dependent on the type of host energy metabolism and redox state[97,98]. The virus-induced ThyA-to-ThyX transition may have played a role

in the metabolic diversification and subsequent ecological expansion of the ANME-1 common ancestors into uncharted territories. Thymidylate synthase plays a part in C_1 anabolism recently found to be more divergent across ANME lineages than their C_1 energy metabolism[6]. It is plausible that viruses and other MGEs may be generally involved in the evolutionary diversification across ANMEs and their alkane-metabolizing relatives.

The expansive virome of ANME-1, as discovered in this study, encompasses novel representatives of all 4 major archaeal virus realms that are thought to have been infecting the last archaeal common ancestor[99]. Whereas varidnaviruses and duplodnaviruses infect hosts in all three domains of life, adnaviruses and spindle-shaped viruses are specific to domain Archaea. The ANME-1-targeting viruses are distant from all known bacterial and archaeal viruses, forming 16 previously undescribed virus families and at least 3 new orders. These families of viruses are characterized by many unique structural and replicative features, including new classes of MCPs, dual MCPs of different origins, and the unexpected presence of protein-primed and RNA-primed polymerase B. For example, whereas several groups of spindle-shaped viruses were previously known to encode protein-primed DNA polymerases (pPolB), 'Itzamnaviridae' and 'Tepeuviridae' represent the first putative spindle-shaped viruses with RNA-primed DNA polymerase (rPolB) genes. These diverse viruses significantly expand our appreciation of the archaeal virus diversity and their ecological significance.

ANME archaea play dominant roles in sequestering the greenhouse gas methane in diverse environments, as well as serving as important primary producers for the corresponding chemosynthetic ecosystems[1-4]. Our results open doors for targeted culture-dependent and culture-independent exploration of ANME virus-host interactions that are expected to play a critical role in the biogeochemical cycling[30,100] in these productive methane-driven ecosystems.

Material and Methods

Sampling and incubations. Four samples of minerals were collected from the 3.7 km-deep Auka vent field in the South Pescadero Basin (23.956094 N 108.86192 W)[36,38,101]. Sample NA091.008 was collected and incubated as described previously[69]. Samples 12019 (S0200-R1), 11719 (S0193-R2) and 11868 (S0197-PC1), the latter representing a lithified nodule recovered from a sediment push core, were collected with *ROV SuBastian* and *R/V Falkor* on cruise FK181031 in November 2018. These samples were processed shipboard and stored under anoxic conditions at 4 °C for subsequent incubation in the laboratory. In the laboratory, mineral sample 12019 and 11719 were broken into smaller pieces under sterile conditions, immersed in N_2-sparged artificial sea water and incubated under anoxic conditions with methane as described previously for NA091.008[69]. Additional sampling information can be found in Supplementary Table 1**.** Mineralogical analysis by XRD identified several of these samples as containing barite (11719, NA091.008), collected from two locations on the western side of the Matterhorn vent, and one sample (12019) recovered from the sedimented flanks from the southern side of Z vent, which was saturated with oil. Our analysis also includes metagenomic data from two sediment cores (DR750-PC67 and DR750-PC80) collected in April 2015 with the *ROV Doc Ricketts* and *R/V Western Flyer* (MBARI2015), previously published[38].

Fluorescence *in situ* hybridization. Samples were fixed shipboard using freshly prepared paraformaldehyde (2 vol% in 3x PBS, EMS) at 4°C overnight, rinsed twice using 3x PBS, and stored in ethanol (50% in 1xPBS) at -20°C until processing. Small pieces (< $1cm^3$) of the mineral sample NA091.008 were gently crushed in a sterile agate mortar and pestle in a freshly prepared, filter sterilized 80% ethanol – 1× PBS solution. About 500 µl of the resulting mixture was sonicated three times in 15 second bursts on a Branson Sonifier W-150 ultrasonic cell disruptor (level 3) on ice with a sterile remote-tapered microtip probe inserted into the liquid. Cells were separated from mineral matrix using an adapted protocol of

Percoll density separation[11]. The density-separated cells were filtered onto 25 mm polycarbonate filters with a pore size of 0.22 µm, and rinsed using 1x PBS. Fluorescence *in situ* hybridizations were carried out as described previously[11] using a 1:1 mixture of an ANME-1 targeted probe (ANME-1-350[14] labeled with Cy3) and the general bacterial probe mix (EUB-338 I-III[102,103] labeled with Alexa-488) at 35% of formamide concentration. Hybridized samples were imaged using a 100x objective using a Zeiss Elyra structured illumination microscope using the Zen Black software.

DNA extraction and sequencing. DNA extraction from the mineral samples followed previously published protocols[69]. Metagenomic analysis from the extracted genomic DNA was outsourced to Quick Biology (Pasadena, CA, USA) for library preparation and sequencing. Libraries were prepared with the KAPA Hyper plus kit using 10 ng of DNA as input. This input was subjected to enzymatic fragmentation at 37°C for 10 min. After end repair and A-tailing, the DNA was ligated with an IDT adapter (Integrated DNA Technologies Inc., Coralville, Iowa, USA). Ligated DNA was amplified with KAPA HiFi HotStart ReadyMix (2x) for 11 cycles. Post-amplification cleanup was performed with 1x KAPA pure beads. The final library quality and quantity were analyzed and measured by Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) and Life Technologies Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) respectively. Finally, the libraries were sequenced using 150 bp paired-end reads on Illumina HiSeq4000 Sequencer (Illumina Inc., San Diego, CA). After sequencing, primers and adapters were removed from all libraries using bbduk[104] with mink=6 and hdist=1 as trimming parameters, and establishing a minimum quality value of 20 and a minimal length of 50 bp. For incubated samples, DNA was amplified using multiple displacement amplification (MDA) with the QIAGEN REPLI-g Midi kit prior to library preparation for nanopore sequencing. Oxford Nanopore sequencing libraries were constructed using the PCR-free barcoding kit and were sequenced on PromethION platform by Novogene Inc.

Metagenomic analysis. The sequencing reads from unincubated rocks were assembled individually and in a coassembly using SPAdes v. 3.12.0[105]. From the de-novo assemblies, we performed manual binning using Anvio v. 6[51]. We assessed the quality and taxonomy affiliation from the obtained bins using GTDB-tk[106] and checkM [107]. Genomes affiliated to ANME-1 and *Syntrophoarchaeales* were further refined via a targeted-reassembly pipeline. In this pipeline, the original reads were mapped to the bin of interest using bbmap, then the mapped reads were assembled using SPAdes and finally the resulting assembly was filtered discarding contigs below 1500 bp. This procedure was repeated during several rounds (between 11-50) for each bin, until we could not see an improvement in the bin quality. Bin quality was assessed using the checkM and considering the completeness, contamination (< 5%), N50 value and number of scaffolds. The resulting bins were considered as metagenome-assembled genomes (MAGs). The sequencing reads for the incubated rocks 12019 and 11719 were assembled as described previously for NA091.R008[69]. Additionally, the assembly of 12019 was then scaffolded using Nanopore reads through two iterations of LRScaf v1.1.10[108]. The final assemblies were binned using metabat2 v2.15[109] using default setting. Automatic metabolic prediction of the MAGs was performed using prokka v. 1.14.6[110] and curated with the identification of Pfam[111] and TIGRFAM[112] profiles using HMMER v. 3.3 (hmmer.org); KEGG orthologs[113] with Kofamscan[114] and of COGs[115] and arCOGs motifs[116] using COGsoft[117]. To identify multiheme cytochromes in our genomes, we searched the motif CXXCH across the amino acid sequences predicted for each MAG. Similar metabolic predictions were carried out with publicly available ANME-1 and *Syntrophoarchaeales* genomes in order to compare the metabolic potential of the whole ANME-1 order. A list of the genomes used in this study can be found in Supplementary Table 2. For the comparison of different genomic features among the ANME-1 genomes, we searched for specific proteins using the assigned COGs, arCOGs and KEGG identifiers (Supplementary Table 5).

Genomic relative abundance analysis. We used the software coverM v. 0.5 (https://github.com/wwood/CoverM) to calculate the genomic relative abundance of the different organisms of our samples using all the MAGs we have extracted from our metagenomic analysis. We ran the software with the following additional parameters for dereplication ("--dereplication-ani 95 --dereplication-prethreshold-ani 90 --dereplication-precluster-method finch"). Results were visualized in R[118] using ggplot [119].

Optimum growth temperature analysis. We calculated the optimum growth temperature for all ANME-1 and *Syntrophoarchaeales* MAGs included in our analysis (Supplementary Table 2) using the OGT_prediction tool described in Sauer and Wang (2019)[39] with the regression models for Archaea excluding rRNA features and genome size.

Analysis of the hydrogenase operon of '*Candidatus* Methanospirare jalkutatii' genomes. Since only two of the five genomes of '*Ca.* Methanospirare jalkutatii' have an operon encoding a hydrogenase, we performed additional analysis to better understand this intraspecies distribution. On the one hand, we mapped the metagenomic reads from samples with genomes of '*Candidatus* Methanospirare jalkutatii' (12019, FW4382_bin126, NA091.008, PR1007, PR1031B) to the MAGs containing the hydrogenase operon (FW4382_bin126, NA091.008_bin1) to check if reads mapping this operon are also present in samples from where MAGs without the hydrogenase were recovered. For mapping the reads, we used bowtie2[120] and then transformed the sam files to bam using samtools[121] and finally extract the coverage depth for each position. Additionally, we performed a genomic comparison of the genomes with a hydrogenase operon (FW4382_bin126, NA091.008_bin1) with the genome FWG175 that was assembled into a single scaffold. For this, we used the genome-to-genome aligner Sibelia[122] and we visualized the results using Circos[123].

Phylogenetic analysis. For the phylogenomic tree of the ANME-1 MAGs, we used the list of genomes present in Supplementary Table 2. As marker genes, we used 31 single copy genes (Supplementary Table 5) that we extracted and aligned from the corresponding genomes using anvi-get-sequences-for-hmm-hits from Anvio v. 6 [51,124] with the parameters "--return-best-hit --max-num-genes-missing-from-bin 7 --partition-file". Seven genomes missed more than 7 marker genes and were not used for the phylogenomic reconstruction present in Figure 1 (ANME-1 UWMA-0191, Syntrophoarchaeum GoM_oil, ANME-1 ERB7, ANME-1 Co_bin174, ANME-1 Agg-C03, PB_MBMC_218, FW4382_bin035). The concatenated aligned marker gene set was then used to calculate a phylogenomic tree with RAxML v. 8.2.12[125] using a partition file to calculate differential models for each gene the following parameters "-m PROTGAMMAAUTO -f a -N autoMRE -k". The tree was then visualized using iTol[126]. For the clustering of the MAGs into different species, we dereplicated the ANME-1 MAGs using dRep v. 2.6.2 with the parameter "-S_ani 0.95"[127]. A smaller phylogenomic tree was calculated with the genomes containing hydrogenase genes (Fig. 3). For this tree we also used Anvio v. 6 and RAxML v. 8.2.12 with the same parameters but excluding the flag "—max-num-genes-missing-from-bin" from the anvi-get-sequences-for-hmm-hits command to include in the analysis those genomes with a lower number of marker genes that still contain hydrogenase genes (PB_MBMC_218, FW4382_bin035, ANME-1 UWMA-0191).

The 16S rRNA gene phylogenetic tree was calculated for the 16S rRNA genes predicted from our genome dataset that were full-length. We included these full-length 16S rRNA genes in the SILVA_132_SSURef_NR99 database[128] and with the ARB software[129] we calculated a 16S phylogenetic tree using the maximum-likelihood algorithm RAxML with GTRGAMMA as the model and a 50% similarity filter. One thousand bootstrap analyses were performed to calculate branch support values. The tree with the best likelihood score was selected.

For the construction of the hydrogenase phylogenetic tree ((Supplementary Table 6), we used the predicted protein sequence for the large subunit of the NiFe hydrogenase present in the genomes of our dataset (Supplementary Table 2), a subset of the large subunit hydrogenases present in the HydDB database[63] and the predicted hydrogenases present in an archaeal database using the COG motif for the large NiFe hydrogenase (COG0374) with the Anvio v. 6 software. For the mcrD gene phylogeny, we used the predicted protein sequences of mcrD in the ANME-1c genomes and in the previously mentioned archaeal database with the TIGR motif TIGR03260.1 using also the Anvio v. 6 software. The list of genomes from the archaeal database used in the analysis can be found in Supplementary Table 6. For both phylogenies, the protein sequences for the analysis were aligned using clustalw v.2.1 with default settings [130]. The aligned file was used to calculate a phylogenetic tree using RAxML v. 8.2.12 [125] with the following parameters "-m PROTGAMMAAUTO -f a -N 100 –k". The tree was then visualized using iTol[126]. For the distribution and phylogenetic analysis of MCP and pPolB, known sequences encoded by various bacterial and archaeal viruses were used to build a Hidden Markov Model (HMM) via hmmer v3.3.2[131]. The HMM was then used to capture the corresponding components in proteomes of ANME-1 viruses and other MGEs. All sequences were then aligned using MAFFT v7.475[132] option linsi and trimmed using trimAl v1.4.1[124] option gappyout for pPolB and 20% gap removal option for MCP. Maximum-likelihood analyses were carried out through IQtree v2.1.12[133] using model finder and ultrafast bootstrap with 2000 replicates. The phylogenetic tree was visualized and prepared using iTOL[126].

For the distribution of ThyX, all ThyX sequences annotated by EggNOG mapper[134] v2 in the genomes of ANME-1 and their MGEs were used to create a HMM as described above, and used to search for close homologs in the GTDB202 database, IMGVR V.3 database, as well as again in the proteomes and ANME-1 and their MGEs in this study. This yielded 261 sequences, which was then aligned and phylogenetically analyzed as described above.

CRISPR analysis. The CRISPR/Cas systems from the ANME-1 genomes and various metagenomic assemblies were annotated using CRISPRCasTyper v.1[68]. CRISPR spacer mapping onto MGEs was carried out as previously described[69] with the following modifications. To filter out unreliable sequences that may have arisen during MAG binning, we took a conservative measure of only retaining CRISPR repeats identified in at least three ANME-1 contigs. We additionally analyzed the CRISPR repeats found in the *Alkanophagales* sister clade to ANME-1 using the same approach, which were found to have no overlap with the ANME-1 CRISPR repeats. To further avoid accidental mapping to unrelated MGEs, we applied a second stringent criteria of only retaining MGEs with at least 3 ANME-1 protospacers. MGEs larger than 10kb in size were retained for further analyses in this study.

Virus protein annotations. Open reading frames in viral contigs were identified using the PATRIC package[135] annotated using sensitive hidden Markov model profile-profile comparisons with HHsearch v3.3.0[136] against the following publicly available databases: Pfam 33.1, Protein Data Bank (25/03/2021), CDD v3.18, PHROG (PMID: 34377978) and uniprot_sprot_vir70 (09/02/2021)[137]. Putative major capsid proteins of 'Chaacviridae' and 'Ixchelviridae' could not be identified using sequence similarity based approaches. Thus, the candidate proteins were subjected to structural modeling using AlphaFold2[81,138] and RoseTTAFold[5]. The obtained models were visualized using ChimeraX[139] and compared to the reference structure of the major capsid protein of corticovirus PM2 (PDB id: 2vvf).

Mobilome network analysis and evaluation. Gene similarity network analyses were done using vCONTACT2 using the default reference, with head-tailed viruses targeting haloarchaea and methanogens added as additional references[72]. Inverted and direct terminal repeats were detected using CheckV and the PATRIC package[135].

Virus genome alignment. The viral genomes were annotated using Prokka v1.14.6[110] to produce genbank files. Select genbank files were then analyzed using Clinker v. 0.0.23[140] to produce the protein sequence clustering and alignments.

Taxonomic description of '*Ca.* Methanoxibalbaceae'

Phylogenomic analysis placed the MAGs belonging to ANME-1c into two different genera, represented by one species in each ('*Candidatus* Methanoxibalbensis ujae' and '*Candidatus* Methanospirare jalkutatii'). Both genera belong to the same family named *'Candidatus* Methanoxibalbaceae' that it is included within the order *'Candidatus* Methanophagales'.

Description of the proposed genus '*Candidatus* Methanoxibalbensis ujae'
(N.L. neut. n. *methanum* methane; N.L. pref. *methano-*, pertaining to methane; N.L. adj. *xibalbensis* from the place called *Xibalba*, the Mayan word for the underworld; N.L. neut. n. *Methanoxibalbensis* methane-cycling organism present in deep-sea hydrothermal sediments; N.L. neut. adj. *ujae,* from the Kiliwa word *ujá* from Baja California meaning rock, referred to the high abundance of this species in rock samples). This organism is not cultured and is represented by three MAGs from sedimented hydrothermal vents in the Gulf of California, one recovered from the Guaymas Basin and two from the South Pescadero Basin (see Supplementary Table 2). This group presumably is meso- or thermophilic and inhabits hydrothermal deep-sea environments, being mostly detected in rock samples. The type material is the genome designated NA091.008_bin2, a MAG comprising 1.96 Mbp in 86 scaffolds. The MAG was recovered from mineral sample (NA091.008) from the hydrothermal environment of South Pescadero Basin. It is proposed to be capable of anaerobic methanotrophy.

Description of the proposed genus '*Candidatus* Methanospirare jalkutatii'
(N.L. neut. n. *methanum* methane; N.L. pref. *methano-*, pertaining to methane; L. v. *spirare* to breathe; N.L. neut. n. *Methanospirare* methane-breathing organism; N.L. masc. n. *jalkutatii,* mythical dragon of the paipai cosmology from Baja California, this dragon inhabited a beautiful place made of rocks and water similar to the Auka vent site). This organism is not cultured and is represented by five MAGs, all of them recovered from the hydrothermal environment of South Pescadero Basin. The type material is the genome designated FWG175, a single-scaffolded MAG comprising 1.99 Mbp in 1 circular scaffolds. This MAG was recovered from a methane-fed incubation of the mineral sample 12019 retrieved from the hydrothermal environment of South Pescadero Basin.

Description of the proposed family '*Candidatus* Methanoxibalbaceae'
(N.L. neut. n. *Methanoxibalbensis* a (Candidatus) genus name; *-aceae* ending to denote a family). N.L. neut. pl. n. *Methanoxibalbaceae* the (Candidatus) family including the (Candidatus) genera of *Methanoxibalbensis* and *Methanospirare*. The description is the same as for the candidate type genus *Methanoxibalbensis*.

Taxonomic description of proposed ANME-1 virus orders and families with representative complete genomes

Order *Coyopavirales*, family *Chaacviridae.* This group of viruses is characterized by novel major capsid protein (MCP), which is predicted using AlphaFold2 to have the double jelly-roll (DJR) fold, the hallmark protein of viruses within the realm *Varidnaviria*. We propose classifying these viruses into a new family, *Chaacviridae*, after Chaac, the god of death in the Mayan mythology. Chaacviruses displays minimal

proteome overlap with other known viruses, and is characterized by a uniform 10-11 kb genome size and a gene encoding protein-primed family B DNA polymerase (pPolB).

Chaacviridae comprises two genera that have relatively similar functional composition in their proteomes, yet exhibit relatively low sequence conservation and distinct gene arrangements. Viruses in the two genera appeared to have undergone a genomic inversion during their evolutionary history. We propose the genus names *Homochaacvirus* and *Antichaacvirus* (from *homo* [same in Greek] and *anti* [opposed in Greek] to emphasize the inversion of a gene module including the *pPolB* gene). Four complete genomes of chaacviruses have been obtained, as judged from the presence of inverted terminal repeats, consistent with the presence of *pPolB* gene. The four viruses share less than 90% average nucleotide identity and represent separate species. *Homochaacvirus* genus will include viruses GBV261, GBV265 and GBV275, whereas genus *Antichaacvirus* will include a single representative PBV266.

While *Chaacviridae* is not closely related to any of the existing virus families, based on the gene complement, chaacviruses resemble members of the class *Tectiliviricetes*. Thus, we propose placing *Chaacviridae* into a new monotypic order *Coyopavirales* (after Coyopa, the god of thunder in Mayan mythology) within the existing class *Tectiliviricetes*.

Order *Nakonvirales*, families *Ahpuchviridae* and *Ekchuahviridae.* Based on ViPTree analysis, All ANME-1 viruses belonging to *Caudoviricetes* form a distinct clade outside of the three existing orders of haloarchaeal and methanogenic archaeal viruses [72] (Fig. 4). We propose to create a new order *Nakonvirales* (after Nakon, the most powerful god of war in Mayan mythology) for unification of two family-level groups that have complete genome representatives.

We propose naming the first of the two groups *Ahpuchviridae* (after Ah Puch, the god of death in the Mayan mythology). This family is represented by one genus *Kisinvirus* (after Kisin, another Mayan god of death) and a single species, *Kisinvirus pescadero*. The species includes virus PBV299, which has a dsDNA genome of 70,925 bp and besides the morphogenetic genes typical of members of the *Caudoviricetes,* encodes an RNA-primed family B DNA polymerase, archaeo-eukaryotic primase and a processivity factor PCNA.

The second proposed family, *Ekchuahviridae* (after Ek Chuah, the patron god of warriors and merchants in the Mayan mythology), is represented by one genus *Kukulkanvirus* (after Kukulkan, the War Serpent in the Mayan mythology). The proposed genus will include two species, *Kukulkanvirus IMGVR0083622* and *Kukulkanvirus IMGVR0540589*, with representative viruses containing genomes of 80,551 bp and 71,795 bp, respectively. Notably, viruses in this group encode two divergent HK97-fold MCPs with their own capsid maturation proteases, but all other canonical head-tailed virus structural proteins are encoded as single copy genes. Their replication modules include RNA-primed family B DNA polymerase and archaeo-eukaryotic primase.

Order *Maximonvirales,* family *Ahmunviridae.* For classification of rod-shaped virus PBV300, we propose the creation of a new genus, *Yumkaaxvirus* (after Yum Kaax, the god of the woods, the wild nature, and the hunt in Mayan mythology) within a new family, *Ahmunviridae* (after Ah Mun, the god of agriculture in Mayan mythology). PBV300 has a linear dsDNA genome of 41,525 bp with 99 bp terminal inverted repeats. It encodes two divergent MCPs homologous to those of viruses in the realm *Adnaviria*. The virus also encodes several other proteins with homologs in members of the family *Rudiviridae*, including the terminal fiber protein responsible for receptor binding. This family is related to members of the class *Tokiviricetes* (realm *Adnaviria*), but outside of the two existing orders, *Ligamenvirales* and *Primavirales*. We thus propose to assign *Ahmunviridae* to a new order *Maximonvirales*, after Maximon, a god of travelers, merchants, medicine men/women, mischief and fertility in Mayan mythology.

Family *Itzamnaviridae.* We propose the family name *Itzamnaviridae* (after Itzamna, lord of the heavens as well as night and day in the Mayan mythology) for the spindle-shaped viruses with complete genomes in this study. The members of this family differ in genome sizes and are subdivided into two genera, which we propose naming *Demiitzamnavirus* and *Pletoitzamnavirus* (after *demi-* for half or partial [derived via French from Latin '*dimedius*'] and *pleto* for full [Latin]). Demiitzamnaviruses have circular genomes sized around 25 kb, each encoding two MCPs homologous to those characteristics of and exclusive to archaeal spindle-shaped viruses. This genus is represented by two species, *Demiitzamnavirus guaymas* and *Demiitzamnavirus IMGVR0402074*. Pletoitzamnaviruses have genome sizes around 45-48 kb, where roughly half of the genome aligns with nearly the entirety of the genomes of demiitzamnaviruses, including the two MCP genes. However, the remaining fraction of pletoitzamnavirus genomes encodes enzymes of diverse functions, including replicative proteins, such as RNA-primed family B DNA polymerase and archaeo-eukaryotic primase, PCNA, large and small subunit of the replication factor C, etc. The content of this genome fraction is relatively flexible, with genes encoding various metabolic and regulatory enzymes swapping in and out, often in the form of multi-gene cassettes. For example, the representative pletoitzamnavirus, PBV082, contains a gene cluster encoding a kinase-phosphatase pair, a thymidylate synthase, and a radical SAM enzyme of unknown function. Except for the characteristic MCP, members of the *Itzamnaviridae* do not encode proteins with appreciable sequence similarity to proteins of other spindle-shaped viruses.

A formal proposal for classification of the ANME-1 viruses discovered in this study has been submitted for consideration by the International Committee for Taxonomy of Viruses (ICTV) and are detailed in Supplementary Table 12.

Data availability

All MAGs and sequence data can be found in the figshare link https://figshare.com/projects/Laso-Perez_and_Wu_-_ANME-1_project/140453 (except some large libraries), and will be deposited onto NCBI database prior to peer-reviewed publication.

Acknowledgements

We are indebted to the crews from R/V Falkor (cruise FK181031) and E/V Nautilus (cruise NA091) and the pilots of ROVs SuBastian and Hercules. Sample collection permits for FK181031 (25/07/2018) were granted by la Dirección General de Ordenamiento Pesquero y Acuícola, Comisión Nacional de Acuacultura y Pesca (CONAPESCA: Permiso de Pesca de Fomento No. PPFE/DGOPA-200/18) and la Dirección General de Geografía y Medio Ambiente, Instituto Nacional de Estadística y Geografía (INEGI: Autorización EG0122018), with the associated Diplomatic Note number 18-2083 (CTC/07345/18) from la Secretaría de Relaciones Exteriores - Agencia Mexicana de Cooperación Internacional para el Desarrollo / Dirección General de Cooperación Técnica y Científica. Sample collection permit for cruise NA091 (18/04/2017) was obtained by the Ocean Exploration Trust under permit number EG0072017. This research used samples provided by the Ocean Exploration Trust's Nautilus Exploration Program, cruise NA091. E/V Nautlius operated by the Ocean Exploration Trust, with cruise NA091 supported by the Dalio Foundation and Woods Hole Oceanographic Institute, and R/V Falkor operated by the Schmidt Ocean Institute. Funding for this work was provided by grants from the National Science Foundation Center For Dark Energy Biosphere Investigations (C-DEBI) and the NOMIS foundation (V.J.O.). VJO contribution was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0020373. This work was also supported by the Deutsche Forschungsgemeinschaft

(DFG, German Research Foundation) under Germany's Excellence Initiative/Strategy through the Cluster of Excellence "The Ocean Floor-Earth's Uncharted Interface" (EXC-2077-390741603 R.L.-P. and V.J.O). M.K. was supported by l'Agence Nationale de la Recherche grant ANR-20-CE20-0009-02. F.W. was supported by the Dutch Research Council Rubicon Award 019.162LW.037, the Human Frontiers Science Program Long-term fellowship LT000468/2017, and a ZJU-HIC Independent PI Startup Grant.

Disclaimer:

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Author contributions

R. L.-P., F.W., A.C., and V.J.O. conceived and designed the study. D.R.S, V.J.O. and J.S.M. retrieved the original samples. A.C. and F.W. carried out rock incubations and FISH microscopy. R.L.-P., F.W. and A.C. performed DNA extraction. R.L.-P. and F.W. performed metagenomic assembly and analysis. F.W. performed CRISPR-based mobilome discovery. M.K. and F.W. performed analyses of viruses. R.L.-P., F.W., M.K. and V.J.O wrote the manuscript with contributions from all coauthors. We declare no competing financial interests.

References

1 Reeburgh, W. S. Oceanic Methane Biogeochemistry. *Chemical Reviews* **107**, 486-513, doi:10.1021/cr050362v (2007).

2 Hinrichs, K.-U. & Boetius, A. in *Ocean Margin Systems* (eds Gerold Wefer *et al.*) 457-477 (Springer Berlin Heidelberg, 2003).

3 Egger, M., Riedinger, N., Mogollón, J. M. & Jørgensen, B. B. Global diffusive fluxes of methane in marine sediments. *Nature Geoscience* **11**, 421-425, doi:10.1038/s41561-018-0122-8 (2018).

4 Hinrichs, K.-U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaebacteria in marine sediments. *Nature* **398**, 802-805, doi:10.1038/19751 (1999).

5 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876, doi:10.1126/science.abj8754 (2021).

6 Chadwick, G. L. *et al.* Comparative genomics reveals electron transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic archaea. *PLOS Biology* **20**, e3001508, doi:10.1371/journal.pbio.3001508 (2022).

7 Holler, T. *et al.* Thermophilic anaerobic oxidation of methane by marine microbial consortia. *The ISME journal* **5**, 1946 (2011).

8 Krukenberg, V. *et al.* Candidatus Desulfofervidus auxilii, a hydrogenotrophic sulfate‐reducing bacterium involved in the thermophilic anaerobic oxidation of methane. *Environmental microbiology* **18**, 3073-3091 (2016).

9 Orphan, V. J., House, C. H., Hinrichs, K.-U., McKeegan, K. D. & DeLong, E. F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proceedings of the National Academy of Sciences* **99**, 7663-7668, doi:10.1073/pnas.072210299 (2002).

10 Knittel, K. & Boetius, A. Anaerobic Oxidation of Methane: Progress with an Unknown Process. *Annual Review of Microbiology* **63**, 311-334, doi:10.1146/annurev.micro.61.080706.093130 (2009).

11 McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531, doi:10.1038/nature15512 (2015).

12 Laso-Pérez, R. *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396, doi:10.1038/nature20152 (2016).

13 Eller, G., Känel, L. & Krüger, M. Cooccurrence of Aerobic and Anaerobic Methane Oxidation in the Water Column of Lake Plußsee. *Applied and Environmental Microbiology* **71**, 8925-8928, doi:doi:10.1128/AEM.71.12.8925-8928.2005 (2005).

14 Knittel, K., Lösekann, T., Boetius, A., Kort, R. & Amann, R. Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. *Applied and environmental microbiology* **71**, 467-479, doi:10.1128/aem.71.1.467-479.2005 (2005).

15 Orcutt, B., Boetius, A., Elvert, M., Samarkin, V. & Joye, S. B. Molecular biogeochemistry of sulfate reduction, methanogenesis and the anaerobic oxidation of methane at Gulf of Mexico cold seeps. *Geochimica et Cosmochimica Acta* **69**, 4267-4281, doi:https://doi.org/10.1016/j.gca.2005.04.012 (2005).

16 Treude, T. *et al.* Consumption of Methane and CO₂ by Methanotrophic Microbial Mats from Gas Seeps of the Anoxic Black Sea. *Applied and Environmental Microbiology* **73**, 2271-2283, doi:doi:10.1128/AEM.02685-06 (2007).

17 House, C. H. *et al.* Extensive carbon isotopic heterogeneity among methane seep microbiota. *Environmental Microbiology* **11**, 2207-2215, doi:https://doi.org/10.1111/j.1462-2920.2009.01934.x (2009).

18 Biddle, J. F. *et al.* Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *The ISME journal* **6**, 1018, doi:10.1038/ismej.2011.164 (2011).

19 Kevorkian, R. T., Callahan, S., Winstead, R. & Lloyd, K. G. ANME-1 archaea may drive methane accumulation and removal in estuarine sediments. *Environmental Microbiology Reports* **13**, 185-194, doi:https://doi.org/10.1111/1758-2229.12926 (2021).

20 Nauhaus, K., Boetius, A., Krüger, M. & Widdel, F. In vitro demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. *Environmental microbiology* **4**, 296-305, doi:doi:10.1046/j.1462-2920.2002.00299.x (2002).

21 Runar, S., Irene, R., Anders, L., Haflidi, H. & H., S. I. Integrated metagenomic and metaproteomic analyses of an ANME‐1‐dominated community in marine cold seep sediments. *Environmental microbiology* **14**, 1333-1346, doi:doi:10.1111/j.1462-2920.2012.02716.x (2012).

22 Wang, F.-P. *et al.* Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *The ISME journal* **8**, 1069, doi:10.1038/ismej.2013.212 (2013).

23 Krukenberg, V. *et al.* Gene expression and ultrastructure of meso- and thermophilic methanotrophic consortia. *Environmental microbiology* **20**, 1651-1666, doi:doi:10.1111/1462-2920.14077 (2018).

24 Beulig, F., Røy, H., McGlynn, S. & Jørgensen, B. Cryptic CH4 cycling in the sulfate–methane transition of marine sediments apparently mediated by ANME-1 archaea. *ISME J* **296**, 10.1038 (2018).

25 Magnuson, E. *et al.* Active lithoautotrophic and methane-oxidizing microbial community in an anoxic, sub-zero, and hypersaline High Arctic spring. *The ISME Journal*, doi:10.1038/s41396-022-01233-8 (2022).

26 Li, Z. *et al.* Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *The ISME Journal* **15**, 2366-2378, doi:10.1038/s41396-021-00932-y (2021).

27 Paul, B. G. *et al.* Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nature Communications* **6**, 6585, doi:10.1038/ncomms7585 (2015).

28 Wang, L. *et al.* Potential metabolic and genetic interaction among viruses, methanogen and methanotrophic archaea, and their syntrophic partners. *ISME Communications* **2**, 50, doi:10.1038/s43705-022-00135-2 (2022).

29 Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology* **18**, 21-34, doi:10.1038/s41579-019-0270-x (2020).

30 Danovaro, R. *et al.* Virus-mediated archaeal hecatomb in the deep seafloor. *Sci Adv* **2**, e1600492, doi:10.1126/sciadv.1600492 (2016).

31 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336, doi:10.1038/s41586-019-1894-8 (2020).

32 Thompson Luke, R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences* **108**, E757-E764, doi:10.1073/pnas.1102164108 (2011).

33 Haaber, J. *et al.* Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nature Communications* **7**, 13333, doi:10.1038/ncomms13333 (2016).

34 San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends in Microbiology* **26**, 978-985, doi:10.1016/j.tim.2018.06.007 (2018).

35 Koonin, E. V., Makarova, K. S., Wolf, Y. I. & Krupovic, M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nature Reviews Genetics* **21**, 119-131, doi:10.1038/s41576-019-0172-9 (2020).

36 Paduan, J. B. *et al.* Discovery of Hydrothermal Vent Fields on Alarcón Rise and in Southern Pescadero Basin, Gulf of California. *Geochemistry, Geophysics, Geosystems* **19**, 4788-4819, doi:https://doi.org/10.1029/2018GC007771 (2018).

37 Dombrowski, N., Teske, A. P. & Baker, B. J. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nature Communications* **9**, 4999, doi:10.1038/s41467-018-07418-0 (2018).

38 Speth, D. R. *et al.* Microbial communities of Auka hydrothermal sediments shed light on vent biogeography and the evolutionary history of thermophily. *The ISME Journal*, doi:10.1038/s41396-022-01222-x (2022).

39 Sauer, D. B. & Wang, D.-N. Predicting the optimal growth temperatures of prokaryotes using only genome derived features. *Bioinformatics* **35**, 3224-3231, doi:10.1093/bioinformatics/btz059 (2019).

40 Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth Temperature and Genome Size in Bacteria Are Negatively Correlated, Suggesting Genomic Streamlining During Thermal Adaptation. *Genome Biology and Evolution* **5**, 966-977, doi:10.1093/gbe/evt050 (2013).

41 Meyerdierks, A. *et al.* Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environmental microbiology* **12**, 422-439, doi:doi:10.1111/j.1462-2920.2009.02083.x (2010).

42 Ermler, U., Grabarse, W., Shima, S., Goubeaud, M. & Thauer, R. K. Crystal Structure of Methyl-Coenzyme M Reductase: The Key Enzyme of Biological Methane Formation. *Science* **278**, 1457-1462, doi:10.1126/science.278.5342.1457 (1997).

43 Bokranz, M. & Klein, A. Nucleotide sequence of the methyl coenzyme M reductase gene cluster from Methanosarcina barkeri. *Nucleic acids research* **15**, 4350 (1987).

44 Prakash, D., Wu, Y., Suh, S.-J. & Duin, E. C. Elucidating the Process of Activation of Methyl-Coenzyme M Reductase. *Journal of Bacteriology* **196**, 2491-2498, doi:doi:10.1128/JB.01658-14 (2014).

45 Zheng, K., Ngo, P. D., Owens, V. L., Yang, X.-p. & Mansoorabadi, S. O. The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* **354**, 339-342, doi:10.1126/science.aag2947 (2016).

46 Mayr, S. *et al.* Structure of an F430 Variant from Archaea Associated with Anaerobic Oxidation of Methane. *Journal of the American Chemical Society* **130**, 10758-10767, doi:10.1021/ja802929z (2008).

47 Shima, S. *et al.* Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature* **481**, 98-101, doi:10.1038/nature10663 (2012).

48 Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: their function and mechanisms. *Applied Microbiology and Biotechnology* **60**, 12-23, doi:10.1007/s00253-002-1107-4 (2002).

49 Tchong, S.-I., Xu, H. & White, R. H. l-Cysteine Desulfidase: An [4Fe-4S] Enzyme Isolated from Methanocaldococcus jannaschii That Catalyzes the Breakdown of l-Cysteine into Pyruvate, Ammonia, and Sulfide. *Biochemistry* **44**, 1659-1670, doi:10.1021/bi0484769 (2005).

50 Kinosita, Y., Uchida, N., Nakane, D. & Nishizaka, T. Direct observation of rotation and steps of the archaellum in the swimming halophilic archaeon Halobacterium salinarum. *Nature Microbiology* **1**, 16148, doi:10.1038/nmicrobiol.2016.148 (2016).

51 Albers, S.-V. & Jarrell, K. F. The archaellum: how Archaea swim. *Frontiers in Microbiology* **6**, doi:10.3389/fmicb.2015.00023 (2015).

52 Albers, S.-V. & Jarrell, K. F. The Archaellum: An Update on the Unique Archaeal Motility Structure. *Trends in Microbiology* **26**, 351-362, doi:https://doi.org/10.1016/j.tim.2018.01.004 (2018).

53 Leigh, J. A. Nitrogen fixation in methanogens: the archaeal perspective. *Current issues in molecular biology* **2**, 125-131 (2000).

54 van Wolferen, M., Orell, A. & Albers, S.-V. Archaeal biofilm formation. *Nature Reviews Microbiology* **16**, 699-713 (2018).

55 Pohlschroder, M. & Esquivel, R. N. Archaeal type IV pili and their involvement in biofilm formation. *Frontiers in microbiology* **6**, 190 (2015).

56 Walker, D. *et al.* The archaellum of Methanospirillum hungatei is electrically conductive. . *mBio* **10**, e00579-00519 (2019).

57 Holmes, D. E., Zhou, J., Ueki, T., Woodard, T. & Lovley, D. R. Mechanisms for Electron Uptake by Methanosarcina acetivorans during Direct Interspecies Electron Transfer. *Mbio* **12**, e02344-02321 (2021).

58 Quemin, E. R. *et al.* First insights into the entry process of hyperthermophilic archaeal viruses. *J Virol* **87**, 13379-13385, doi:10.1128/jvi.02742-13 (2013).

59 Baquero, D. P. *et al.* in *Advances in Virus Research* Vol. 108 (eds Margaret Kielian, Thomas C. Mettenleiter, & Marilyn J. Roossinck) 127-164 (Academic Press, 2020).

60 Pihl, T. D., Schicho, R. N., Kelly, R. M. & Maier, R. J. Characterization of hydrogen-uptake activity in the hyperthermophile Pyrodictium brockii. *Proceedings of the National Academy of Sciences* **86**, 138-141, doi:10.1073/pnas.86.1.138 (1989).

61 Laska, S., Lottspeich, F. & Kletzin, A. Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon Acidianus ambivalens. *Microbiology* **149**, 2357-2371, doi:https://doi.org/10.1099/mic.0.26455-0 (2003).

62 Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *The ISME Journal* **10**, 761-777, doi:10.1038/ismej.2015.153 (2016).

63 Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Scientific reports* **6**, 34212-34212, doi:10.1038/srep34212 (2016).

64 Bertram, S. *et al.* Methanogenic capabilities of ANME-archaea deduced from 13C-labelling approaches. *Environmental Microbiology* **15**, 2384-2393, doi:https://doi.org/10.1111/1462-2920.12112 (2013).

65 Sørensen, K. B., Finster, K. & Ramsing, N. B. Thermodynamic and kinetic requirements in anaerobic methane oxidizing consortia exclude hydrogen, acetate, and methanol as possible electron shuttles. *Microbial Ecology* **42**, 1-10, doi:10.1007/s002480000083 (2001).

66 Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biology* **8**, R61, doi:10.1186/gb-2007-8-4-r61 (2007).

67 Shmakov, S. A. *et al.* The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *mBio* **8**, e01397-01317, doi:10.1128/mBio.01397-17 (2017).

68 Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *The CRISPR Journal* **3**, 462-469, doi:10.1089/crispr.2020.0059 (2020).

69 Wu, F. *et al.* Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. *Nature Microbiology* **7**, 200-212, doi:10.1038/s41564-021-01039-y (2022).

70 Roux, S. *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Research* **49**, D764-D775, doi:10.1093/nar/gkaa946 (2021).

71 Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* **37**, 632-639, doi:10.1038/s41587-019-0100-8 (2019).

72 Liu, Y. *et al.* Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. *PLOS Biology* **19**, e3001442, doi:10.1371/journal.pbio.3001442 (2021).

73 Koonin Eugene, V. *et al.* Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews* **84**, e00061-00019, doi:10.1128/MMBR.00061-19 (2020).

74 Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Current Opinion in Virology* **36**, 9-16, doi:https://doi.org/10.1016/j.coviro.2019.02.001 (2019).

75 Krupovič, M. & Bamford, D. H. Virus evolution: how far does the double β-barrel viral lineage extend? *Nature Reviews Microbiology* **6**, 941-948, doi:10.1038/nrmicro2033 (2008).

76 Krupovic, M. *et al.* Adnaviria: a New Realm for Archaeal Filamentous Viruses with Linear A-Form Double-Stranded DNA Genomes. *Journal of Virology* **95**, e00673-00621, doi:10.1128/JVI.00673-21 (2021).

77 Wang, F. *et al.* Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. *Proceedings of the National Academy of Sciences* **117**, 19643-19652, doi:10.1073/pnas.2011125117 (2020).

78 Baquero, D. P. *et al.* New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *The ISME Journal* **14**, 1821-1833, doi:10.1038/s41396-020-0653-z (2020).

79 Krupovic, M., Quemin Emmanuelle, R. J., Bamford Dennis, H., Forterre, P. & Prangishvili, D. Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea. *Journal of Virology* **88**, 2354-2358, doi:10.1128/JVI.02941-13 (2014).

80 Wang, F. *et al.* Spindle-shaped archaeal viruses evolved from rod-shaped ancestors to package a larger genome. *Cell* **185**, 1297-1307.e1211, doi:10.1016/j.cell.2022.02.019 (2022).

81 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

82 Abrescia, N. G. *et al.* Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Molecular cell* **31**, 749-761 (2008).

83 Medvedeva, S. e. a. Three families of Asgard archaeal viruses identified in metagenome-assembled genomes. *Nature Microbiology*, doi:10.1101/2021.07.29.453957 (in press).

84 Medvedeva, S. *et al.* Viruses of Asgard archaea. *bioRxiv*, 2021.2007.2029.453957, doi:10.1101/2021.07.29.453957 (2021).

85 Yutin, N., Bäckström, D., Ettema, T. J. G., Krupovic, M. & Koonin, E. V. Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virology Journal* **15**, 67, doi:10.1186/s12985-018-0974-y (2018).

86 Tamarit D, C. E., Krupovic M, Nijland R, Eme L, Robinson NP, Ettema TJG. A closed Candidatus Odinarchaeum genome exposes Asgard archaeal viruses. *Nature Microbiology*, doi:10.1101/2021.09.01.458545 (in press).

87 Rinke, C. *et al.* A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature Microbiology* **6**, 946-959, doi:10.1038/s41564-021-00918-8 (2021).

88 Nishimura, Y. *et al.* ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379-2380, doi:10.1093/bioinformatics/btx157 (2017).

89 Fokine, A. *et al.* Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proceedings of the National Academy of Sciences* **102**, 7163-7168, doi:10.1073/pnas.0502164102 (2005).

90 Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. & Venclovas, Č. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res* **48**, 10142-10156, doi:10.1093/nar/gkaa760 (2020).

91 Muda, M., Rao, N. N. & Torriani, A. Role of PhoU in phosphate transport and alkaline phosphatase regulation. *Journal of Bacteriology* **174**, 8057-8064, doi:10.1128/jb.174.24.8057-8064.1992 (1992).

92 Mizuno, C. M. *et al.* Novel haloarchaeal viruses from Lake Retba infecting Haloferax and Halorubrum species. *Environmental Microbiology* **21**, 2129-2147, doi:https://doi.org/10.1111/1462-2920.14604 (2019).

93 Mara, P. *et al.* Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *The ISME Journal* **14**, 3079-3092, doi:10.1038/s41396-020-00739-3 (2020).

94 Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *The ISME Journal* **9**, 472-484, doi:10.1038/ismej.2014.143 (2015).

95 Dennehy, J. J. Evolutionary ecology of virus emergence. *Annals of the New York Academy of Sciences* **1389**, 124-146, doi:https://doi.org/10.1111/nyas.13304 (2017).

96 Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood–Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. *Genome Biology and Evolution* **8**, 1706-1711, doi:10.1093/gbe/evw114 (2016).

97 Goldbeck, O., Eck, A. W. & Seibold, G. M. Real Time Monitoring of NADPH Concentrations in Corynebacterium glutamicum and Escherichia coli via the Genetically Encoded Sensor mBFP. *Frontiers in Microbiology* **9**, doi:10.3389/fmicb.2018.02564 (2018).

98 Blacker, T. S. *et al.* Separating NADH and NADPH fluorescence in live cells and tissues using FLIM. *Nature Communications* **5**, 3936, doi:10.1038/ncomms4936 (2014).

99 Krupovic, M., Dolja, V. V. & Koonin, E. V. The LUCA and its complex virome. *Nat Rev Microbiol* **18**, 661-670, doi:10.1038/s41579-020-0408-x (2020).

100 Rahlff, J. *et al.* Lytic archaeal viruses infect abundant primary producers in Earth's crust. *Nature Communications* **12**, 4642, doi:10.1038/s41467-021-24803-4 (2021).

101 Goffredi, S. K. *et al.* Mixotrophic chemosynthesis in a deep-sea anemone from hydrothermal vents in the Pescadero Basin, Gulf of California. *BMC Biology* **19**, 8, doi:10.1186/s12915-020-00921-1 (2021).

102 Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and environmental microbiology* **56**, 1919-1925 (1990).

103 Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The Domain-specific Probe EUB338 is Insufficient for the Detection of all Bacteria: Development and Evaluation of a more Comprehensive Probe Set. *Systematic and applied microbiology* **22**, 434-444, doi:https://doi.org/10.1016/S0723-2020(99)80053-8 (1999).

104 Bushnell, B. BBMap short read aligner. *University of California, Berkeley, California. URL http://sourceforge. net/projects/bbmap* (2016).

105 Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

106 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927, doi:10.1093/bioinformatics/btz848 (2020).

107 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

108 Qin, M. *et al.* LRScaf: improving draft genomes using long noisy reads. *BMC Genomics* **20**, 955, doi:10.1186/s12864-019-6337-2 (2019).

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

110 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

111 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419, doi:10.1093/nar/gkaa913 (2021).

112 Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Research* **31**, 371-373, doi:10.1093/nar/gkg128 (2003).

113 Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).

114 Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252, doi:10.1093/bioinformatics/btz859 (2020).

115 Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**, 33-36, doi:10.1093/nar/28.1.33 (2000).

116 Makarova, K. S., Sorokin, A. V., Novichkov, P. S., Wolf, Y. I. & Koonin, E. V. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biology Direct* **2**, 33, doi:10.1186/1745-6150-2-33 (2007).

117 Kristensen, D. M. *et al.* A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**, 1481-1487, doi:10.1093/bioinformatics/btq229 (2010).

118 Computing, R. R: A language and environment for statistical computing. *Vienna: R Core Team* (2013).

119 Wickham, H. & Wickham, M. H. The ggplot package. *Google Scholar. http://ftp. uni-bayreuth. de/math/statlib/R/CRAN/doc/packages/ggplot. pdf* (2007).

120 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

121 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

122 Minkin, I., Patel, A., Kolmogorov, M., Vyahhi, N. & Pham, S. in *Algorithms in Bioinformatics.* (eds Aaron Darling & Jens Stoye) 215-229 (Springer Berlin Heidelberg).

123 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645 (2009).

124 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

125 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

126 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

127 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME journal* **11**, 2864-2868 (2017).

128 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590-D596, doi:10.1093/nar/gks1219 (2013).

129 Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Research* **32**, 1363-1371, doi:10.1093/nar/gkh293 (2004).

130 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948, doi:10.1093/bioinformatics/btm404 (2007).

131 Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **7**, e1002195, doi:10.1371/journal.pcbi.1002195 (2011).

132 Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490-2492, doi:10.1093/bioinformatics/bty121 (2018).

133 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

134 Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309-D314, doi:10.1093/nar/gky1085 (2019).

135 Davis, J. J. *et al.* The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Research* **48**, D606-D612, doi:10.1093/nar/gkz943 (2020).

136 Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473, doi:10.1186/s12859-019-3019-7 (2019).

137 Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* **72**, e108, doi:https://doi.org/10.1002/cpbi.108 (2020).

138 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods*, doi:10.1038/s41592-022-01488-1 (2022).

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

140 Gilchrist, C. L. M. & Chooi, Y. H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics*, doi:10.1093/bioinformatics/btab007 (2021).

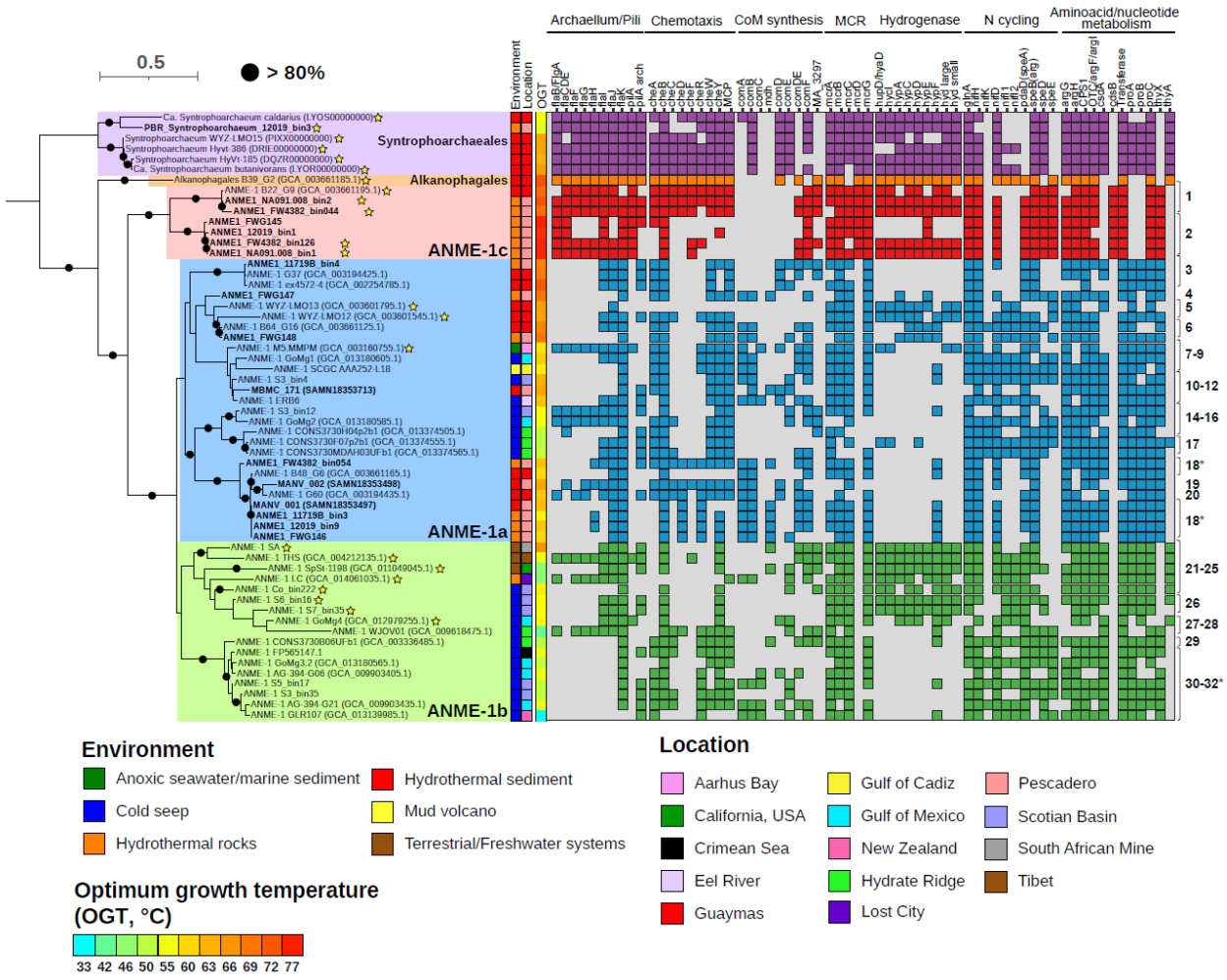


Figure 1. Phylogenomic tree and lineage differentiation of the ANME-1 (*Methanophagales*) order. In bold genomes retrieved from the South Pescadero Basin. The color bars indicate from left to right: environment, location, predicted optimum growth temperature (OGT, in °C) and a genomic comparison of some metabolic features commented on in the text. Numbers on the far right indicate the assigned species for each MAG after dereplication and the stars after the name denote MAGs containing at least the large subunit of a NiFe hydrogenase. For species 18, the corresponding MAGs were clustered in two separated groups. Dereplicated species 30-32 were very closely related and the corresponding genomes appeared intercalated in the tree. Species 13 does not appear in the tree since its genome (ANME-1 Co_bin174) did not reach the minimum number of 24 marker genes for the analysis. Black circles indicate bootstrap support values over 80%. The scale bar represents the number of nucleotide substitutions per site.

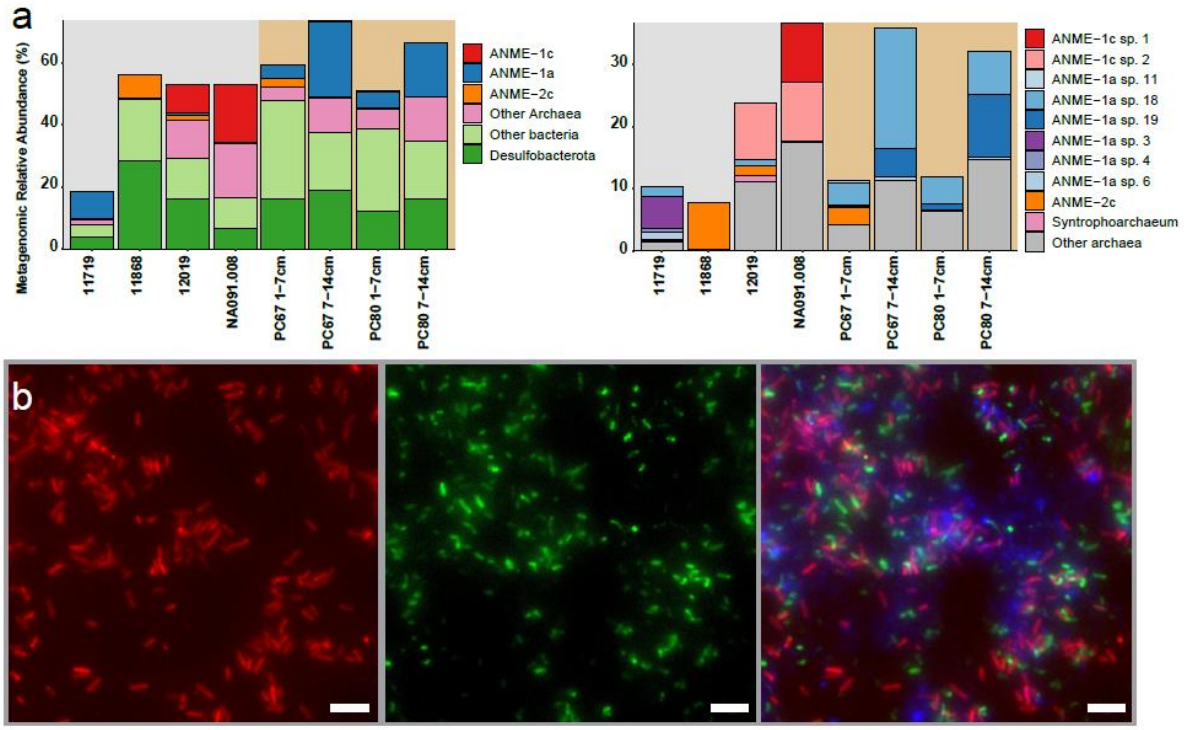


Figure 2. Distribution and morphology of ANME-1 in S. Pescadero sediment and rock samples. a) Relative abundance of MAGs from ANME and other bacteria and archaeal lineages (**left**). Genomic abundance for archaea at the species level is on the right, highlighting a variation ANME-1 lineages in rocks and sediments (**right**). Color background indicates rock (grey) or sediment (brown) samples. The total abundance does not reach 100%, since unmapped reads are not included. Note the scale of the vertical axis is different between panels. **b)** Fluorescence *in situ* hybridization of ANME-1 cells recovered from rock sample NA091.008. Cells targeted by the general ANME-1-350 probe are shown in red, **left**). Cells targeted by the general bacterial probe 338 are in green, **middle**). A composite overlay showing bacteria (green), ANME-1 cells (red) and DAPI staining of all microbial cells in blue (**right**). Scale bar = 5 µm.

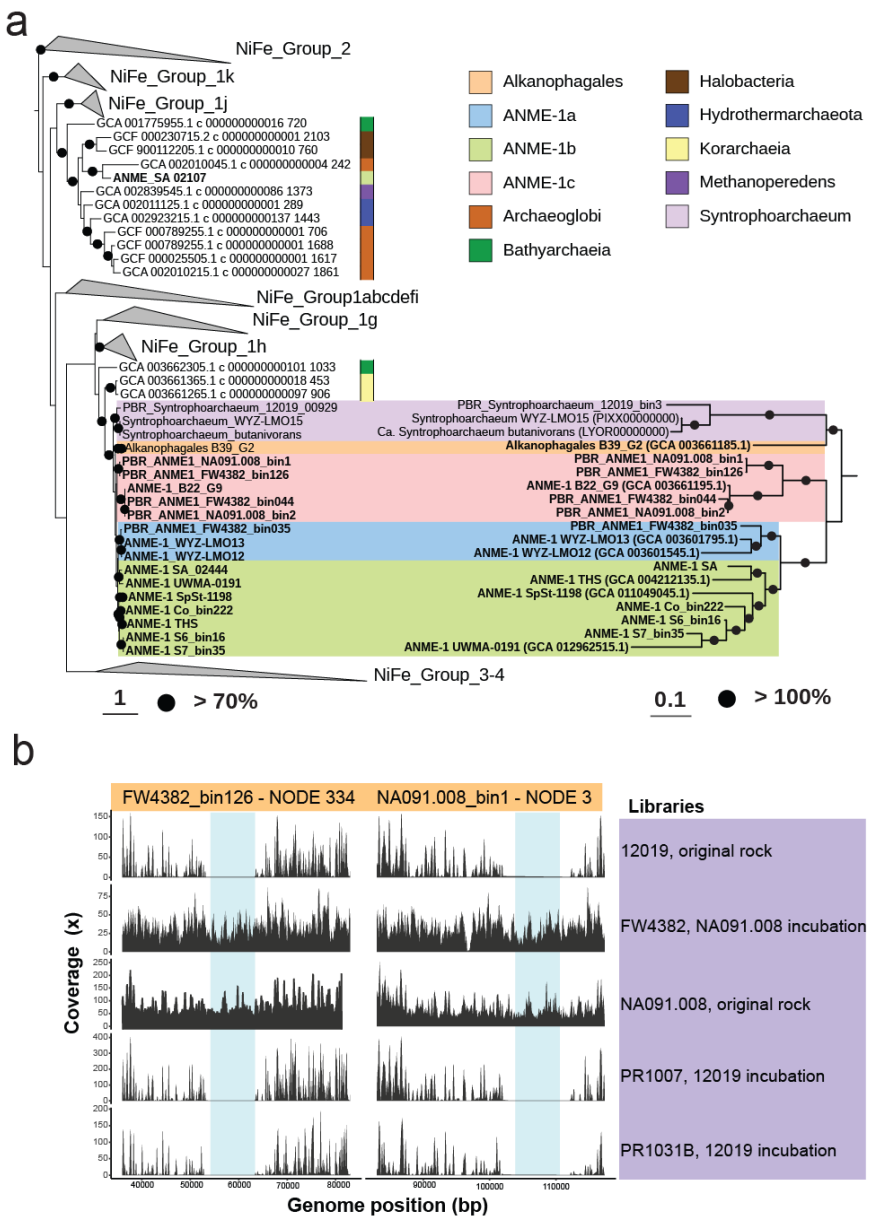


Figure 3. Vertical inheritance and differential loss of hydrogenases across ANME-1. a) Phylogenetic tree of the large subunit of the NiFe hydrogenase present in ANME-1 genomes (left) and the corresponding phylogenomic tree of those genomes (right). A few NiFe hydrogenases of ANME-1 genomes were also affiliated with NiFe Group 3 and 4 (not shown, see Supplementary Figure 5). Color bar/backgrounds indicates the phylogenetic affiliation of hydrogenases of interest. Black circles indicate bootstrap support values over 70% (left) and equal to 100% (right). Scale bars represent the number of amino acid substitutions per site. **b)** Read coverage distribution of the hydrogenase operon of ANME-1c genomes FW4382_bin126 and NA091.008_bin1. Metagenomic read libraries are indicated on the right. The blue shade indicates where the hydrogenase operon is located within the corresponding contig.

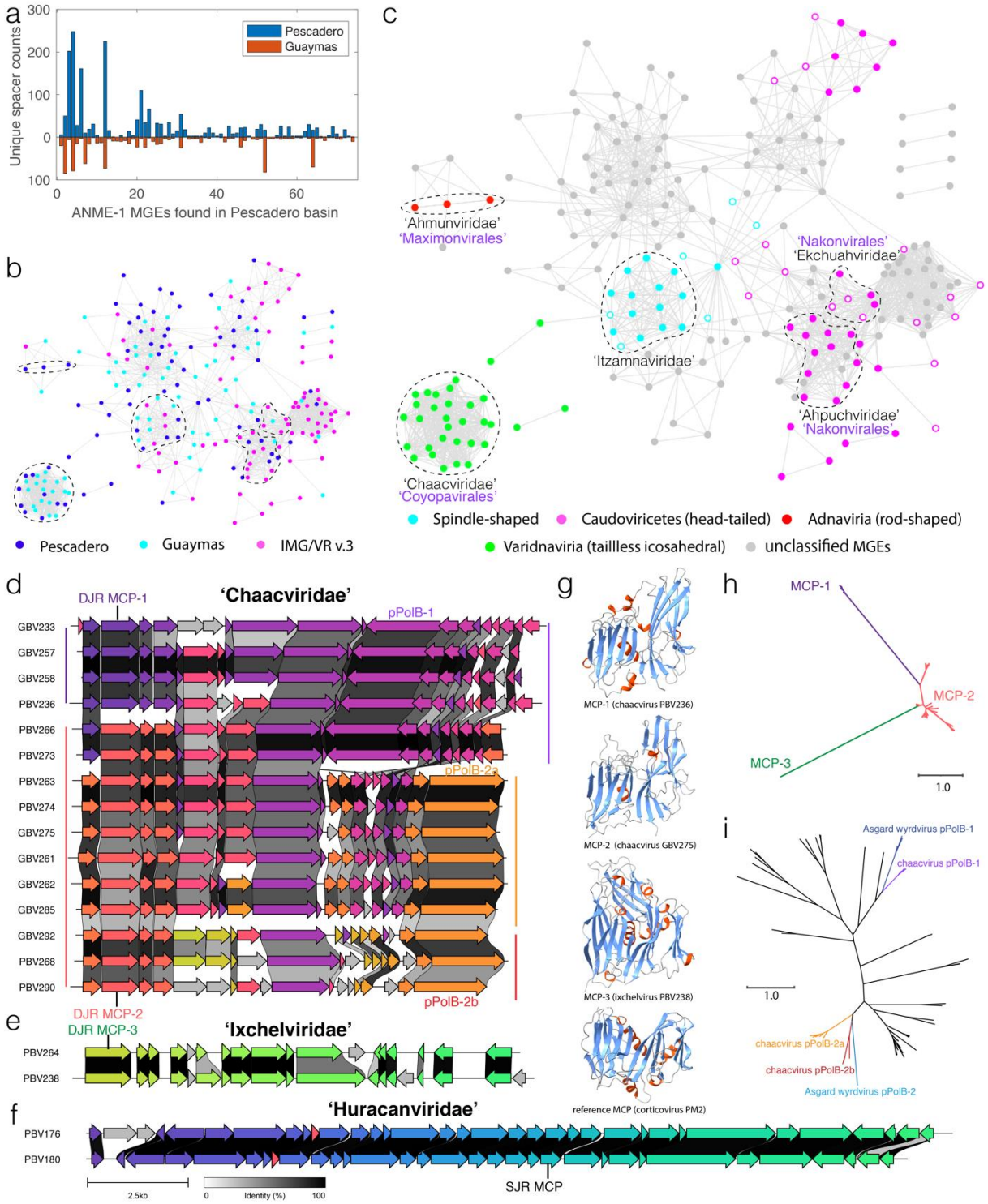


Figure 4. Expansive ANME-1 mobilome includes 16 new viral families and new types of major capsid proteins. a) Histograms showing the number of CRISPR spacers from S. Pescadero and Guaymas basin metagenomes matching the S. Pescadero ANME-1 MGEs. **b)**. Gene sharing network of diverse ANME-1 MGEs of different origins. **c**, ANME-1 MGEs, exhibited in the same network as b, are found to encompass major archaeal virus diversity and non-viral elements. Solid/open circles indicate viral assemblies

with/without identifiable MCPs. In **b** and **c**, dashed lines encircle five new viral families containing complete genome representatives. The newly proposed names of viral families (black) and orders (purple) are indicated in c. **d-f**. Gene synteny of three new families of tailless icosahedral viruses targeting ANME-1. Different colors indicate 83 different protein groups. The scale bar and precent identity shading are indicated in f. **g**. alphafold2-predicted structures of newly discovered DJR MCPs in ANME-1 viruses shown in panels d and e. Blue indicates β barrels, and red α helices. **h.** Maximum-likelihood analysis of new MCP families indicates their long evolutionary distances. **i**. Maximum-likelihood analysis of PolB found in different clades of the *Chaacviridae* are related to two clades of spindle-shaped Wyrdviruses targeting Asgard archaea. Abbreviations: MCP, major capsid protein. DJR, double jellyroll. SJR, single jellyroll. pPolB, Protein-primed DNA Polymerase family B.

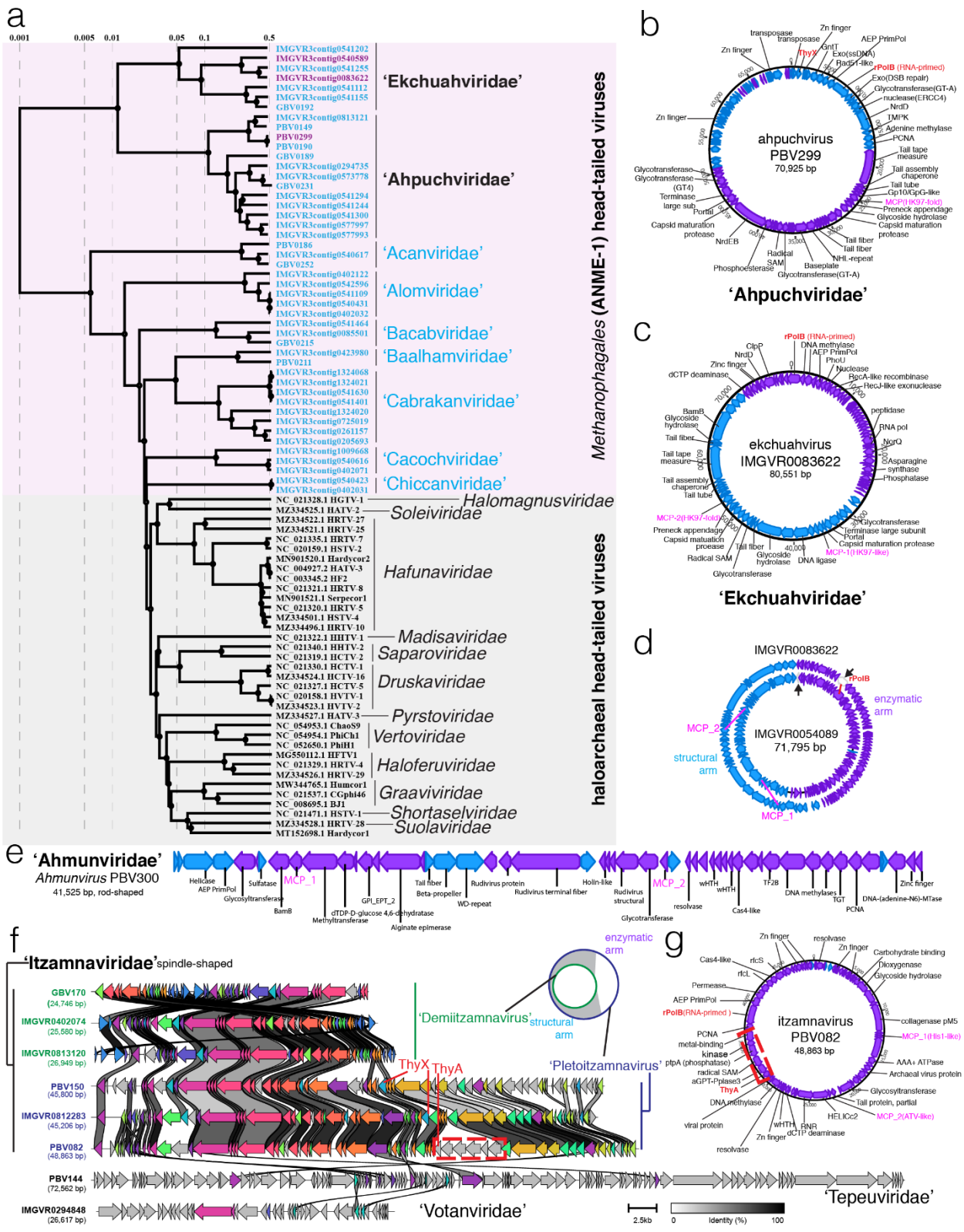


Figure 5. ANME-1 viral genomes encode complex structures. a) Evolutionary division between head-tailed viruses targeting ANME-1 and haloarchaea revealed by global proteome-based phylogenetic

analyses. ANME-1 viruses with complete circular genomes are highlighted in purple, those with unconfirmed completeness are in blue. **b)** and **c)** genome organization and gene content of the complete genomes representing two new families of ANME-1 head-tailed viruses. Blue and purple shading represents forward and reverse strands, respectively. MCP, PolB and ThyX genes are highlighted in pink and red. **d**) circular alignment of the two genomes of *Ekchuahviruses*. Black arrowheads indicate the original contig start/end sites in each assembly. **e**) Gene content of the complete linear genome of a representative of the rod-shaped virus family *Ahmunviridae*. **f**) Gene synteny of three new families of spindle-shaped viruses targeting ANME-1, where complete, circularized genomes of Itzamnaviridae were found to occur in two genome sizes, where Demiitzamnavirus representatives align with a section of the larger Pletoitzamnavirus genomes (illustrated on the top-right). Different colors indicate 76 different protein groups. Grey shading denotes singletons. The scale bar and percent identity shading are indicated in the bottom right. **g**) Gene content of the complete linear genome of a representative of the spindle-shaped virus family *Itzamnaviridae*. Dashed red box in f and g highlights an example of a multi-gene cluster insertion. In d and f, structural arm denotes the genome fraction where all viral structural genes reside; enzymatic arm denotes the fraction where no structural genes but only enzyme-encoding genes reside.

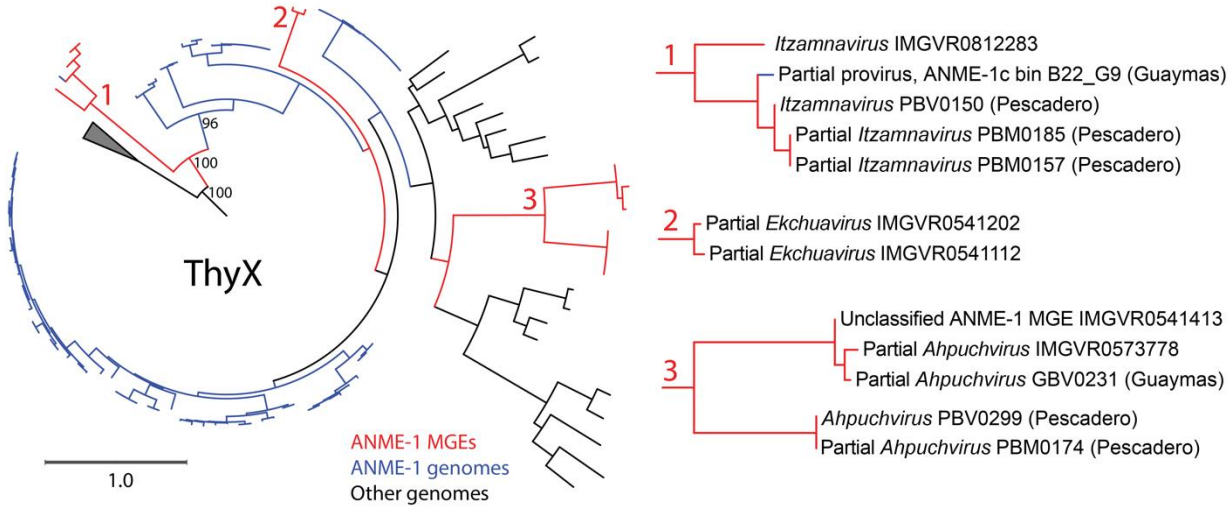
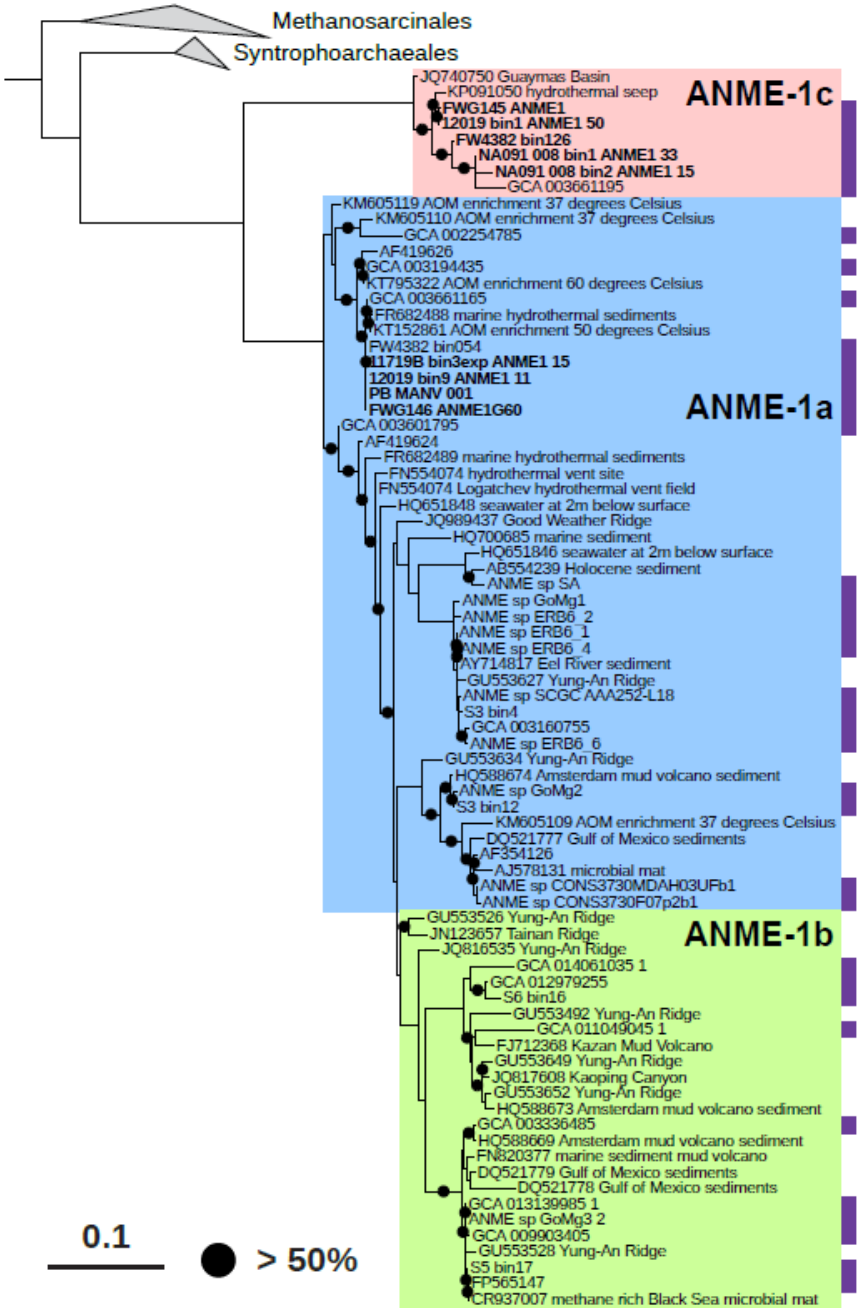
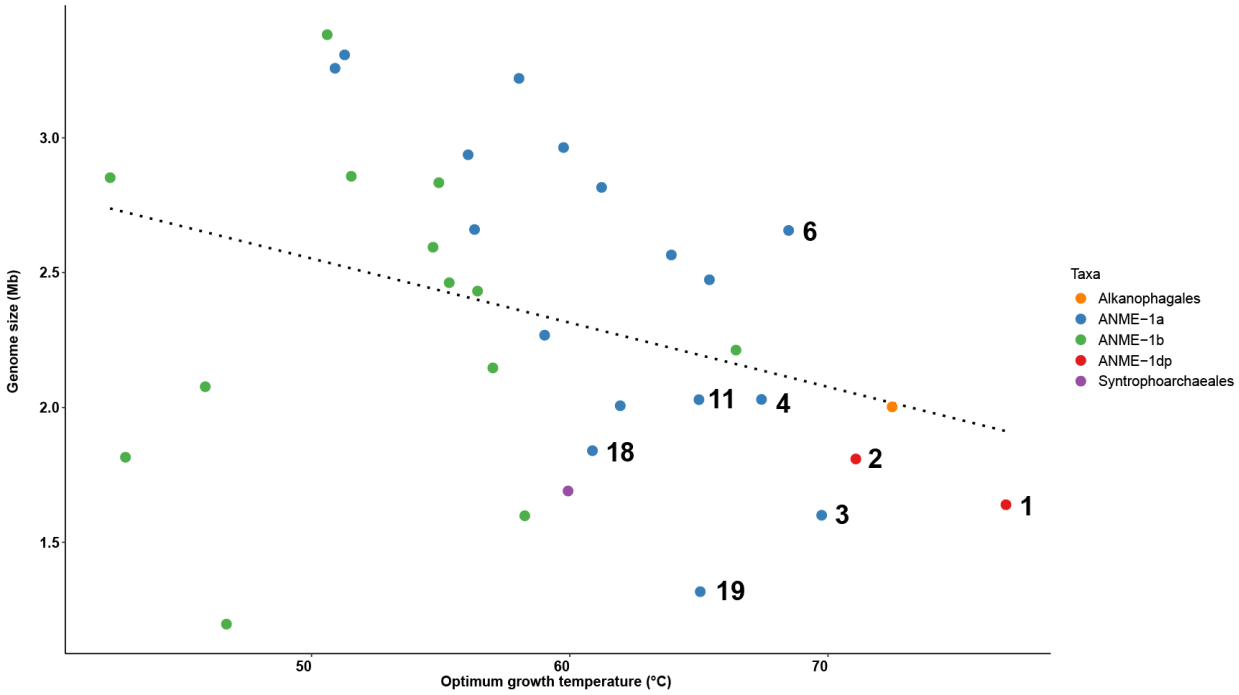


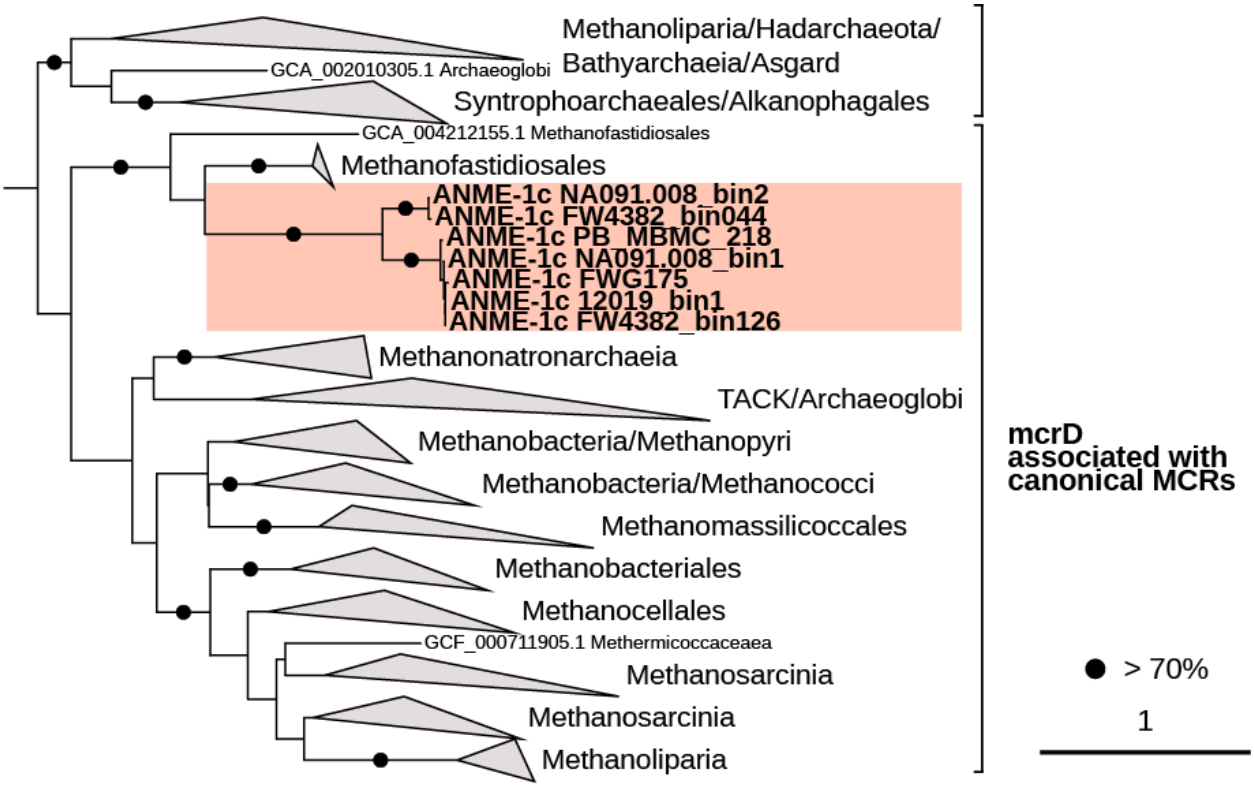
Figure 6. A viral origin of thymidylate synthase in ANME-1. Maximum likelihood analysis of ThyX related to ANME-1 encoded ThyX proteins, with expanded views of ThyX from ANME-1 viruses on the right. Legends for the branch colors for ThyX from mobile genetic elements (MGEs) and ANME-1 genomes are indicated below the main phylogenetic tree.



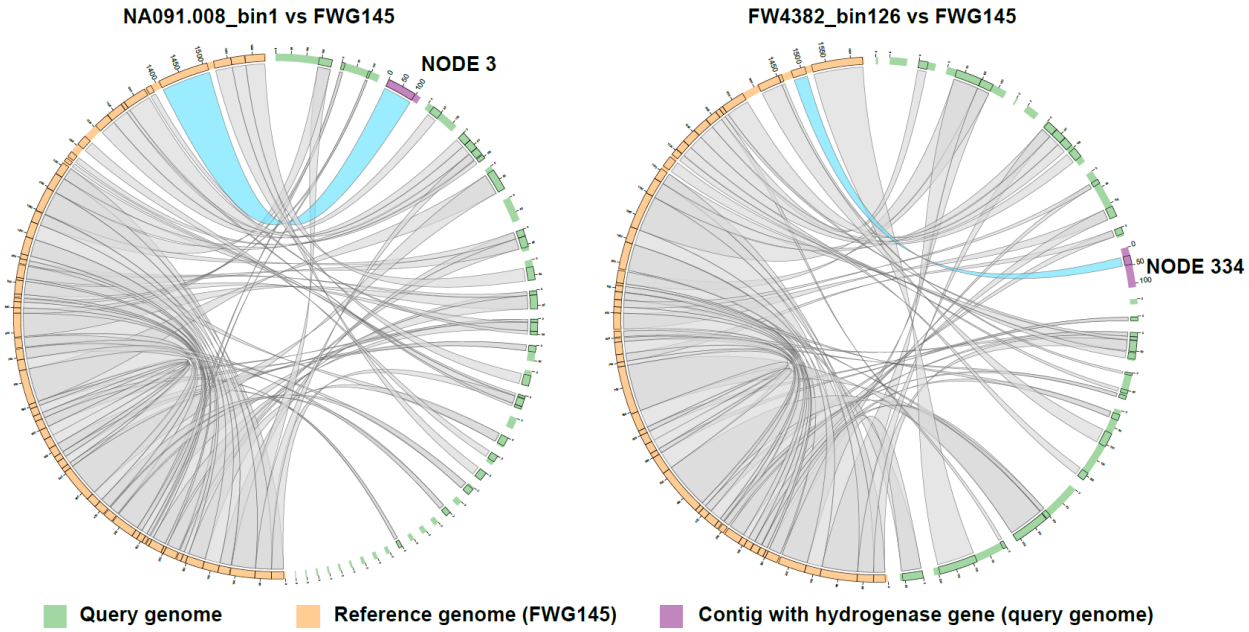
Supplemental Figure 1. 16S rRNA gene phylogeny for the ANME-1 clade (*Methanophagales*). Color shading highlights the three main groups of ANME-1 archaea. The purple bars note 16S rRNA gene sequences retrieved from MAGs shown in Figure 1. Sequences retrieved from Pescadero MAGs are in bold. Bootstrap values over 50% are indicated with a black circle. Scale bar indicates the number of nucleotide substitutions per site.



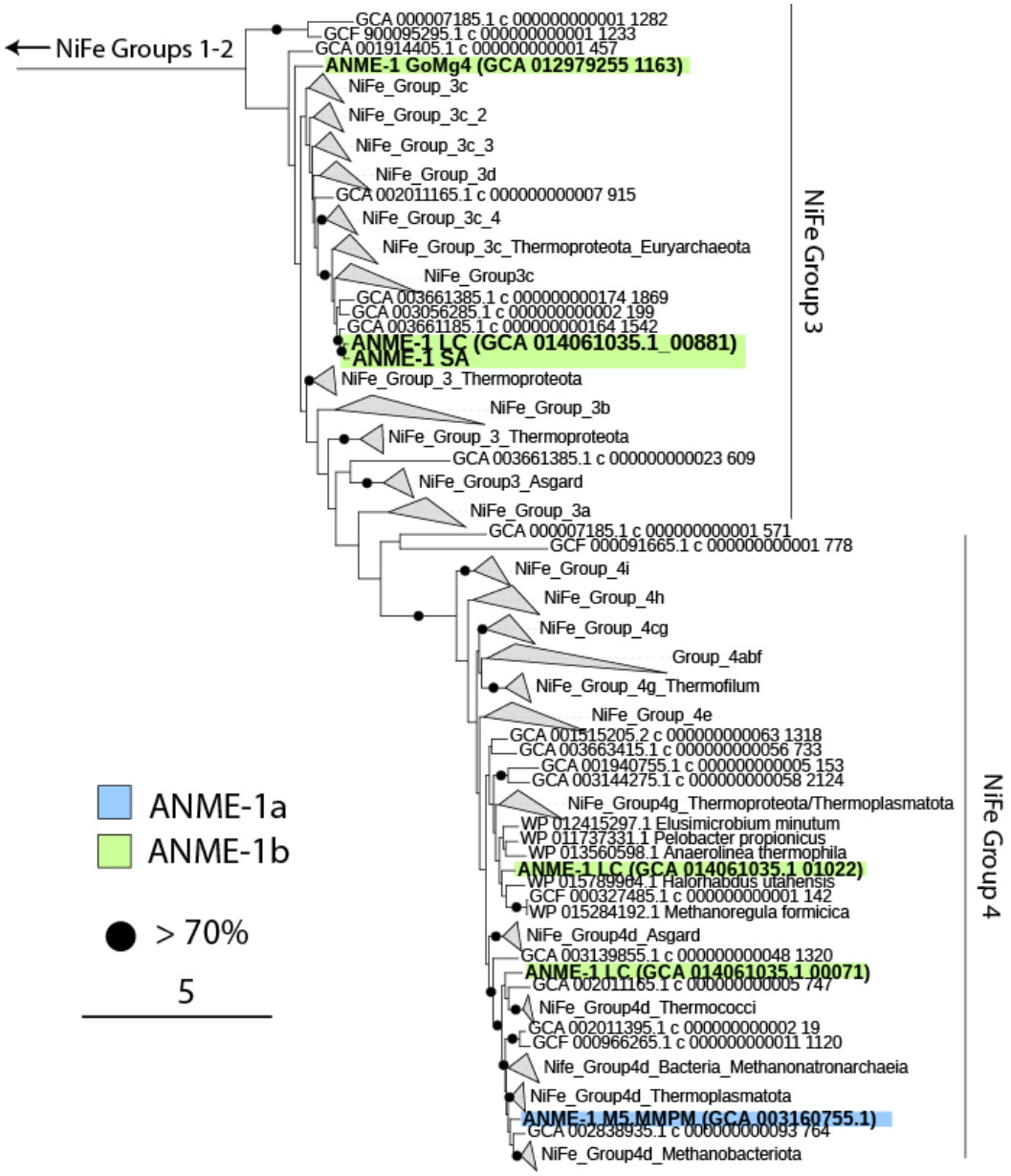
Supplemental Figure 2. Correlation between estimated genome size (in Mb and after calculation considering contamination and completeness see Material and Methods) and the predicted optimum growth temperature (°C). Each point represents the average values for one ANME species after dereplication, except in the case of Syntrophoarchaeales where the point represent the average values for the whole family. Numbers indicate the corresponding ANME-1 species present in S. Pescadero. Color indicates the corresponding taxonomy. Dotted line indicates the regression model (R^2=0.11).



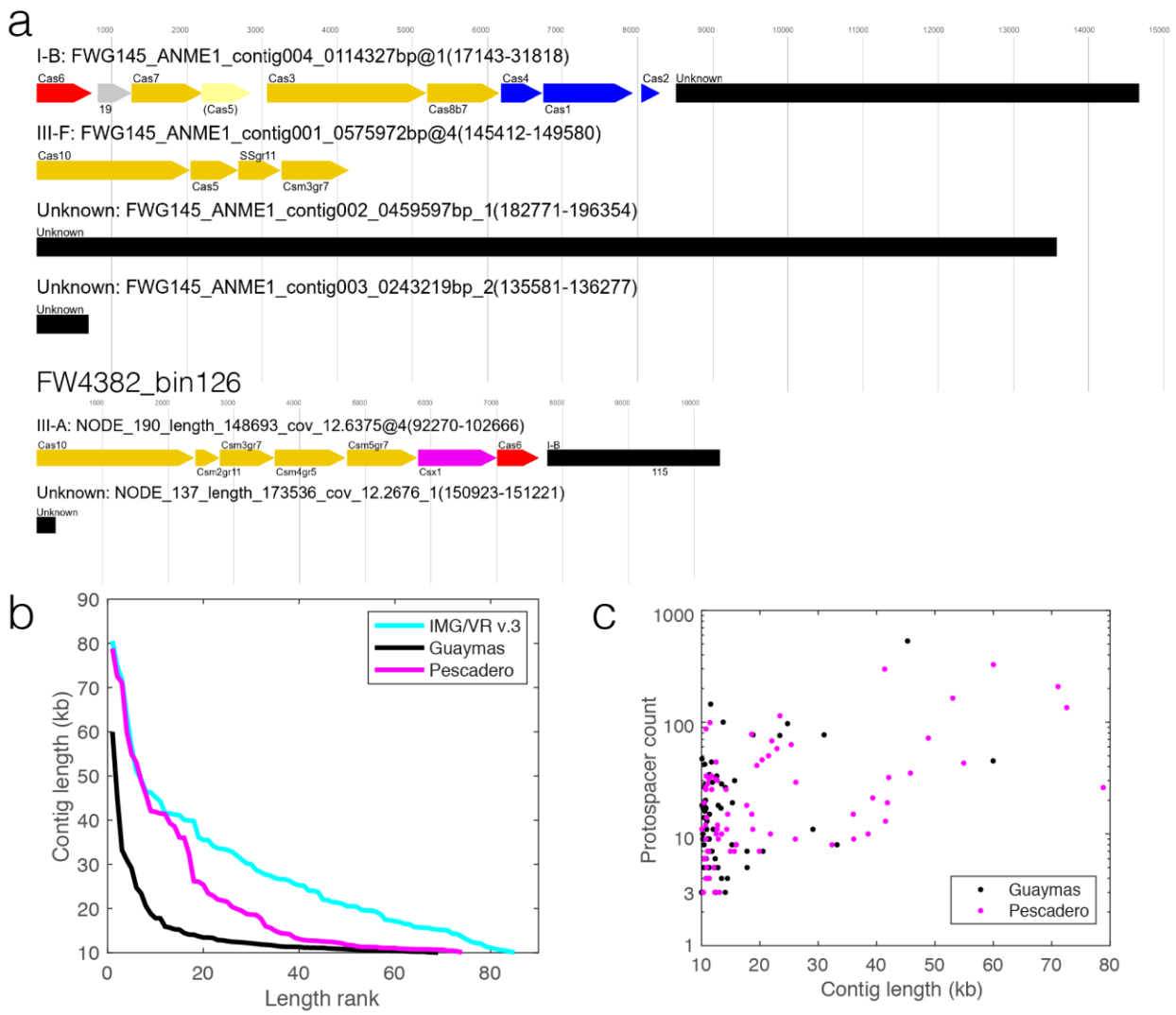
Supplemental Figure 3. Phylogenetic tree of McrD genes from archaea, including the McrD in ANME-1 genomes (only found in ANME-1c). Black circles indicate bootstrap support values over 70%. Scale bar represents the number of amino acid substitutions per site.



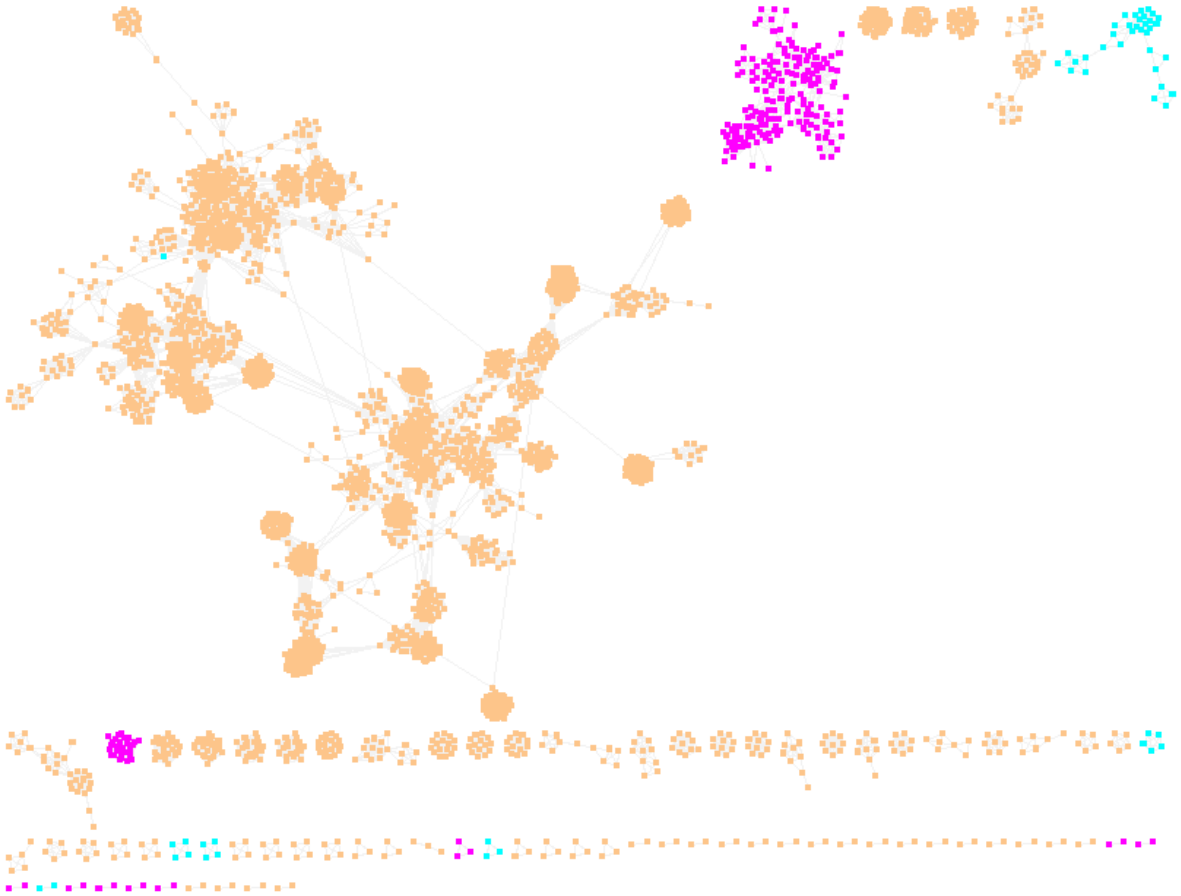
Supplementary Figure 4. Circos plot comparing homologous regions of the ANME-1c genomes, NA091.008_bin1 and FW4382_bin126 (both with hydrogenase operons) to the predicted completed genome FWG175 that was assembled as a single contiguous scaffold and belongs to the same species. Contigs corresponding to the query genomes (NA091.008, FW4382_bin126) are marked in green and contigs from genome FWG175 are in orange. The contig containing the hydrogenase operon is shown in purple and the corresponding homology sections between the reference and query genomes are highlighted in blue. The region between these homology sections corresponds to the hydrogenase operon that was not detected in genome FWG145.



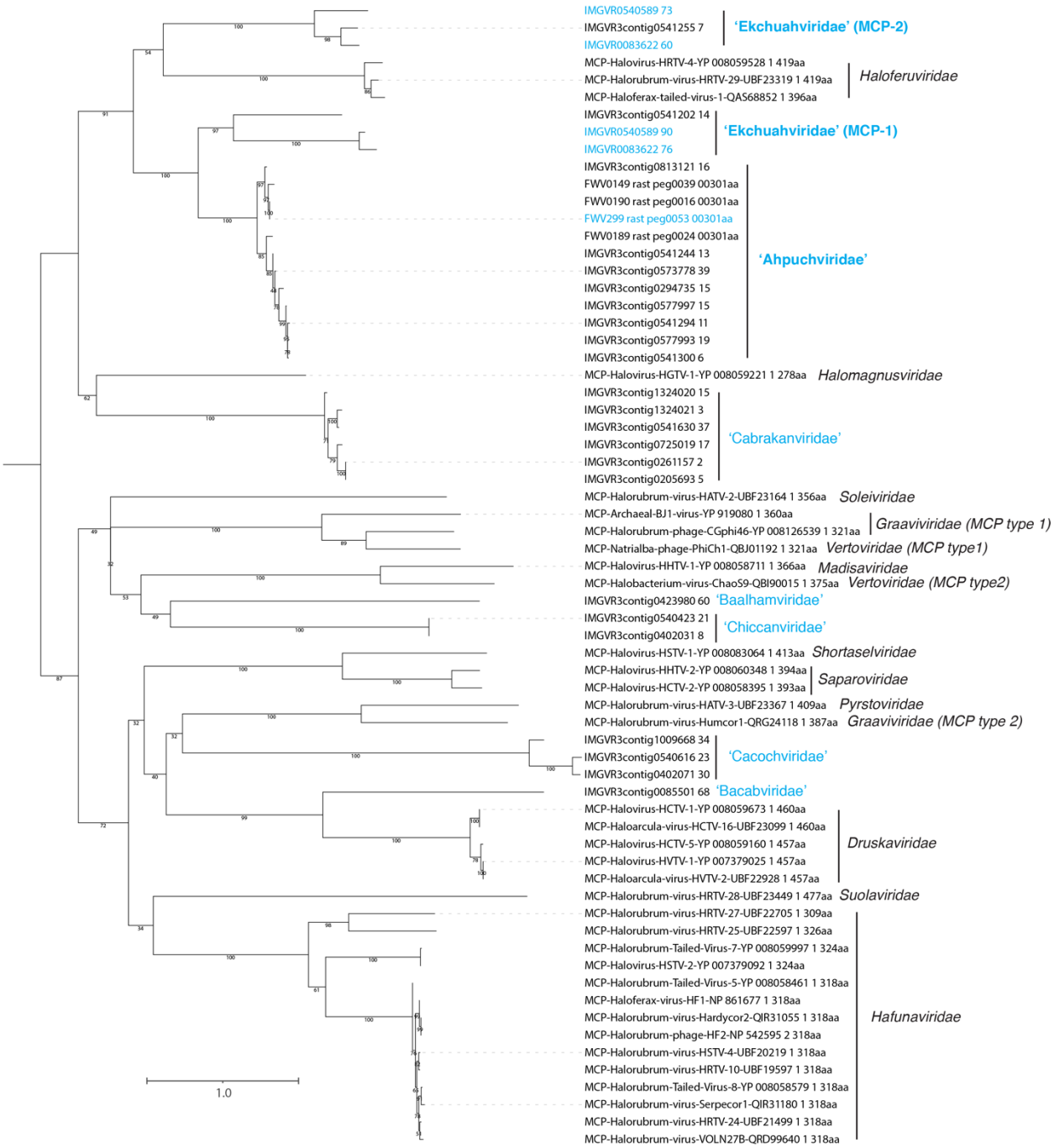
Supplementary Figure 5. Phylogenetic tree of the large subunit of the NiFe hydrogenase present in ANME-1 genomes associated with NiFe Groups 3 and 4. The green and blue shading indicates the taxonomic identity of the ANME-1 MAG containing the corresponding hydrogenase. Black circles indicate bootstrap support values over 70% (left). The scale bar represents the number of amino acid substitutions per site.



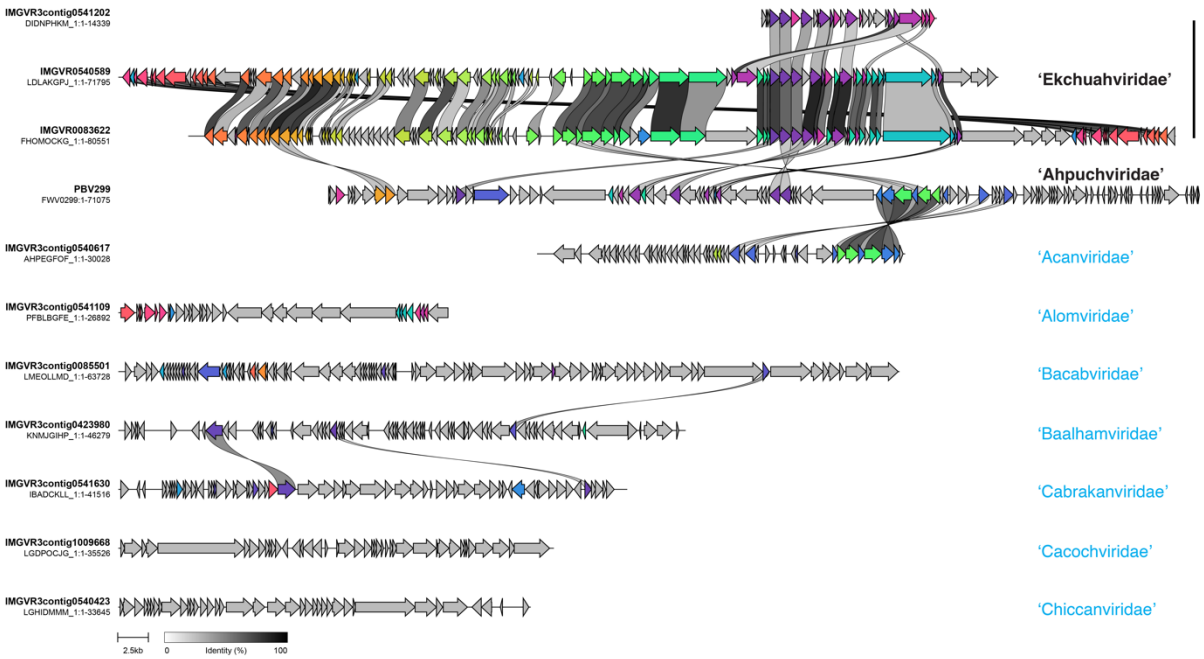
Supplemental Fig 6. **Features of ANME-1 CRISPR/Cas and spacer-mobilome mapping. a**) CRISPR/Cas features in the two most contiguous ANME-1c MAGs characterized using CCtyper. Black bars indicate CRISPR arrays. **b**) Contig lengths of all ANME-1 mobile genetic elements (MGEs) found in this study. Note that the distribution does not necessarily indicate completeness as IMG/VR v.3 is more enriched in head-tailed viruses (with genomes sized up 80 kb) whereas the Pescadero/Guaymas basin dataset contains many tailless icosahedral viruses whose genomes are sized around 10 kb. **c**) Distribution of protospacers within the ANME-1 mobile elements found in S. Pescadero and Guaymas basins.



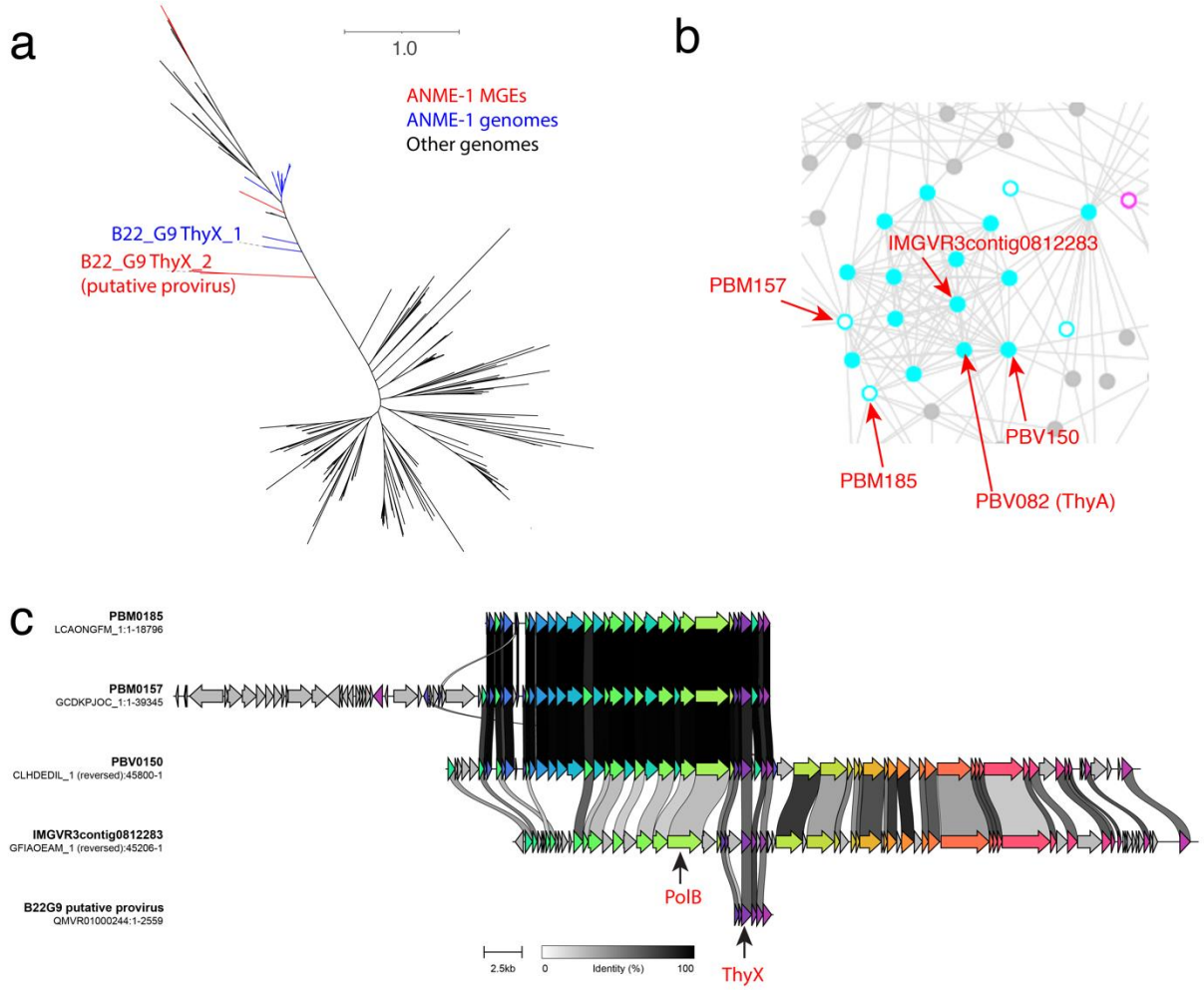
Supplemental Fig 7. **Gene-sharing networks produced via vCONTACT2 indicate that all ANME-1 mobile genetic elements (Magenta) are well distinguished from the known *haloarchaeal* viruses, or *Haloviruses*, (Blue) and other viruses with known hosts (orange).**



Supplementary Figure 8. Maximum likelihood analyses of head-tailed viruses. Blue indicates ANME-1 virus families and haloarchaeal virus families are shown in black italics. MCPs from complete genomes of ANME-1 viruses are indicated in blue, and their respective families in bold.



Supplementary Figure 9. Sequence alignment of representatives of ANME-1 viruses. Colors indicates protein families encoded by at least 2 representative viral genomes here. Grey indicates singleton proteins without apparent homologs. Scale bars for protein identity scores and genome sizes are indicated at the bottom. Viral contig names and sizes are indicated on the left side and their respective family and Candidatus (Ca.) family names are indicated on the right.



Supplementary Figure 10. Spindle-shaped viruses encode ThyX at the root of ANME-1 ThyX. a, unrooted phylogeny suggests that ANME-1 *thyX* may have evolved from *thyX* genes originated in ANME-1 viruses. The two versions of of ThyX in the ANME-1c bin B22_G9 is highlighted. b, two MGEs containing thyX genes are found to be highly related to the spindle-shaped viruses with identified MCPs. c, Sequence alignment indicate that the two MGEs and a genome contig that encode ThyX are all fragmented from spindle-shaped viruses.