

MINIREVIEW – Incubator

# Genomic diversity, lifestyles and evolutionary origins of DPANN archaea

Nina Dombrowski<sup>1,2</sup>, Jun-Hoe Lee<sup>3</sup>, Tom A. Williams<sup>4</sup>, Pierre Offre<sup>1</sup> and Anja Spang<sup>1,3,\*</sup>

<sup>1</sup>NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, NL-1790 AB Den Burg, The Netherlands, <sup>2</sup>Department of Marine Science, University of Texas at Austin, Marine Science Institute, 750 Channel View Drive, Port Aransas, TX 78373, USA, <sup>3</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, P.O. Box 596, Husargatan 3, SE-75123 Uppsala, Sweden and <sup>4</sup>School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, Bristol BS8 1TQ, UK

\*Corresponding author: NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, Landsdiep 4, NL-1790 AB Den Burg, The Netherlands. Tel: +31 (0)222 369 526; E-mail: [anja.spang@nioz.nl](mailto:anja.spang@nioz.nl)

**One sentence summary:** We review current knowledge on the diversity and genomic potential of the only recently discovered enigmatic and potentially symbiotic DPANN archaea, discuss insights gained from functional studies of host–symbiont systems involving DPANN archaea as well as summarize controversies regarding the placement of the various DPANN lineages in the tree of life and thus the role of this putative radiation in the early evolution of life on Earth.

Editor: Daniel Tamarit

## ABSTRACT

Archaea—a primary domain of life besides Bacteria—have for a long time been regarded as peculiar organisms that play marginal roles in biogeochemical cycles. However, this picture changed with the discovery of a large diversity of archaea in non-extreme environments enabled by the use of cultivation-independent methods. These approaches have allowed the reconstruction of genomes of uncultivated microorganisms and revealed that archaea are diverse and broadly distributed in the biosphere and seemingly include a large diversity of putative symbiotic organisms, most of which belong to the tentative archaeal superphylum referred to as DPANN. This archaeal group encompasses at least 10 different lineages and includes organisms with extremely small cell and genome sizes and limited metabolic capabilities. Therefore, many members of DPANN may be obligately dependent on symbiotic interactions with other organisms and may even include novel parasites. In this contribution, we review the current knowledge of the gene repertoires and lifestyles of members of this group and discuss their placement in the tree of life, which is the basis for our understanding of the deep microbial roots and the role of symbiosis in the evolution of life on Earth.

**Keywords:** DPANN; archaea; evolution; symbiosis; genomics; metabolism

## INTRODUCTION

Antoni van Leeuwenhoek is often credited for the discovery of bacteria, which he visualized for the first time through a microscope of his own design in 1676 (Leeuwenhoek 1677). It took another 300 years before Carl Woese and George Fox inferred

that methane-producing microorganisms (i.e. methanogens)—thought to be bacteria—in fact represent members of a separate domain of life, now referred to as Archaea (Woese and Fox 1977; Woese, Kandler and Wheelis 1990). Burgeoning methodologies for the sequencing of nucleic acids and the

Received: 19 September 2018; Accepted: 7 January 2019

© FEMS 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

reconstruction of phylogenies showed that, in addition to methanogens, archaea included a range of extremophilic organisms: halophiles, acidophiles and hyperthermophiles. The assumption that most Archaea inhabit extreme environments of limited global significance prevailed until the early 1990s, which saw the first reports of archaeal organisms detected in marine waters (DeLong 1992; Fuhrman, McCallum and Davis 1992). Various lineages of archaea are now known to be globally distributed and prevalent in marine pelagic and benthic ecosystems as well as soils (DeLong, Pace and Kane 2001) and were shown to be of utmost importance for our understanding of the origin of eukaryotes (Spang et al. 2015; Eme et al. 2017; Zaremba-Niedzwiedzka et al. 2017). Surprisingly, archaea were not known to include pathogens or endosymbionts and were for a long time thought to comprise predominantly free-living organisms. This view changed with the discovery of the ultrasmall ectosymbiotic archaeon *Nanoarchaeum equitans*, whose growth is obligately dependent on its archaeal host *Ignicoccus hospitalis* (Huber et al. 2002). In agreement with its host dependency, *N. equitans* was found to have a small reduced genome encoding a limited set of metabolic functions and was suggested to be the first member of a separate archaeal phylum referred to as Nanoarchaeota (Huber et al. 2002; Waters et al. 2003). Since then, the application of single-cell and metagenomic approaches has helped to gradually refine our picture of archaeal phylogenetic diversity (Adam et al. 2017; Spang, Caceres and Ettema 2017) and unveiled the genomes of a large amount of additional nanosized and/or genome-reduced archaeal lineages (Fig. 1, Supplementary Fig. S1, Supplementary File 1, Supporting Information) (Rinke et al. 2013; Castelle et al. 2015; Castelle and Banfield 2018; Probst et al. 2018). In initial analyses, these genome-reduced archaea were suggested to form a monophyletic and deep-branching archaeal superphylum, which included Nanoarchaeota and was collectively referred to as DPANN—an acronym for the different phyla known at the time, the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (Rinke et al. 2013). In contrast to *N. equitans*, which was enriched in a co-culture with its hosts and has been studied extensively (Huber et al. 2002; Waters et al. 2003; Jahn et al. 2004, 2008; Burghardt et al. 2009; Giannone et al. 2015; Mohanty et al. 2015; Heimerl et al. 2017), most of our knowledge of the biology of other DPANN archaea is derived from genomic data (Castelle and Banfield 2018). Therefore, much has to be learned about these enigmatic archaea and the coming years will certainly witness a reappraisal of the extent of symbiotic interactions involving archaea.

In this contribution, we review the history of the discovery of DPANN archaea, our current understanding of their metabolic potential, features of their reduced genomes, their host-symbiont interactions as well as their phylogenetic diversity and evolutionary history. This will provide a foundation for our understanding of the symbiotic nature of these archaea and can guide prospective efforts to unveil both the functional importance of these extremely diverse but understudied organisms and their role in the early evolution and diversification of Archaea.

## THE DISCOVERY OF ARCHAEA WITH SMALL GENOMES AND CELL SIZES

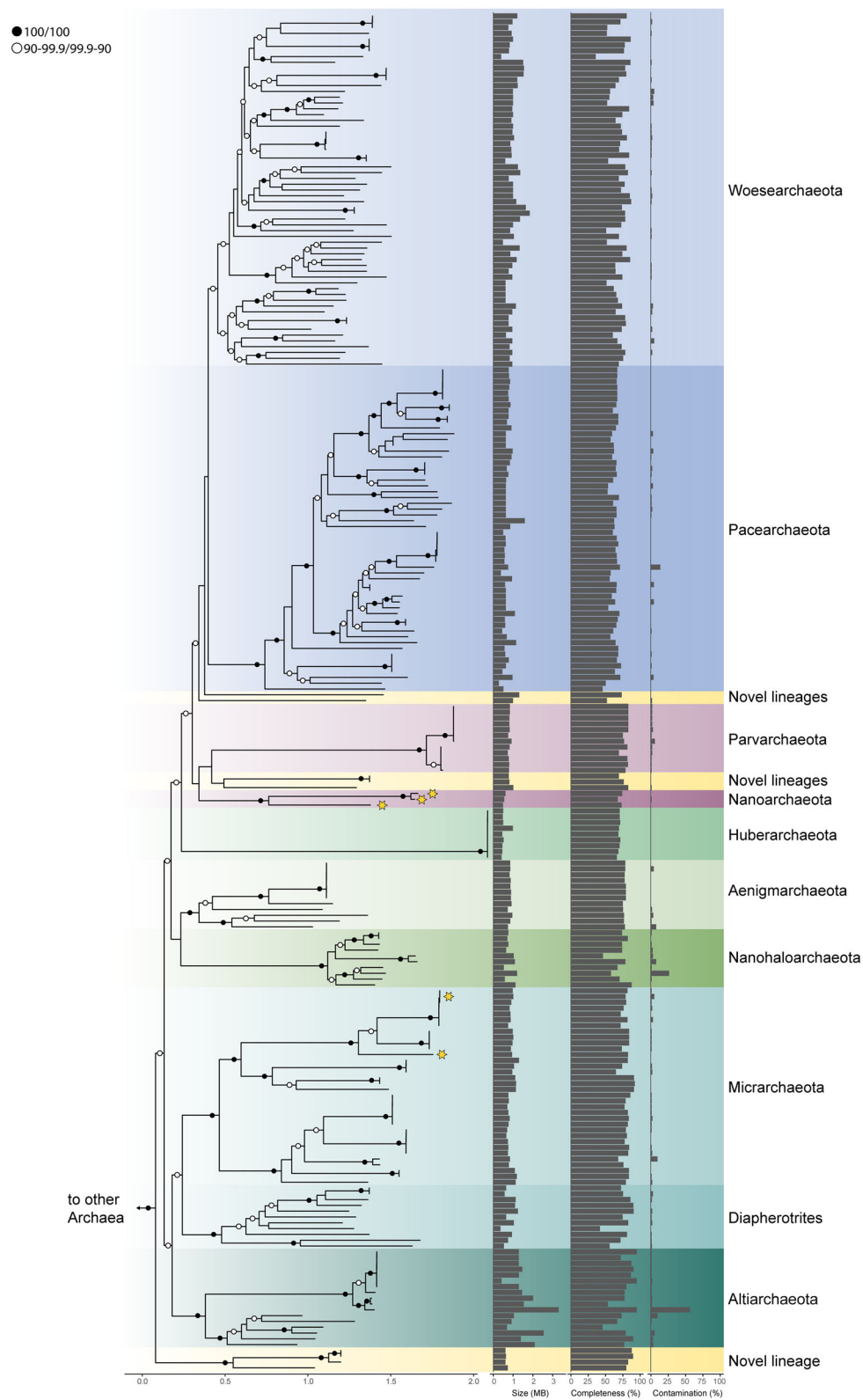
The first member of the Nanoarchaeota, *N. equitans*, was discovered in an enrichment culture established from samples collected at a hydrothermal site north of Iceland (Huber et al. 2002).

This organism is characterized by small cells of just 400 nm in diameter and encodes one of the smallest known archaeal genomes (0.49 Mb) (Huber et al. 2002; Waters et al. 2003) (Fig. 1, Supplementary Table S1, Supplementary Fig. S1, Supplementary File 1, Supporting Information). A few years later, slightly larger genomes (0.64–1.08 Mb) of other ultrasmall archaea with cell volumes as low as 0.009  $\mu\text{m}^3$  were recovered from an acid mine drainage and initially referred to as archaeal Richmond Mine acidophilic nanoorganisms (ARMAN), but later renamed to Parvarchaeota and Micrarchaeota (Comolli et al. 2009; Baker et al. 2010; Baker and Dick 2013). An additional lineage of nanosized archaea (0.1–0.8  $\mu\text{m}$ ) named Nanohaloarchaeota was found in hypersaline environments (Ghai et al. 2011; Narasingarao et al. 2012; Rinke et al. 2013; Vavourakis et al. 2016) and initially thought to comprise a sister lineage of Haloarchaea (Narasingarao et al. 2012). Subsequent analyses indicated alternative positions for this lineage and raised the possibility that Nanohaloarchaeota belong to DPANN (Rinke et al. 2013); however, their phylogenetic placement in the Euryarchaeota is still debated (Aouad et al. 2018) (see sections below). The use of single-cell genomics unveiled an even larger diversity of small archaea in brackish/freshwater and hydrothermal environments and led to the description of Diapherotrites and Aenigmarchaeota (Rinke et al. 2013), the latter of which were originally known as DSEG archaea (Takai et al. 2001). DHVE-5 and 6 archaeal groups, which were first described by 16S rRNA gene analyses (Takai and Horikoshi 1999; Durbin and Teske 2012), were renamed to Pearchaeota and Woearchaeota, upon the recovery of the first metagenome-assembled genomes (MAGs) from an aquifer (Castelle et al. 2015). Pearchaeota and Woearchaeota seem to represent the most ubiquitously distributed lineages within the DPANN and have been detected in groundwater (Castelle et al. 2015), freshwater lakes (Ortiz-Alvarez and Casamayor 2016), ocean sediments (Durbin and Teske 2010) and hydrothermal vents (Takai and Horikoshi 1999). Recently, Woearchaeota have even been identified in permafrost samples and the human microbiome (Shcherbakova et al. 2016; Koskinen et al. 2017). The most recent addition to the DPANN superphylum is the Huberarchaeota, for which MAGs have been recovered from a CO<sub>2</sub>-driven geyser (Probst et al. 2018). While the Altiarchaeota have been first placed within the Euryarchaeota, several phylogenetic analyses indicate that they may affiliate with the DPANN archaea (Spang, Caceres and Ettema 2017; Castelle and Banfield 2018; Castelle et al. 2018) (see sections below). Their first representative, *Cand. Altiarchaeum hamiconexum* (formerly known as SM1 Euryarchaeon), was discovered in a sulfidic spring (Probst et al. 2013, 2014), where it forms ‘string-of-pearls’, a biofilm community where the archaeon inhabits the interior and a filamentous bacterial species dominates the exterior of the ‘pearls’ (Rudolph, Wanner and Huber 2001). However, not all Altiarchaeota appear to form such biofilms as suggested by the recovery of altiarchaeotal genomes from river sediments, springs and lakes (Bird et al. 2016). Altogether, DPANN archaea represent an extremely diverse putative superphylum that comprises more than 10 phylum-level lineages most of which share small cell and genome sizes as common features (Fig. 1).

## WHAT IS ENCODED BY THE SMALL GENOMES OF DPANN ARCHAEA?

### The metabolic potential of DPANN archaea

While the metabolic potential of DPANN archaea appears to vary considerably both between and within DPANN lineages, most of



**Figure 1.** Phylogenetic tree displaying currently available taxonomic diversity of DPANN genomes. Tree was generated using the RP15 pipeline published previously in Zaremba-Niedzwiedzka et al. (2017) and is based on maximum-likelihood analyses in IQ-tree using the LG + C60 + F model of evolution. SH-values and bootstraps were calculated using a SH-like approximate likelihood ratio test (Guindon et al. 2010) and ultrafast bootstrap support values (Minh et al. 2013), respectively. Black and white circles indicate support values of 100/100 and 99.9-90/90-99.9, respectively. Scale bars indicate the average number of substitutions per site. The phylogeny represents bins available at NCBI before 23 March 2018 and indicates the genome size (for closed genomes), the bin completeness and contamination (for MAGs). The stars highlight genomes from DPANN members existing in co-cultures. Corresponding information for these genomic bins is available in Supplementary Table S1 (Supporting Information), the tree including taxa names is available as Supplementary Fig. S1 (Supporting Information) and the tree file is provided in Supplementary Data 1 (Supporting Information).

these organisms are characterized by sparse metabolisms with limited catabolic and anabolic capabilities indicating that at least some members may represent obligate symbionts (Castelle and Banfield 2018; Castelle et al. 2018). In this section, we will briefly summarize the main metabolic features predicted for DPANN archaea and refer the interested reader to excellent publications offering comprehensive overviews for more details (Castelle et al. 2015, 2018; Castelle and Banfield 2018; Chen et al. 2018; Liu et al. 2018).

#### Energy metabolism

Most of the currently available DPANN genomes lack genes encoding known components of electron transport chains capable of generating a proton/sodium-motive force, which drives ATP production via a membrane-bound ATP synthase (Castelle and Banfield 2018; Castelle et al. 2018). In contrast, genes encoding enzymes involved in the formation of fermentation products including lactate, formate, ethanol, acetate (Castelle and Banfield 2018) as well as butyrate (Chen et al. 2018) are detected in various genomes of DPANN archaea, suggesting that many of the currently known members are anaerobes with the potential to use substrate-level phosphorylation as main mode of energy conservation. Representatives of the Altiarchaeota could constitute a notable exception (Probst et al. 2014; Bird et al. 2016). These organisms have been hypothesized to rely on a putative ferredoxin-dependent complex 1-like oxidoreductase to generate a proton-motive force and drive ATP synthesis via a membrane-bound ATP synthase. Although several DPANN archaea encode subunits of putative ATP synthases, these protein complexes might not always be functional as suggested by the structural and biophysical investigations of the  $A_3B_3$  core complex of *N. equitans* (Mohanty et al. 2015). Yet, some genomes assigned to Parvarchaeota and Micrarchaeota have been reported to encode putative components of an aerobic electron transport chain and a canonical A-type ATP synthase besides the fermentation pathways common to most DPANN, which indicate an ability for both aerobic and anaerobic metabolism (Baker et al. 2010; Castelle and Banfield 2018; Castelle et al. 2018; Chen et al. 2018).

#### Catabolism

Some fermentation pathways predicted to operate in DPANN archaea are fed by pyruvate and others by acetyl-CoA. Homologs of enzymes converting pyruvate to acetyl-CoA (i.e. pyruvate dehydrogenase and pyruvate-ferredoxin oxidoreductase) appear to be broadly distributed across the DPANN superphylum (Castelle and Banfield 2018; Castelle et al. 2018). Yet, components of this minimal energy metabolism (i.e. pyruvate-metabolizing enzymes) were absent in some DPANN genomes assigned to Woesearchaeota (Castelle et al. 2015), Mamarchaeota (Castelle and Banfield 2018) and Nanoarchaeota (Podar et al. 2008), raising questions about the potential source of energy in these organisms: e.g. is acetyl-CoA directly taken up by these organisms? In contrast, various organisms across the DPANN superphylum were predicted to operate additional catabolic pathways leading to the production of pyruvate and acetyl-CoA (Narasimgarao et al. 2012; Castelle et al. 2015, 2018; Castelle and Banfield 2018; Chen et al. 2018; Liu et al. 2018). This includes the Embden–Meyerhof–Parnas pathway, an incomplete Entner–Doudoroff pathway, the beta-oxidation pathway and a RubisCO-dependent nucleoside degradation pathway (Sato, Atomi and Imanaka 2007; Aono et al. 2015), suggesting that at least some DPANN archaea have the ability to conserve energy from the oxidation of hexoses, fatty acids and nucleosides. Further inferred catabolic capabilities include the depolymerization of oligosaccharides and polysaccharides in Nanohaloarchaeota, Micrarchaeota, Pacearchaeota,

Parvarchaeota and Woesearchaeota (Narasimgarao et al. 2012; Castelle et al. 2015; Chen et al. 2018); the depolymerization of proteins in Diapherotrites, Micrarchaeota, Parvarchaeota and Woesearchaeota (Castelle et al. 2015; Youssef et al. 2015; Chen et al. 2018; Liu et al. 2018); the utilization of glycerol in Parvarchaeota (Chen et al. 2018); the degradation of polyhydroxybutyrate in Diapherotrites (Youssef et al. 2015); and, tentatively, the oxidation of ferrous iron in Parvarchaeota (Chen et al. 2018). Catabolic pathways inferred to be encoded by altiarchaeotal genomes generate  $CO_2$  instead of organic acids. The former is suggested to derive from the oxidation of C1 carbon compounds, i.e. carbon monoxide and formate (Probst et al. 2014; Bird et al. 2016).

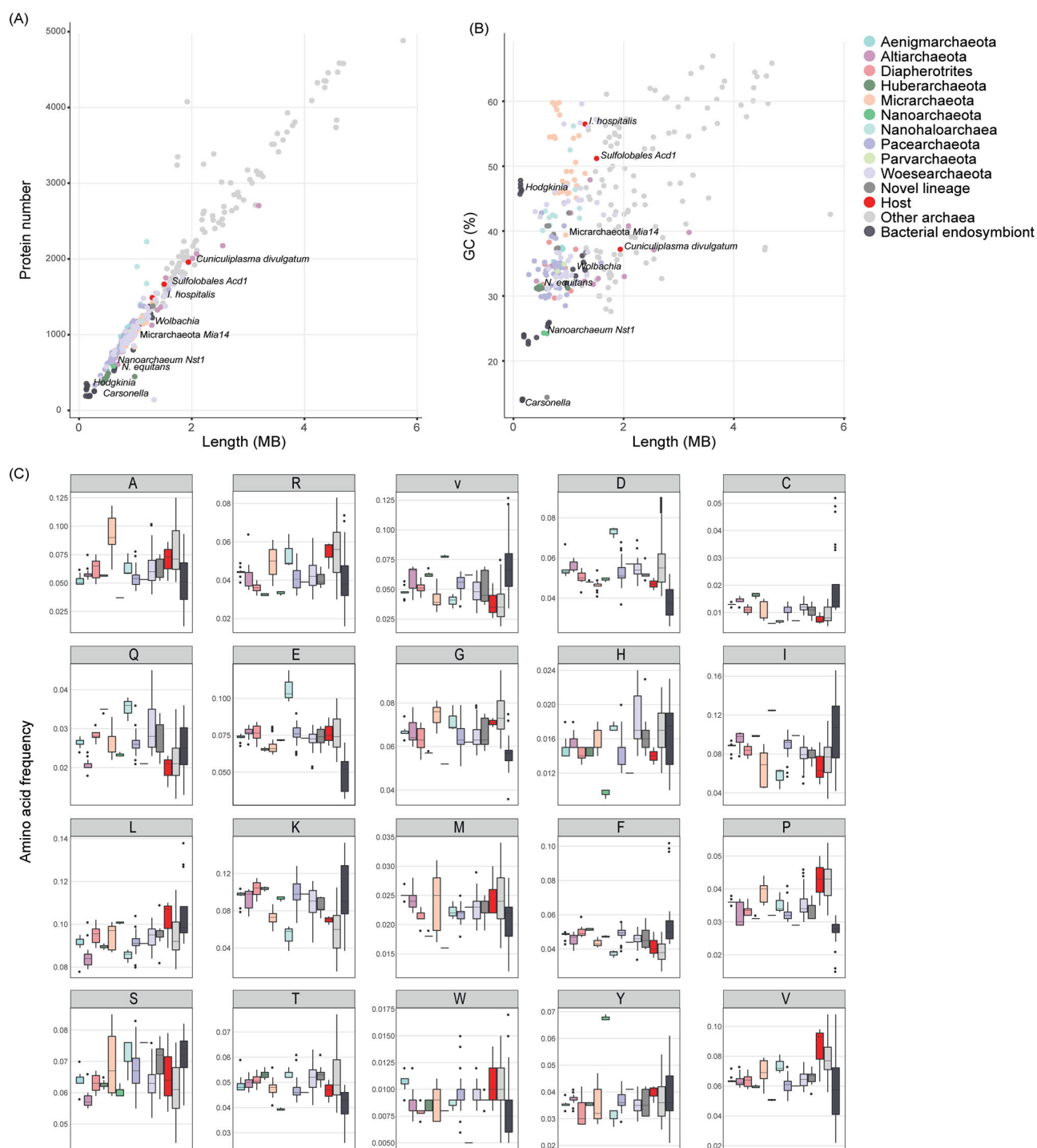
#### Anabolism

Many genomes of the DPANN archaea are characterized by the absence of genes encoding the enzymes of canonical, primary biosynthetic pathways for amino acids, purines, pyrimidines, lipids and vitamins (Castelle and Banfield 2018; Castelle et al. 2018). Furthermore, many genomes lack genes encoding known components of the tricarboxylic acid (TCA) cycle and both oxidative and non-oxidative variants of the pentose phosphate pathway. In spite of these reduced biosynthetic capabilities, some members of the DPANN superphylum may synthesize some macromolecule building blocks (e.g. purine and pyrimidine biosynthetic pathways may operate in some Diapherotrites, Micrarchaeota, Aenigmarchaeota, Nanoarchaeota and Woesearchaeota), and some genomes assigned to Micrarchaeota and Parvarchaeota were recently reported to encode near-complete TCA cycles (Baker et al. 2010; Krause et al. 2017; Chen et al. 2018). Contrasting with the apparent dependence of most DPANN organisms on reduced carbon compounds, representatives of the Altiarchaeota are autotrophs using a modified Wood–Ljungdahl pathway to fix carbon dioxide (Probst et al. 2014).

The general paucity of genes encoding homologs of known metabolic enzymes in at least some genomes of DPANN organisms is remarkable considering that comprehensive inventories of metabolic pathways can be reconstructed from genomes of bacterial symbionts, such as *Mycoplasma pneumoniae* (Yus et al. 2009; Wodke et al. 2013) or *M. genitalium* (Karr et al. 2012), which possess, like DPANN archaea (Fig. 2A), few protein-coding genes. Current observations suggest that many members of the DPANN archaea must acquire multiple essential metabolites externally, consistent with the experimentally validated host-associated lifestyle of some DPANN organisms (see below) (Huber et al. 2002; Jahn et al. 2008; Golyshina et al. 2017; Krause et al. 2017). Nevertheless, the extent of the dependency of DPANN on the provision of essential metabolites remains unclear since many of the proteins encoded in their genomes have no known function, raising the question of whether DPANN organisms encode novel enzymes driving canonical or entirely new metabolic pathways.

#### Other genomic features of DPANN archaea

In spite of these reduced metabolic capabilities and in contrast to findings in some bacterial endosymbionts (Moran and Bennett 2014), reductive genome evolution in DPANN archaea does not seem to coincide with the loss of the informational processing genes as most members of this group encode core informational processing machineries (Castelle and Banfield 2018). However, the occurrence of split genes encoding some proteins involved in informational processing, such as reverse gyrase and tRNA synthetases described in Nanoarchaeota and Micrarchaeota (Waters et al. 2003; Randau, Pearson and Söll 2005; Randau et al. 2005; Baker et al. 2010), or the diphthamide



**Figure 2.** General genome characteristics of DPANN archaea. Dot plots displaying total protein numbers (A) and GC content (B) versus genome size for DPANN genomes as compared to other archaeal representatives as well as selected endosymbiotic bacteria (see also Supplementary Table S1, Supporting Information). (C) Average amino acid frequencies of members of the DPANN archaea compared to their hosts, other archaea and bacterial endosymbionts (see also Supplementary Table S2, Supporting Information). Frequencies are shown for each individual amino acid, which are listed in their single letter code. Boxplot shows the median, the first and third quartiles, the upper/lower whiskers that extend from the hinge to the largest/smallest value no further than  $1.5 \times$  of the interquartile range from the hinge as well as outliers, which are represented as dots. Boxplots were generated with R v3.5.0 using the package ggplot2.

biosynthesis protein Dph1/2 in some members of the Parvarchaeota, Pacearchaeota, Woesearchaeota and Altiarchaeota (Narrowe et al. 2018), may represent signs of reductive genome evolution in some of these lineages. Initially, split genes in Nanoarchaeota were proposed to represent an ancestral fea-

ture (Di Giulio 2006), thought to coincide with an early divergence of Nanoarchaeota (Waters et al. 2003; Podar et al. 2013). However, more recent analyses of additional members of Nanoarchaeota have led to the suggestion that split genes originated later in only some representatives of Nanoarchaeota with

particularly reduced genomes (Podar et al. 2013). This view is supported by the observation that some of the proteins encoded by split genes in *N. equitans*, such as DNA polymerase I, topoisomerase I and alanyl-tRNA synthetase, are encoded by a single gene in the close relative *Nanoarchaeota* Nst1 (Podar et al. 2013) and do not reveal a deep-branching position in phylogenetic analyses (Andersson, Sarchfield and Roger 2005; Furukawa et al. 2017). On the other hand, the primase found in various DPANN including *N. equitans*, *Parvarchaeum* ARMAN-5 and *Cand. Nanosalarum* sp. (Raymann et al. 2014) consists of only one subunit encoded by a single gene, while this enzyme consists of a catalytic subunit (PriS) and an accessory subunit (PriL) encoded by two distinct genes in other archaea. While *Micrarchaeota* ARMAN-2 has a canonical archaeal primase encoded by two genes, the atypical primase of most DPANN archaea may represent a putative synapomorphy (a shared derived character) for a subset of DPANN lineages. However, subsequent phylogenetic analyses that include a wider representation of primases encoded by the various recently described DPANN lineages will be necessary to resolve the evolutionary history of this protein family and thereby shed more light onto replication machineries of DPANN archaea.

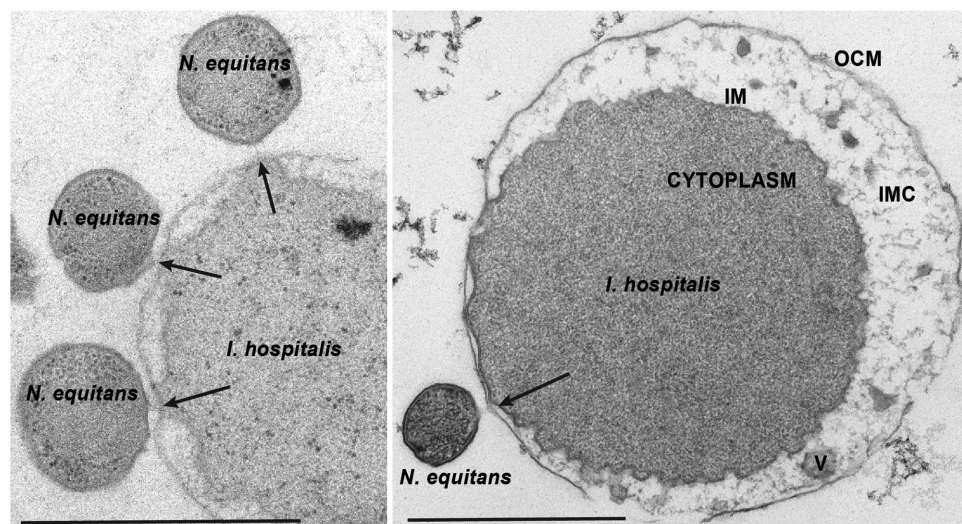
## HOST-SYMBIONT SYSTEMS INVOLVING DPANN ARCHAEA

The limited metabolic capacities of most DPANN archaea described above suggest that various of these organisms have a predominately symbiotic lifestyle. However, to date only few stable co-cultures of DPANN archaea with their hosts have been obtained: *N. equitans* with *I. hospitalis* (Huber et al. 2002; Waters et al. 2003), *Nanoarchaeota* Nst1 (later renamed to *Nanosidianus stetteri*) with *Sulfolobales* Acd1 (Podar et al. 2013; Munson-McGee et al. 2015), *Cand. Nanopusillus acidilobi* with *Acidilobus* sp. 7A (Wurch et al. 2016), *Cand. Micrarchaeota* Mia14 with *Cuniculiplasma divulgatum* PM4 (Golyshina et al. 2017), *Micrarchaeota* (ARMAN-1) A.DKE with *Cuniculiplasma* sp. C.DKE (Krause et al. 2017) and *Cand. Nanoclepta minutus* Ncl-1 with *Zestosphaera tikiterensis* NZ3T (St. John et al. 2018). Furthermore, a potential interaction of *Cand. Huberarchaeum crystalense* with members of the Altiarchaeota was suggested based on co-varying cell abundance profiles and microscopic imaging (Probst et al. 2018). A recent meta-analysis of publicly available archaeal 16S rRNA gene sequences revealed the potential co-occurrence of operational taxonomic units (OTUs) derived from Woesearchaeota with those of Methanomicrobia and Methanobacteria and may indicate interactions between members of these groups (Liu et al. 2018). Interestingly, some protein sequences of DPANN archaea, such as aminoacyl tRNA synthetases of *Nanoarchaeota*, *Parvarchaeota*, *Woesearchaeota* and *Micrarchaeota* and the diphthamide biosynthesis protein Dph5 of *Woesearchaeota*, cluster with eukaryotic homologs in phylogenetic analyses (Andersson, Sarchfield and Roger 2005; Furukawa et al. 2017; Narrowe et al. 2018). It is tempting to speculate that this could imply a symbiotic relationship of some DPANN archaea with eukaryotes, considering that horizontal gene transfer seems to be common between DPANN symbionts and their hosts, such as between *N. equitans* and *I. hospitalis* (Podar et al. 2008) or *Cand. Micrarchaeota* Mia14 and *C. divulgatum* PM4 (Golyshina et al. 2017). In light of these findings, it is interesting to note that a recent study detected potential *Nanoarchaeota*-related organisms in an enrichment culture consisting of a few bacterial species as

well as the protist *Carpediemonas frisia* (Hamann et al. 2017). Finally, 16S rRNA gene sequences assigned to *Pacearchaeota* and *Woesearchaeota* were found to positively correlate with bacterial communities (Ortiz-Alvarez and Casamayor 2016). These findings might suggest that some DPANN archaea interact with either bacterial or eukaryotic partners. However, further analyses are necessary to confirm these observations and determine whether they indicate metabolic interactions or physical associations between specific bacteria or eukaryotes with DPANN. Intriguingly, current analyses suggest that at least some DPANN archaea, for example members of *Nanohaloarchaeota* and *Cand. Iainarchaeum andersonii* (phylum *Diapherotrites*) (Narasimgarao et al. 2012; Youssef et al. 2015) as well as *Altiarchaeota* (Bird et al. 2016), may be capable of leading an independent lifestyle. Altogether, this highlights the multitude of lifestyles found across DPANN archaea and suggests that adaptations to symbiotic growth modes may have evolved several times independently.

The metabolic dependencies of DPANN archaea on their hosts have been studied extensively using *N. equitans* with *I. hospitalis* as a model system. *N. equitans* displays a high host specificity and can only be grown in co-culture with *I. hospitalis*. Furthermore, while *N. equitans* can be separated from its host, its cells appear unable to proliferate (Huber et al. 2002; Jahn et al. 2008). These findings suggest a strong host dependency and the existence of a specific, yet unknown, recognition system. It is still debated whether this interaction is of a mutualistic or parasitic nature as there have not been any experimental setups that study this interaction under natural conditions. In agreement with the limited gene repertoire encoding for core metabolic pathways of *N. equitans* (Waters et al. 2003), this organism seems to obtain various metabolites from *I. hospitalis* rather than from the environment (Jahn et al. 2004, 2008; Hamerly et al. 2015). Both organisms share essentially the same lipid composition (Jahn et al. 2004) and amino acid labeling studies were unable to distinguish their labeling patterns (Jahn et al. 2008), suggesting that *N. equitans* obtains its membrane lipids and amino acids from its host. Furthermore, the likely inactive ATPase of *N. equitans* raises the question of how it obtains ATP (Lewalter and Müller 2006; Mohanty et al. 2015). The unique membrane system of *I. hospitalis*, consisting of an inner- and outer membrane separated by a large periplasmic space and being one of the few examples of an energy-conserving outer membrane, is debated to play an essential role in energy conservation of *N. equitans* (Küper et al. 2010). Specifically, the presence of an ATPase and H<sub>2</sub>:sulfur oxidoreductase in the outer membrane of *I. hospitalis* suggests that ATP is generated in the periplasm of this organism and might be accessible to *N. equitans* (Küper et al. 2010; Mayer et al. 2012). Around 10% of the proteome changes upon interaction of *I. hospitalis* with *N. equitans* including an upregulation of proteins related to energy conservation, cell cycle control and membrane modification (Giannone et al. 2011, 2015). More specifically, the upregulation of an ATP synthase, NiFe-hydrogenase or pyruvate:ferredoxin oxidoreductase might reflect the higher energy demands imposed on the host. While *N. equitans* cells do not appear to contain significant amounts of host proteins (Giannone et al. 2011), it was shown that the overall metabolite pool recovered from the co-culture is less concentrated than in the host alone, implicating that *N. equitans* stimulates the consumption of *I. hospitalis* metabolites (Hamerly et al. 2015).

Even though studies of the interactions between *I. hospitalis* and *N. equitans* have increased our understanding of the associations involving DPANN archaea, it is still unclear how



**Figure 3.** Transmission electron micrographs (TEM) of *I. hospitalis* and *N. equitans* interaction. This TEM of ultrathin sections of *I. hospitalis* and *N. equitans* display the interactions between this host–symbiont system and are based on a slightly modified and updated version of Fig. 4 originally published in Jahn et al. (2008) and kindly provided by Harald Huber. OCM: outer cellular membrane, IMC: intermembrane compartment, IM: inner membrane. Bars: 1  $\mu\text{m}$ .

metabolites are interchanged. In the case of *N. equitans*, two modes of interaction with *I. hospitalis* are proposed: (a) direct periplasmic contact and (b) indirect contact between cells via thin fibers (Junglas et al. 2008) (Fig. 3). The isolation of the contact side has identified hypothetical subunits of the Sec protein translocase complex (SecD), the A1A0-ATPase as well as potential transporters of *I. hospitalis* and *N. equitans*. These components might mediate the transfer of metabolites (Burghardt et al. 2009) or enzymes, such as the fatty acid-CoA ligase encoded by *I. hospitalis*, into the cytoplasmic space of *N. equitans* (Heimerl et al. 2017). Intriguingly, hosts of DPANN archaea can differ dramatically in their membrane architecture, as is the case for *I. hospitalis* and *Sulfolobales Acd1*, the hosts of *N. equitans* and *Nanoarchaeota Nst1*, respectively. While *I. hospitalis* has a double membrane, *Acd1* likely encodes an S-layer (Podar et al. 2013), indicating that even closely related DPANN archaea might have evolved different means to interact with their respective hosts. In line with these observations, a multitude of cell surface structures have been observed in different DPANN lineages. For example, ARMAN-like cells (likely belonging to both Micrarchaeota and Parvarchaeota) might establish direct cytoplasmic contacts via synapse-like or tubular structures or utilize needle-like penetration mechanisms (Comolli et al. 2009; Baker et al. 2010; Comolli and Banfield 2014). Furthermore, hami (‘grappling hook’-like structures) of Altiarchaeota were suggested to be employed for the attachment to other cells (Probst and Moissl-Eichinger 2015). It is however unknown whether these structures also play a role in the interaction between members of the Altiarchaeota and *Cand. Huberarchaeum crystalense* (Probst et al. 2018).

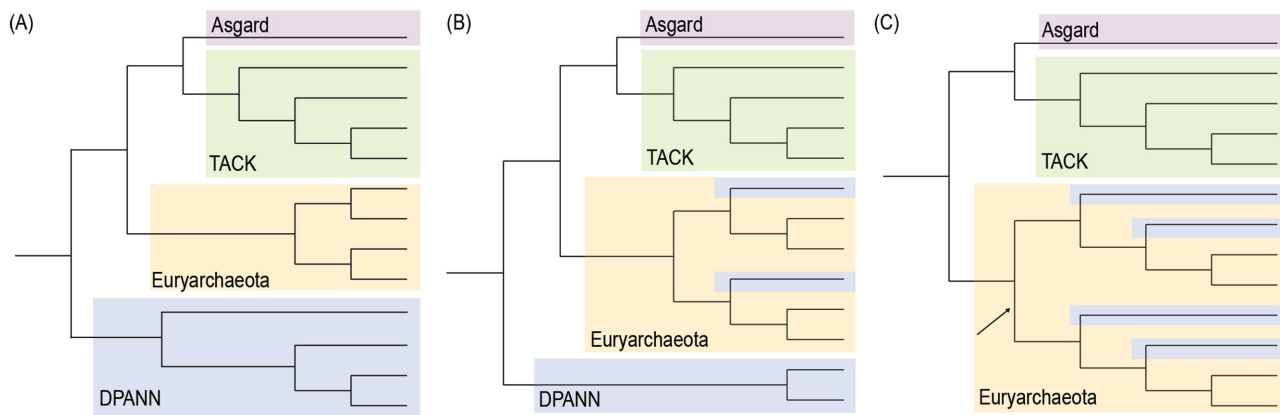
Altogether, studying the symbiotic lifestyle of DPANN archaea has provided us with a better understanding of their metabolic dependencies, multitude of potential hosts and means of interactions. Nevertheless, and also considering the large diversity of DPANN-affiliated lineages, much has to be learned about the interactions these organisms are engaged in. Studying the variety of symbiotic lifestyles found across DPANN archaea using both genomics and microbiological approaches will provide additional insights into the diversity of mechanisms characterizing archaeal symbioses.

## THE EVOLUTION OF DPANN ARCHAEA AND THEIR POSITION IN THE TREE OF LIFE

Ever since the discovery of *N. equitans*, the phylogenetic placement of DPANN lineages in the archaeal tree has been uncertain and is the subject of controversies. Initial phylogenies placed *N. equitans* as an outgroup to the Euryarchaeota and Crenarchaeota, the two main lineages of archaea known at the time (Waters et al. 2003). Subsequent work found some tentative support for a relationship of *N. equitans* with Thermococcales (Brochier et al. 2005), although this placement was not recovered in later analyses (Brochier-Armanet et al. 2008). Improved taxonomic sampling usually helps to resolve the phylogenetic placement of unstable taxa (Graybeal and Cannatella 1998), and the discovery of a broad diversity of additional DPANN lineages (Rinke et al. 2013; Castelle et al. 2015) provided an injection of much-needed genome data to the debate. Various phylogenies of this expanded genome sampling suggested that DPANN archaea form a monophyletic group at the base of the archaeal tree (Rinke et al. 2013; Spang et al. 2013; Castelle et al. 2015; Saw et al. 2015; Spang, Caceres and Ettema 2017; Williams et al. 2017), although alternative topologies have been observed (Petitjean et al. 2015). Thus, none of these analyses has definitively resolved the phylogenetic placement of the different DPANN lineages, hampering our understanding of the metabolic gene repertoire of the last common ancestor of the Archaea as well of the role of symbioses in the evolution and diversification of the Archaea.

### The challenge of determining the phylogenetic placement of symbionts

Based on published analyses, at least three scenarios for DPANN phylogeny seem possible (Fig. 4). First, DPANN might represent a monophyletic clade branching at the base of the Archaea; this position is consistent with most of the published trees that include representatives from all known DPANN groups (Rinke et al. 2013; Spang et al. 2013; Castelle et al. 2015; Saw et al. 2015; Spang, Caceres and Ettema 2017; Williams et al. 2017). Secondly, some DPANN lineages may form a clade at the base of Archaea,



**Figure 4.** Illustration of controversially discussed placements of DPANN lineages in the archaeal part of the tree of life. While most studies indicate that DPANN represent a deep-branching archaeal superphylum (A), few studies have suggested that some (B) or all DPANN lineages (C) could represent fast-evolving archaeal taxa, which are artificially attracted to the base of the archaeal tree while in fact affiliating with other archaeal lineages.

while others would be secondarily derived lineages that group erroneously with the basal clade as a result of phylogenetic artifacts (Aouad *et al.* 2018). Finally, all of the DPANN lineages might be secondarily derived from within other archaeal groups, with their apparent monophyly being the result of phylogenetic artifacts such as long branch attraction (LBA). In LBA, long branches are erroneously grouped together in a tree even though they are not closely related (Bergsten 2005). The causes of LBA and the conditions under which it occurs are poorly understood, but the basic problem is that evolution along long branches is difficult to model. When multiple changes occur at a single amino acid site, it can be difficult to determine whether identical states represent evidence of recent common ancestry—that is, synapomorphies—or are the result of convergent evolution. In particular, parasites and symbionts are thought to experience different selective pressures compared to free-living organisms (Moran and Bennett 2014). For example, the switch to a host-associated lifestyle in bacteria often coincides with gene loss, genomic and cellular reduction, an elevated rate of sequence evolution and mutation-driven drift towards very low or very high GC contents (Clark, Moran and Baumann 1999; Moran 2003; Toft and Andersson 2010; Moran and Bennett 2014). In general, these characteristics can cause phylogenetic artifacts that can impede the exact phylogenetic placement of symbionts in the tree of life if not modeled carefully (Moran 1996). In the case of DPANN, the suspicion is that a large number of convergent amino acid changes associated with the transition from a free-living to host-associated lifestyle might cause these organisms to artifactually group together in a deep-branching clade. If so, it has been argued (Aouad *et al.* 2018) that increased sampling of DPANN might worsen, rather than ameliorate, the problems of LBA.

While GC content and amino acid frequencies of DPANN archaea do not appear to be as strongly biased as those of bacterial endosymbionts, they differ to some extent from the average of other archaea (Fig. 2B and C, Supplementary Table S2, Supporting Information). It remains to be determined whether these differences lead to compositional effects that could cause systematic phylogenetic artifacts. In addition, the long branches of DPANN representatives in published phylogenetic trees may indicate that genomes of members of this group evolve faster than non-symbiotic archaea. Genomic features of DPANN archaea that could potentially contribute to accelerated evolutionary rates include diversity-generating elements (DGRs), CRISPR systems and transposons. In particular, DGRs allow to create mas-

sive amounts of sequence variation in selected ligand-binding sites of target proteins (Doulatov *et al.* 2004; McMahon *et al.* 2005) and seem to be prevalent in DPANN lineages such as Nanoarchaeota (Paul *et al.* 2015), Pacearchaeota and Woesearchaeota (Handa *et al.* 2016; Paul *et al.* 2017) where they might play a role in cell-cell attachment and thus might be beneficial for dynamic host responses (Paul *et al.* 2017). Furthermore, at least some members of the DPANN encode CRISPR-Cas9 systems (Burstein *et al.* 2017), which coincides with the observation that some representatives of this group are subjected to viral infections (Comolli *et al.* 2009; Martínez-García *et al.* 2014). Frequent viral infections may influence genome dynamics due to an evolutionary arms race between viruses and hosts (Stern and Sorek 2011) and could thereby contribute to increased rates of evolution. Finally, the recent discovery of a duplicated transposon and localized single nucleotide polymorphisms in *Cand. Forterra* multitransposor (Diapherotrites) suggests that homologous recombination could also affect genome evolution in DPANN archaea (Probst and Banfield 2018). However, these characteristics are not universally conserved across the DPANN superphylum. Additionally, the observed loss of genes coding for proteins involved in DNA repair machineries in bacterial endosymbionts (Moran, McCutcheon and Nakabachi 2008) does not seem to represent a defining feature of DPANN archaea (Castelle *et al.* 2018), suggesting that means to control mutation rates might be retained in at least some members. Furthermore, several lineages that affiliate with DPANN in phylogenies, such as the Altiarchaeota, do not seem to represent obligate symbionts or are characterized by reductive genome evolution (Fig. 1, Supplementary Fig. S1, Supplementary File 1, Supporting Information). Therefore, it remains to be determined whether genome evolution of DPANN lineages is indeed characterized by fundamentally different rates as compared to other archaeal lineages and whether the potential host-associated lifestyle of members of this diverse clade contributes to the observed topology or whether DPANN monophyly represents a genuine signal.

### Current insights from phylogenetic analyses on the monophyly of DPANN archaea

To evaluate potential artifacts regarding the placement of DPANN in the tree of Archaea carefully, a number of authors have investigated the robustness of the DPANN clade using analyses that attempt to mitigate LBA. The CAT + GTR substitution model has been shown to be less susceptible to LBA artifacts



(Lartillot, Brinkmann and Philippe 2007) because the probabilities of change among amino acid states are calculated per site, rather than averaged over the entire sequence alignment as in standard phylogenetic models (Lartillot and Philippe 2004). The monophyly of DPANN received maximal support, both when CAT + GTR was used to analyze raw amino acid sequences and when the alignment was recoded into four Dayhoff categories of biochemically similar amino acids (Williams et al. 2017). Although the latter reduces information, this type of data recoding is a useful exploratory tool as it reduces substitutional saturation and compositional variation among the sequences and thereby allows a more accurate modeling of the data (Hrdy et al. 2004; Susko and Roger 2007). The placement of the different DPANN clades relative to each other was also assessed by comparing DPANN phylogenies, obtained by excluding or including representatives of other archaeal lineages, respectively (Williams et al. 2017). If DPANN monophyly is an LBA artifact, one may expect that the relationships within the group should be essentially random, and there is no reason to expect the same tree structure in the DPANN-only analysis. Yet, the results of both analyses were very similar and seem thus consistent with DPANN monophyly. However, it cannot be excluded that shared compositional biases between particular DPANN lineages could have caused the similar topologies of DPANN lineages relative to each other (Fig. 2B and C).

To date, the strongest evidence against DPANN monophyly comes from analyses that attempt to place DPANN taxa or sublineages into the tree of Archaea one at a time (Williams et al. 2017; Aouad et al. 2018). The assumption of this approach is that, if DPANN are indeed monophyletic, each individual member of the group should connect to a backbone phylogeny of other archaea in the same position when analyzed in isolation. If DPANN monophyly is instead an LBA artifact, individual taxa might be easier to place than the group as a whole, and different placements of the individual lineages might suggest that DPANN monophyly is artifactual. Williams et al. (2017) found that the Diapherotrites, Aenigmarchaeota and Woesearchaeota lineages branched basally when analyzed individually, but Nanoarchaeota, Nanohaloarchaeota and Pacearchaeota instead grouped within Euryarchaeota. Consistent with these results, Aouad et al. (2018) reported that Nanohaloarchaeota branched from the euryarchaeotal stem when the entire concatenated alignment was analyzed, but grouped with the Methanocellales (that is, within the Euryarchaeota) when including only slow-evolving sites. Finally, in this context it may also be worth considering the phylogenetic placement of Altiarchaeota. When analyzed individually, Altiarchaeota group within the Euryarchaeota (Probst et al. 2014; Adam et al. 2017), but this clade groups with DPANN when these are included in the analysis (Bird et al. 2016; Hug et al. 2016; Spang, Caceres and Ettema 2017). Taken together, these results demonstrate that the phylogenetic resolution of DPANN lineages is sensitive both to taxon sampling and to the methods of analysis used, and further work is needed to robustly place the DPANN archaea in the tree of life.

Previous difficult phylogenetic problems were ultimately solved through improved taxon sampling and the use of better-fitting phylogenetic models (Embley and Martin 2006), which account for both site-specific biochemical constraints and across-branch compositional heterogeneity. Thus, such approaches may also be of help to further assess the placement of DPANN lineages. While site- and branch-heterogeneous models have been described (Blanquart and Lartillot 2008; Heaps et al. 2014;

Williams et al. 2015; Cherlin et al. 2018) and implemented, they are currently computationally intractable even on modestly sized phylogenomic datasets. Therefore, speed-ups in the implementations of these methods, and the development of new, more efficient branch-and-site models, will be essential to help resolve DPANN phylogeny. Furthermore, the inclusion of a taxonomically broader dataset of potential DPANN lineages (Fig. 1) in careful phylogenetic analyses will help towards resolving the evolution of this enigmatic group of archaea. In particular, the assessment of the placement of Altiarchaeota, which do not show substantial indications for genome reduction, will be essential for anchoring the more derived DPANNs in the tree and for determining the degree to which convergent processes affect reductive genome evolution in Archaea.

## CONCLUSION AND OUTLOOK

The astounding diversity of available genomes of DPANN archaea (Fig. 1) and the currently limited set of characterized host-symbiont systems emphasize how much we still need to learn about the basic biology of these organisms. For example, important questions that need to be addressed in the coming years will be to determine which lineages are host-dependent and which are free-living; what are their host organisms and/or interaction partners and how much variation in terms of lifestyle, metabolism and gene content exists among these lineages. Certainly, the development and application of new methods will be essential to recover novel host-symbiont systems involving DPANN archaea (Jarett et al. 2018) and start answering some of these questions. Insights gained will shed light on what appears to be—at the very least—a major way of living among archaea and provide a better basis for assessing the functional importance and ecological role of members of the DPANN. Clearly, the vast diversity of DPANN archaea, most of which may depend on interacting partners and the presence of members of this group in environments across the entire biosphere (Castelle and Banfield 2018), indicate that they may represent a non-negligible component of microbial food webs (Probst et al. 2018).

Currently, we still face major difficulties to unequivocally determine the evolutionary origins of DPANN and thus the evolution of symbiosis in the archaeal domain of life. Taken at face value, recent phylogenies (Hug et al. 2016; Castelle and Banfield 2018; Castelle et al. 2018) suggest that the deepest split within both the archaeal and bacterial domains appears to be between a clade of relatively small, apparently genome-reduced lineages with various shared genetic features—i.e. the DPANN archaea on one side and the bacterial candidate phyla radiation (CPR) on the other side. This remains true, even if the taxonomic level of these lineages still represents a matter of debate (Parks et al. 2018). At least in the case of the bacterial tree, it is important to note that the many previously described highly reduced, fast-evolving symbiotic and parasitic bacterial lineages do not group with CPR, suggesting that current phylogenetic methods do not invariably succumb to LBA when parasites and symbionts are included in the analysis. Clearly, determining whether the monophyletic clustering and deep placement of genome reduced organisms in the tree of life represent a genuine phylogenetic signal is one of the most challenging and important current questions in microbial evolution. Tackling these questions will allow us to illuminate and potentially considerably change our current understanding of the nature of the last common ancestor of Archaea and Bacteria as well as of the role of symbiosis in the evolution of life on Earth.

## SUPPLEMENTARY DATA

Supplementary data are available at [FEMSLE](#) online.

## ACKNOWLEDGEMENTS

We would also like to thank Thijs Ettema for kindly hosting Jun-Hoe in the Ettema lab at Uppsala University and for providing infrastructure. Furthermore, we would like to thank Brett Baker for supporting Nina Dombrowski. Finally, we are thankful to Harald Huber for sharing EM pictures on the Nanoarchaeum-Ignicoccus host-symbiont system.

## FUNDING

This work was supported by a grant of the Swedish Research Council (VR starting grant 2016-03559 to AS), the NWO-I foundation of the Netherlands Organisation for Scientific Research (WISE fellowship to AS), and a Royal Society University Research Fellowship to TAW.

**Conflict of interest.** None declared.

## REFERENCES

- Adam PS, Borrel G, Brochier-Armanet C et al. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *Nature* 2017;**11**:353–8.
- Andersson JO, Sarchfield SW, Roger AJ. Gene transfers from Nanoarchaeota to an ancestor of diplomonads and parabasalids. *Mol Biol Evol* 2005;**22**:85–90.
- Aono R, Sato T, Imanaka T et al. A pentose bisphosphate pathway for nucleoside degradation in Archaea. *Nat Chem Biol* 2015;**11**:355–60.
- Aouad M, Taib N, Oudart A et al. Extreme halophilic archaea derive from two distinct methanogen class II lineages. *Mol Phylogenet Evol* 2018;**127**:46–54.
- Baker BJ, Comolli LR, Dick GJ et al. Enigmatic, ultrasmall, uncultivated Archaea. *Proc Natl Acad Sci USA* 2010;**107**:8806–11.
- Baker BJ, Dick GJ. Omic approaches in microbial ecology; charting the unknown. *Microbe* 2013;**8**:353–9.
- Bergsten J. A review of long-branch attraction. *Cladistics* 2005;**21**:163–93.
- Bird JT, Baker BJ, Probst AJ et al. Culture independent genomic comparisons reveal environmental adaptations for Altiarchaeales. *Front Microbiol* 2016;**7**:1221.
- Blanquart S, Lartillot N. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 2008;**25**:842–58.
- Brochier C, Gribaldo S, Zivanovic Y et al. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* 2005;**6**:R42.
- Brochier-Armanet C, Boussau B, Gribaldo S et al. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 2008;**6**:245–52.
- Burghardt T, Junglas B, Siedler F et al. The interaction of *Nanoarchaeum equitans* with *Ignicoccus hospitalis*: proteins in the contact site between two cells. *Biochem Soc Trans* 2009;**37**:127–32.
- Burstein D, Harrington LB, Strutt SC et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* 2017;**542**:237–41.
- Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 2018;**172**:1181–97.
- Castelle CJ, Brown CT, Anantharaman K et al. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat Rev Microbiol* 2018;**16**:629–45.
- Castelle CJ, Wrighton KC, Thomas BC et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* 2015;**25**:690–701.
- Chen L-X, Méndez-García C, Dombrowski N et al. Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J* 2018;**12**:756–75.
- Cherlin S, Heaps SE, Nye TMW et al. The effect of nonreversibility on inferring rooted phylogenies. *Mol Biol Evol* 2018;**35**:984–1002.
- Clark MA, Moran NA, Baumann P. Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* 1999;**16**:1586–98.
- Comolli LR, Banfield JF. Inter-species interconnections in acid mine drainage microbial communities. *Front Microbiol* 2014;**5**:367.
- Comolli LR, Baker BJ, Downing KH et al. Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *ISME J* 2009;**3**:159–67.
- DeLong EF. Archaea in coastal marine environments. *Proc Natl Acad Sci USA* 1992;**89**:5685–9.
- DeLong EF, Pace NR, Kane M. Environmental diversity of bacteria and archaea. *Syst Biol* 2001;**50**:470–8.
- Di Giulio M. The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). *J Theor Biol* 2006;**240**:343–52.
- Doulatov S, Hodes A, Dai L et al. Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature* 2004;**431**:476–81.
- Durbin AM, Teske A. Sediment-associated microdiversity within the Marine Group I Crenarchaeota. *Environ Microbiol Rep* 2010;**2**:693–703.
- Durbin AM, Teske A. Archaea in organic-lean and organic-rich marine subsurface sediments: an environmental gradient reflected in distinct phylogenetic lineages. *Front Microbiol* 2012;**3**:168.
- Embley TM, Martin W. Eukaryotic evolution, changes and challenges. *Nature* 2006;**440**:623–30.
- Eme L, Spang A, Lombard J et al. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 2017;**15**:711–23.
- Fuhrman JA, McCallum K, Davis AA. Novel major archaeobacterial group from marine plankton. *Nature* 1992;**356**:148–9.
- Furukawa R, Nakagawa M, Kuroyanagi T et al. Quest for ancestors of eukaryal cells based on phylogenetic analyses of aminoacyl-tRNA synthetases. *J Mol Evol* 2017;**84**:51–66.
- Ghai R, Pašić L, Fernández AB et al. New abundant microbial groups in aquatic hypersaline environments. *Sci Rep* 2011;**1**:135.
- Giannone RJ, Huber H, Karpinets T et al. Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans*-*Ignicoccus hospitalis* relationship. *PLoS One* 2011;**6**:e22942.
- Giannone RJ, Wurch LL, Heimerl T et al. Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. *ISME J* 2015;**9**:101–14.
- Golyshina OV, Toshchakov SV, Makarova KS et al. 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. *Nat Commun* 2017;**8**:60.
- Graybeal A, Cannatella D. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst Biol* 1998;**47**:9–17.

- Guindon S, Dufayard J-F, Lefort V et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–21.
- Hamann E, Tegetmeyer HE, Riedel D et al. Syntrophic linkage between predatory *Carpodimonas* and specific prokaryotic populations. *ISME J* 2017;11:1205–17.
- Hamerly T, Tripet BP, Tigges M et al. Untargeted metabolomics studies employing NMR and LC-MS reveal metabolic coupling between *Nanoarchaeum equitans* and its archaeal host *Ignicoccus hospitalis*. *Metabolomics* 2015;11:895–907.
- Handa S, Paul BG, Miller JF et al. Conservation of the C-type lectin fold for accommodating massive sequence variation in archaeal diversity-generating retroelements. *BMC Struct Biol* 2016;16:13.
- Heaps SE, Nye TMW, Boys RJ et al. Bayesian modelling of compositional heterogeneity in molecular phylogenetics. *Stat Appl Genet Mol* 2014;13:589–609.
- Heimerl T, Flechsler J, Pickl C et al. A complex endomembrane system in the archaea *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. *Front Microbiol* 2017;8:1072.
- Hrdy I, Hirt RP, Dolezal P et al. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* 2004;432:618–22.
- Huber H, Hohn MJ, Rachel R et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 2002;417:63–67.
- Hug LA, Baker BJ, Anantharaman K et al. A new view of the tree of life. *Nat Microbiol* 2016;1:16048.
- Jahn U, Gallenberger M, Paper W et al. *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J Bacteriol* 2008;190:1743–50.
- Jahn U, Summons R, Sturt H et al. Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch Microbiol* 2004;182:404–13.
- Jarett JK, Nayfach S, Podar M et al. Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* 2018;6:161.
- Junglas B, Briegel A, Burghardt T et al. *Ignicoccus hospitalis* and *Nanoarchaeum equitans*: ultrastructure, cell–cell interaction, and 3D reconstruction from serial sections of freeze-substituted cells and by electron cryotomography. *Arch Microbiol* 2008;190:395–408.
- Karr JR, Sanghvi JC, Macklin DN et al. A whole-cell computational model predicts phenotype from genotype. *Cell* 2012;150:389–401.
- Koskinen K, Pausan MR, Perras AK et al. First Insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin. *mBio* 2017;8:e00824–00817.
- Krause S, Bremges A, Münch PC et al. Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci Rep* 2017;7:3289.
- Küper U, Meyer C, Müller V et al. Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic Archaeon *Ignicoccus hospitalis*. *Proc Natl Acad Sci USA* 2010;107:3152–6.
- Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 2007;7:S4.
- Lartillot N, Philippe H. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 2004;21:1095–109.
- Leeuwenhoek AV. Observations, communicated to the publisher by Mr. Antony van Leewenhoek, in a dutch letter of the 9th Octob. 1676. here English'd: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infused. *Phil Trans* 1677;12: 821–31.
- Lewalter K, Müller V. Bioenergetics of archaea: Ancient energy conserving mechanisms developed in the early history of life. *BBA-Bioenergetics* 2006;1757:437–45.
- Liu X, Li M, Castelle CJ et al. Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. *Microbiome* 2018;6:102.
- Martínez-García M, Santos F, Moreno-Paz M et al. Unveiling viral–host interactions within the ‘microbial dark matter’. *Nat Commun* 2014;5:4542.
- Mayer F, Küper U, Meyer C et al. AMP-forming acetyl Coenzyme A synthetase in the outermost membrane of the hyperthermophilic Crenarchaeon *Ignicoccus hospitalis*. *J Bacteriol* 2012;194:1572–81.
- McMahon SA, Miller JL, Lawton JA et al. The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat Struct Mol Biol* 2005;12:886–92.
- Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013;30:1188–95.
- Mohanty S, Jobichen C, Chichili VPR et al. Structural basis for a unique ATP synthase core complex from *Nanoarchaeum equitans*. *J Biol Chem* 2015;290:27280–96.
- Moran NA. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 1996;93:2873–8.
- Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol* 2003;6:512–8.
- Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol* 2014;68:195–215.
- Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 2008;42:165–90.
- Munson-McGee JH, Field EK, Bateson M et al. Nanoarchaeota, their Sulfolobales host, and Nanoarchaeota virus distribution across Yellowstone National Park hot springs. *Appl Environ Microb* 2015;81:7860–8.
- Narasimarao P, Podell S, Ugalde JA et al. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* 2012;6:81–93.
- Narrowe AB, Spang A, Stairs CW et al. Complex evolutionary history of translation Elongation Factor 2 and diphthamide biosynthesis in Archaea and parabasalids. *Genome Biol Evol* 2018;10:2380–93.
- Ortiz-Alvarez R, Casamayor EO. High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. *Environ Microbiol Rep* 2016;8:210–7.
- Parks DH, Chuvochina M, Waite DW et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
- Paul BG, Bagby SC, Czornyj E et al. Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nat Commun* 2015;6:6585.
- Paul BG, Burstein D, Castelle CJ et al. Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat Microbiol* 2017;2:17045.
- Petitjean C, Deschamps P, López-García P et al. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new Kingdom Proteoarchaeota. *Genome Biol Evol* 2015;7:191–204.

- Podar M, Anderson I, Makarova KS et al. A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*. *Genome Biol* 2008;**9**:R158.
- Podar M, Makarova KS, Graham DE et al. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol Direct* 2013;**8**:9.
- Probst AJ, Banfield JF. Homologous recombination and transposon propagation shape the population structure of an organism from the deep subsurface with minimal metabolism. *Genome Biol Evol* 2018;**10**:1115–9.
- Probst AJ, Holman H-YN, DeSantis TZ et al. Tackling the minority: sulfate-reducing bacteria in an archaea-dominated subsurface biofilm. *ISME J* 2013;**7**:635–51.
- Probst AJ, Ladd B, Jarett JK et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* 2018;**3**:328–36.
- Probst AJ, Moissl-Eichinger C. “Altiarchaeales”: Uncultivated Archaea from the subsurface. *Life* 2015;**5**:1381–95.
- Probst AJ, Weinmaier T, Raymann K et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat Commun* 2014;**5**:5497.
- Randau L, Münch R, Hohn MJ et al. *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5′- and 3′-halves. *Nature* 2005;**433**:537–41.
- Randau L, Pearson M, Söll D. The complete set of tRNA species in *Nanoarchaeum equitans*. *FEBS Lett* 2005;**579**:2945–7.
- Raymann K, Forterre P, Brochier-Armanet C et al. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol* 2014;**6**:192–212.
- Rinke C, Schwientek P, Sczyrba A et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;**499**:431–7.
- Rudolph C, Wanner G, Huber R. Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a String-of-Pearls-like morphology. *Appl Environ Microb* 2001;**67**:2336–44.
- Sato T, Atomi H, Imanaka T. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 2007;**315**:1003–6.
- Saw JH, Spang A, Zaremba-Niedzwiedzka K et al. Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Phil Trans R Soc B* 2015;**370**:20140328.
- Shcherbakova V, Yoshimura Y, Ryzhmanova Y et al. Archaeal communities of Arctic methane-containing permafrost. *FEMS Microbiol Ecol* 2016;**92**, DOI: 10.1093/femsec/fiw135.
- Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 2017;**357**:eaaf3883.
- Spang A, Martijn J, Saw JH et al. Close encounters of the third domain: the emerging genomic view of archaeal diversity and evolution. *Archaea* 2013;**2013**:202358.
- Spang A, Saw JH, Jørgensen SL et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015;**521**:173–9.
- St. John E, Liu Y, Podar M et al. A new symbiotic Nanoarchaeote (*Candidatus Nanoclepta minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst Appl Microbiol* 2018.
- Stern A, Sorek R. The phage-host arms-race: Shaping the evolution of microbes. *Bioessays* 2011;**33**:43–51.
- Susko E, Roger AJ. On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol* 2007;**24**:2139–50.
- Takai K, Horikoshi K. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 1999;**152**:1285–97.
- Takai K, Moser DP, DeFlaun M et al. Archaeal diversity in waters from deep South African gold mines. *Appl Environ Microb* 2001;**67**:5750–60.
- Toft C, Andersson SGE. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 2010;**11**:465–75.
- Vavourakis CD, Ghai R, Rodriguez-Valera F et al. Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front Microbiol* 2016;**7**:211.
- Waters E, Hohn MJ, Ahel I et al. The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 2003;**100**:12984–8.
- Williams TA, Heaps SE, Cherlin S et al. New substitution models for rooting phylogenetic trees. *Phil Trans R Soc B* 2015;**370**:20140336.
- Williams TA, Szöllösi GJ, Spang A et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci USA* 2017;**114**:E4602–11.
- Wodke JAH, Puchałka J, Lluch-Senar M et al. Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol Syst Biol* 2013;**9**:653.
- Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 1977;**74**:5088–90.
- Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990;**87**:4576–9.
- Wurch L, Giannone RJ, Belisle BS et al. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat Commun* 2016;**7**:12115.
- Youssef NH, Rinke C, Stepanauskas R et al. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum ‘Diapherotrites’. *ISME J* 2015;**9**:447–60.
- Yus E, Maier T, Michalodimitrakis K et al. Impact of genome reduction on bacterial metabolism and its regulation. *Science* 2009;**326**:1263–8.
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017;**541**:353–8.