Title

Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life

Journal

Cell, 172(6)

ISSN

0092-8674

Authors

Castelle, Cindy J

Banfield, Jillian F

Publication Date

2018-03-01

DOI

10.1016/j.cell.2018.02.016

Peer reviewed

Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life

Cindy J. Castelle[1,2,3] and Jillian F. Banfield[1,2,3,4,5,6,*]

[1]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA [2]Innovative Genomics Institute, Berkeley, CA, USA [3]Chan Zuckerberg Biohub, San Francisco, CA, USA [4]University of Melbourne, Melbourne, VIC, Australia [5]Lawrence Berkeley National Laboratory, Berkeley, CA, USA [6]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA, USA

*Correspondence: jbanfield@berkeley.edu

The recent recovery of genomes for organisms from phyla with no isolated representative (candidate phyla) via cultivation-independent genomics enabled delineation of major new microbial lineages, namely the bacterial candidate phyla radiation (CPR), DPANN archaea, and Asgard archaea. CPR and DPANN organisms are inferred to be mostly symbionts, and some are episymbionts of other microbial community members. Asgard genomes encode typically eukaryotic systems, and their inclusion in phylogenetic analyses results in placement of eukaryotes as a branch within Archaea. Here, we illustrate how new genomes have changed the structure of the tree of life and altered our understanding of biology, evolution, and metabolic roles in biogeochemical processes.

Introduction

The tree of life is arguably the most important organizing principle in biology and perhaps the most widely understood depiction of the evolutionary process. It explains to us how we are related to other organisms and where we may have come from. The tree has undergone some tremendous revolutions since the first version was sketched by Charles Darwin. A major innovation was the construction of phylogenetic trees using DNA sequence information, which opened the way for classification of microbial life. As implemented by Carl Woese and collaborators, this work enabled the definition of three domains: Bacteria, Archaea, and Eukaryotes (Woese and Fox, 1977, Woese et al., 1990). More recently, the three-domain topology has been questioned, and eukaryotes—our own branch of life—potentially relocated into the archael domain (Spang et al., 2015, Williams et al., 2013). Beyond this, and as described here, cultivation-independent genomic methods that access sequences from laboratory-intractable organisms have added many new lineages to the tree. Their inclusion completely clarifies the extreme minority of life’s diversity that is represented by macroscopic organisms and underscores that our place in biology is dwarfed by bacteria and archaea.

Genome Resolution of Uncultivated Organisms

Cultivation-independent (e.g., 16S rRNA gene-based) survey methods uncovered evidence for vast microbial diversity that was not represented in the set of organisms available in pure culture (e.g., Dojka et al., 2000, Pace, 1997). This motivated the development of new genomic approaches that can provide comprehensive metabolic insights without the requirement for laboratory growth. The first of these is referred to as community genomics or, more commonly, metagenomics. As metagenomics is sometimes considered to include rRNA gene surveys, we adopt the term “genome-resolved metagenomics” to describe the specific approach that can yield direct information about metabolic capacity on an organism-by-organism basis. In genome-resolved metagenomics, DNA is extracted from a whole community or an enrichment and shotgun sequenced, and the short sequences are assembled into larger genome fragments that are ultimately assigned to genome bins (draft genomes; Tyson et al., 2004). Typically, metabolic predictions are only undertaken for reasonably high-quality genomes (>70% complete with low contamination by fragments from other organisms). In a few cases, further assembly curation has generated complete (closed) genomes (e.g., Albertsen et al., 2013, Kantor et al., 2013). Importantly, the approach is not limited to Bacteria and Archaea but can provide draft genomes for Eukaryotes and partial or complete genomes of phage, viruses, and plasmids (e.g., Paez-Espino et al., 2016). The disadvantage of this method is that the sampled cells comprise a natural population. Thus, the genome is, to differing extents, a composite of sequence variants distinguished by single-nucleotide polymorphisms and insertions/deletions. On the other hand, the reads can be mapped back to the genome sequence to provide a snapshot of the form and extent of variation within the population (Simmons et al., 2008).

The second method, single-cell genomics, involves the sequencing of fragments of DNA amplified from a single cell or collection of cells with similar rRNA gene sequences (e.g., Podar et al., 2007). The method has been especially useful for study of eukaryotic cells (Gawad et al., 2016). The advantage of a method in which a single microbial cell is sequenced is that all reported nucleotides derive from one genome so that linkage patterns can be established; however, binning may be required to remove contaminating sequences. A comparative study that targeted both methods to one microbial community revealed that the single-cell method is of much lower throughput than genome-resolved metagenomics, and the genomes are, on average, significantly less complete. However, the sequences from single-cell genomics and metagenomics analyses generally agree well (Probst et al., 2018). Single-cell sequencing can be implemented in a targeted way following screening of rRNA genes (Rinke et al., 2013). Thus, the resulting sequences directly augment the collection of genomes for organisms belonging to lineages that are unsampled or undersampled.

An Updated Topology of the Tree of Life

Hug et al. (2016) presented a new version of the tree of life that captured information for organisms whose draft genomes had been reconstructed without prior cultivation. Although the deep branching order is not well supported, it is possible that the root of the bacterial domain is placed between well-known bacteria (including the commonly known major lineages, such as Proteobacteria, Actinobacteria, Firmicutes, and Cyanobacteria, with well-studied representatives) and the bacteria of a recently described and seemingly monophyletic group referred to as the candidate phyla radiation (CPR) (Brown et al., 2015). Recent detailed phylogenetic analyses have considered the placement of the DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota), a newly recognized superphylum that is described in more detail below, within the archaeal domain (Williams et al., 2017). To root the archaeal tree, Williams et al. applied a new approach that is based on single-gene tree reconciliation with no need for the use of an outgroup. They proposed that the root of the archaeal domain is placed between the all other archaea and the DPANN. Their analyses suggest the monophyly of the DPANN superphylum and the Euryarchaeota phylum. However, due to low sampling of the DPANN superphylum at the time of their analysis, new sequences and approaches are needed to further evaluate this conclusion. Sequencing and bioinformatics methods are evolving rapidly, and new genomes have recently become available. Figure 1 presents an updated view of the three-domain tree that includes new sequences from DPANN archaea.

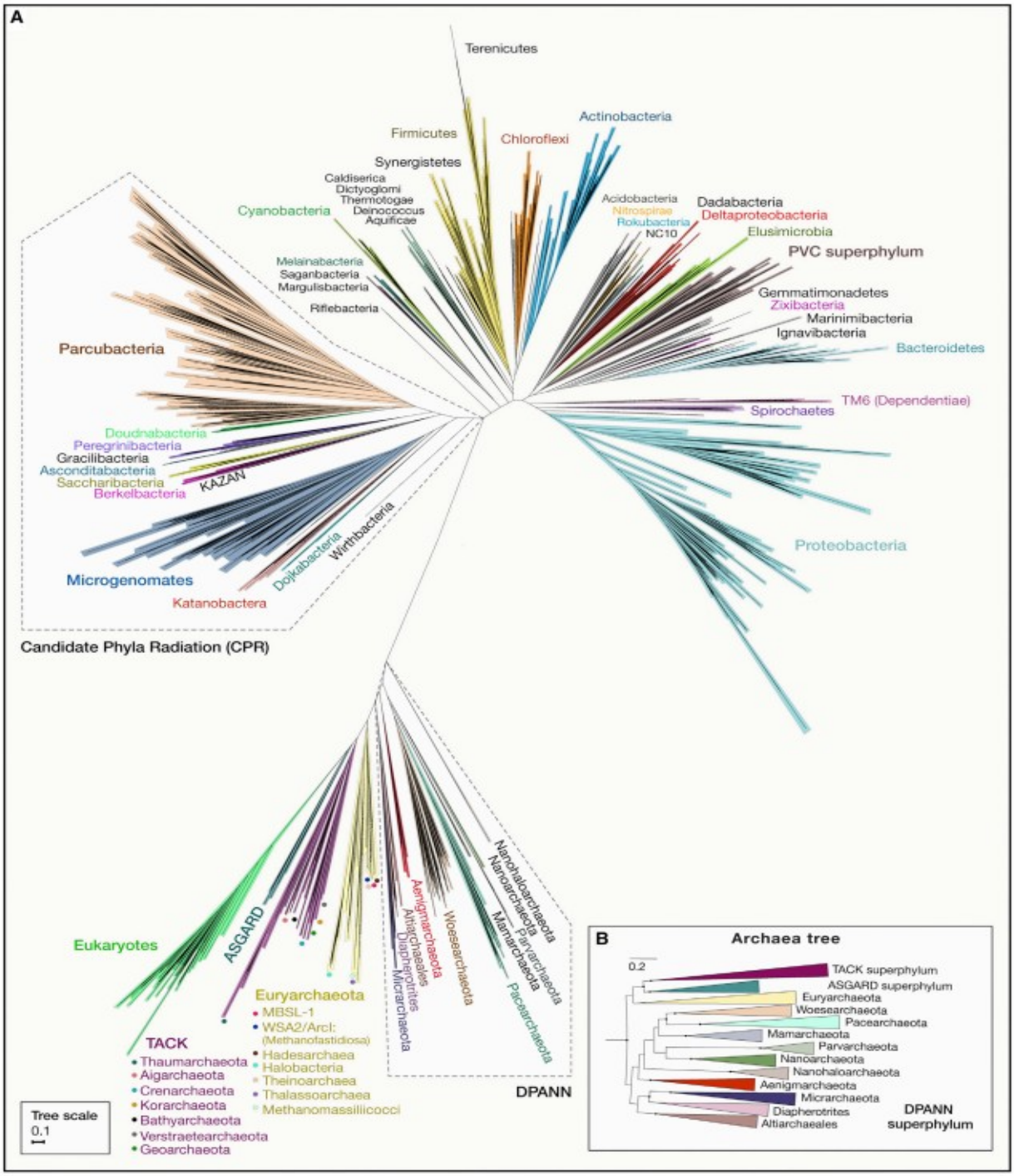

Figure 1. Phylogenetic Tree

(A) Updated from Hug et al. (2016) by addition of new DPANN archaeal sequences (many but not all phyla are indicated on the tree). Note that with the use of a significant number of bacterial sequences in a three-domain tree topology (Bacteria, Archaea, and Eukaryotes), the DPANN superphylum do not form a monophyletic clade. As previously discussed, the placement of nanosized lineages and the monophyly of the DPANN within the tree of life remain unclear and important open issues that need to be addressed with the development of new methodological approaches.

(B) We reconstructed a phylogenetic tree with only the archaeal sequences (with bacteria as an outgroup) in which the monophyly of the DPANN clade is retrieved but the branching order of the deepest branches cannot be confidently resolved. Bootstrap support values in are indicated by circles on nodes (black for support of 89% and above and gray for support from 50% to 89%).

The trees were calculated using a maximum-likelihood algorithm (RAxML with PROTCATLG model) based on genome sequences containing 14 ribosomal proteins (ribosomal proteins L2, L3, L4, L5, L6, L14, L15, L18, L22, L24, S3, S8, S17, and S19). The concatenated ribosomal protein alignment was constructed as described previously (Hug et al., 2016). The updated tree of life is available with full bootstrap values in Newick format in Data S1.

The extent of diversity within the CPR is a topic of current debate. Analyses based on 16S rRNA or concatenation of ribosomal proteins suggest that the CPR may be comparable in breadth to all other bacteria (Hug et al., 2016). However, it remains uncertain whether this is due to early divergence or rapid evolution (or both). Other estimates of the scale of the CPR predict that they could comprise a maximum of 26% (Parks et al., 2017) or 25% (Schulz et al., 2017) of bacterial diversity. A reanalysis of this question using the full set of available genomes, genes universal to both groups, and improved phylogenetic methods is required.

After the explosive growth in the number of major branches either discovered or genomically resolved via cultivation-independent methods between 2012 and 2016, one may wonder if the appearance of new phylum-level groups will continue indefinitely. An analysis of the rate of discovery of new groups would suggest that this is not the case. Around 8,000 genomes were reconstructed *de novo* by exploiting the public short-read archive (*circa* mid-2016). The archive includes samples from a huge array of environment types (Parks et al., 2017). The majority of genomes were for bacteria and archaea from clades already known to exist. A subset was assigned to 44 putative phyla that were defined by cultivation-independent genomic methods over the past 5 years. 3 potentially additional archaeal phyla and 17 possible bacterial phyla were identified (although some of these may have been independently described in the period between download of the archive and publication). In fact, 2017 saw reports of the first genomic sampling of several new groups, many of which are in the previously undersampled archaeal domain (Table S1). Once found, the major phylum-level lineages of bacteria and archaea are often identified across multiple ecosystems (see below). We suspect that the phylum content of the bacterial and archaeal domains will soon approach saturation, although the topology of the tree may change as algorithms to analyze it improve. The scope for discovery of new classes and groups at finer scales of taxonomic resolution is immense.

If we are approaching full delineation of the major branches of the tree of life, one may wonder why it took so long and then happened so quickly. It is our perspective that this is a reflection of relatively little prior focus on subsurface environments (which constitute a huge part of the biosphere but are difficult to sample) and challenges associated with the application of cultivation-independent genomic methods to the most complicated ecosystems (e.g., soil). Many subsurface environments are anaerobic, and therefore may be refuges for organisms from lineages that diverged early from primitive life forms. These environments are also enriched for organisms not commonly associated with humans, animals, or crops (thus viewed as lower-priority research targets). Some subsurface environments are low in nutrients and thus harbor slow-growing organisms with complex interdependencies that ensure retention of resources within the ecosystem (Probst et al., 2018). Overall, these more recently studied environments are

enriched in organisms that have been difficult (or impossible) to cultivate, so their discovery and genomic characterization awaited the development of community-wide cultivation-independent methods.

New Phyla and Phylum Radiations

16S rRNA gene surveys perfomed by Norman Pace's group uncovered sequences from organisms from a "candidate division" (analogous to a candidate phylum) labeled OP11 that were detected in hot springs of Yellowstone National Park (Hugenholtz et al., 1998). As the diversity of sequences expanded, it was apparent that the OP11 radiation includes multiple potentially phylum-level groups, including the OD1 superphylum (Harris et al., 2004) now referred to as Parcubacteria (Rinke et al., 2013) and the OP11 superphylum now referred to as Microgenomates (Rinke et al., 2013). Metagenomics-derived rRNA and protein sequences grouped these into the CPR, and many new major lineages were added (Brown et al., 2015). As discussed below, bacteria of the CPR consistently have small genomes and cell sizes, and most are predicted to have symbiotic lifestyles. Importantly, some groups within the CPR are not detected in rRNA gene surveys due to primer mismatch, introns, or both (Brown et al., 2015). Currently, we estimate that 73 CPR groups have been identified (Figure 1 and Table S1).

Rinke et al. (2013) proposed the term Patescibacteria for a superphylum comprising Microgenomates (OP11), Parcubacteria (OD1), and Gracilibacteria (GNO2-BD1-5), but the meaning of the term has become confused and even misused as synonymous with the CPR (Parks et al., 2017). The term CPR is simply a description of a huge monophyletic radiation of phyla and superphlya that includes the group that ultimately may be referred to as Patscibacteria and dozens of additional phyla and superphyla. Community-wide consultation should be undertaken before a formal name for this radiation is proposed.

It should be noted that a widely accepted definition of what constitutes a superphylum or even a new phylum is not available. In our experience, new phylum-level branches are distinct, apparently monophyletic, and have 16S rRNA genes that share <75%–80% identity with the most closely related groups. Ultimately, appropriate phylum and superphylum definitions require strong evidence for deep branch placements, and this is not yet available.

Based on the current summary, 62 non-CPR bacterial phylum-level groups have representatives that have been genomically described via cultivation-independent methods (Table S1). Given that around 30 traditionally have been described bacterial phyla (a handful of which were candidate phyla that now have recently isolated representatives), the scope of the bacterial domain exclusive of CPR may have approximately tripled. When CPR and non-CPR groups are taken together, we estimate that cultivation-independent genomic methods have expanded the scale of domain Bacteria by over five times.

There were early indications for the existence of a group of novel archaea potentially analogous to the CPR organisms. These include the co-cultivation of *Nanoarchaeum equitans* with its host (Huber et al., 2002), genomic description and imaging of what were initially referred to as archaeal Richmond Mine acidophilic nanoorganisms (ARMAN) archaea (Baker et al., 2006, Baker et al., 2010), and the genomic characterization of extremely salt-adapted nanohaloarchaea (Narasingarao et al., 2012). Based on a collection of draft genomes from single cells (or concentrates), a segment of the archaeal domain referred to as DPANN was proposed (Rinke et al., 2013). Since then, new genomes and the definition of at least nine major groups have greatly expanded the DPANN lineages (Castelle et al., 2015).

Archaea now include at least three other major supergroups: the Euryarchaeota, the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota; proposed name Eocyta; Lake, 2015), and Asgard archaea, all of which comprise potentially phylum- or superphylum-level clades (see below; Figure 1). In addition, 6 potential phylum-level groups have been proposed within the TACK superphylum, and 15 additional groups affiliate approximately with the Euryarchaeota. Although the phylum-level delineation of Archaea is especially complicated (Adam et al., 2017, Spang et al., 2017), it appears that new archaeal candidate phyla (35 groups listed in Table S1) expand the scope of archaeal diversity by a factor similar to that noted for bacteria.

New Discoveries Regarding Potential Roles of Non-DPANN Archaea in Biogeochemical Cycles

Genomic data for uncultivated archaea resulted in key discoveries related to their roles in the carbon, nitrogen, and sulfur geochemical cycles. Excellent recent reviews have been published on the phylogeny and ecological roles of archaea in diverse ecosystems (Adam et al., 2017, Spang et al., 2017). Thus, we focus here on a few examples related to ammonia and methane metabolism to illustrate the overarching principle that genome-resolved metagenomics can inform our understanding of evolution and biogeochemistry.

An important early result of metagenomics was the discovery of homologs of ammonia monooxygenase genes in archaeal genome fragments reconstructed from a Sargasso Sea metagenome (Venter et al., 2004). The existence of an unknown organism responsible for much of the ammonia oxidation in the ocean had been suspected based on *in situ* measurements that indicated that ammonia oxidation often proceeds at substrate concentrations significantly below the growth threshold of cultured ammonia-oxidizing bacteria. Subsequently, autotrophic ammonia-oxidizing archaea were isolated and their critical role in the global nitrogen cycle demonstrated (Pester et al., 2011, Stahl and de la Torre, 2012). In 2008, they were proposed as the third archaeal phylum and named Thaumarchaeota (Brochier-Armanet et al., 2008).

Sibling to Thaumarchaeota and Aigarchaeota are the Bathyarchaeota, which appear to be key players in the global carbon cycle in terrestrial and marine anoxic sediments. Some Bathyarchaeota possess the archaeal Wood-Ljungdahl pathway (He et al., 2016, Lazar et al., 2016), suggesting a capacity for CO_2 fixation via acetogenesis, a process thought to be unique to bacteria (He et al., 2016). Importantly, some Bathyarchaeota may be methanogens, given key genes that could produce methane from methanol, methyl sulfides, and methylated amines (Evans et al., 2015). Verstraetearchaeota (TACK superphylum) are also found in anoxic environments with high methane fluxes and, like some Bathyarchaeota, are predicted to conserve energy via methylotrophic methanogenesis (Vanwonterghem et al., 2016).

Metagenomics has been critical for the study of methane oxidation. Anaerobic methane-oxidizing euryarchaea-2d (ANME-2d) are often detected in the sulfate-methane transition zone in marine sediments and use a modified and reverse-methanogenesis pathway for growth (Hallam et al., 2004). Certain ANME-2d from granitic groundwater were incubated with ^{13}C-labeled methane to demonstrate that methane oxidation is linked to microbial sulfate reduction (Ino et al., 2018). Others are predicted to conduct reverse methanogenesis using nitrate as the terminal electron acceptor, and this has been experimentally demonstrated using bioreactors and ^{13}C and ^{15}N labeling (Haroon et al., 2013). It has also been predicted that some ANME-2d oxidize methane using manganese (Beal et al., 2009) or iron (Hernsdorf et al., 2017) as the electron acceptor. The genomes of anaerobic Methanomassiliicoccales also encode key enzymes involved in methyl-dependent methanogenesis, unlike other members of Thermoplasmata-related lineages (Borrel et al., 2014). WSA2/Arc1, sibling to both the non-methanogenic Hadesarchaeota (Baker et al., 2016) and MSBL1 (Mwirichia et al., 2016), may be methanogens and were recently named “Candidatus Methanofastidiosa.” Based on metabolic predictions from genome sequences from a wastewater treatment bioreactor (Nobu et al., 2016), they lack pathways for CO_2-reducing and aceticlastic methanogenesis, but they may be capable of methane formation via methylated thiol reduction with H_2.

Overall, recent metagenomic studies have substantially expanded the diversity of archaea predicted to be involved in methane metabolism. The inference that Bathyarchaeota and Verstraetearchaeota, members of the TACK superphylum, are methanogens is important because previously, methanogenesis was unknown outside of the Euryarchaeota. New groups such as the Methanomassiliicoccales and Methanofastidiosa also expand the diversity of Euryarchaeotes predicted to be involved in methanogenesis. However, it should be noted that the genes for methanogenesis and methane oxidation via reverse methanogenesis are essentially the same. This observation underlines the importance of experimental (e.g., bioreactor) studies to appropriately define metabolism *in situ*.

Discoveries Regarding DPANN Archaea

Understanding of the diversity, distribution, and metabolisms of nano-sized archaea that are members of the DPANN superphylum has been substantially advanced via cultivation-independent genomics studies. DPANN archaea are found in various extreme (e.g., hot, acidic, hypersaline) and temperate (e.g., lake and marine sediments) ecosystems, although they have been relatively rarely detected in soil and the open ocean. The first ARMAN archaeal groups were discovered in acid mine drainage biofilms via genomics-based approaches (Baker et al., 2006, Baker et al., 2010). The coverage of these groups and the diversity of the environments in which they occur has expanded (Chen et al., 2017), and the phyla have been renamed as Parvarchaeota and Micrarchaeota (Baker and Dick, 2013, Castelle et al., 2015). Both lineages are found in association with other archaea (see below). Single-cell genomics and metagenomics uncovered and defined additional phylum-level lineages within the DPANN superphylum. These include Diapherotrites and Aenigmarchaeota (Rinke et al., 2013), Nanoarchaeota (which includes the previously described *N. equitans*; Huber et al., 2002), Nanohaloarchaeota (Narasingarao et al., 2012), Woesearchaeota, and Pacearchaeota (Castelle et al., 2015). Common features unifying DPANN archaea are their small genomes, small cell sizes and limited metabolic repertoires: many lack core biosynthetic pathways for nucleotides, amino acids, and lipids. Most DPANN archaea depend on other microbes to meet their biological requirements. However, some appear to have the genetic potential to be free living with a heterotrophic aerobic and/or fermentative lifestyle.

A potential phylum-level clade represented by the Altiarchaeales (formerly SM1; Probst et al., 2014) appears to branch deeply within the DPANN superphylum based on recent studies (Bird et al., 2016, Spang et al., 2017) and our analyses (Figure 1). However, this placement requires further analysis. Altiarchaeales dominate some cold subsurface anaerobic groundwater habitats (Probst et al., 2014). Unlike most archaea, they have an outer membrane and unique surface-attached grappling hooks known as hami. They form biofilms and appear to grow autotrophically on carbon monoxide, acetate, or formate (via a modified archaeal Wood-Ljungdahl pathway; Probst et al., 2014).

Some New Discoveries Regarding Potential Roles of Non-CPR Bacteria in Biogeochemical Cycles

Many bacterial groups have new genomically described representatives (Table S1), but it is beyond the scope of this Perspective to provide details for all of these. Many of the bacterial genomes used to define these putative candidate phyla were reported by nantharaman et al. (2016, who provided an extensive table in the Supplemental Information that predicts the biogeochemically relevant capacities for each genome. Here, we review a few select examples from the literature to illustrate the types of insights into ecological roles and evolutionary histories that have been obtained.

A previously unknown bacterium was found to dominate aquifer sediments. A complete curated genome for one population was reconstructed from a metagenome, and the lineage was named Zixibacteria (Castelle et al., 2013). Gene-by-gene analysis yielded a detailed metabolic prediction for this organism, an overview of which is presented in Figure 2. Notably, the genome encodes an extensive repertoire of redox enzymes that likely indicate roles in iron and arsenic oxidation/reduction, nitrogen compound transformations, hydrogen metabolism, and fermentation. This mixture of aerobic and anaerobic pathways likely confers metabolic versatility, enabling this bacterium to proliferate under changing conditions close to the water table (Castelle et al., 2013).

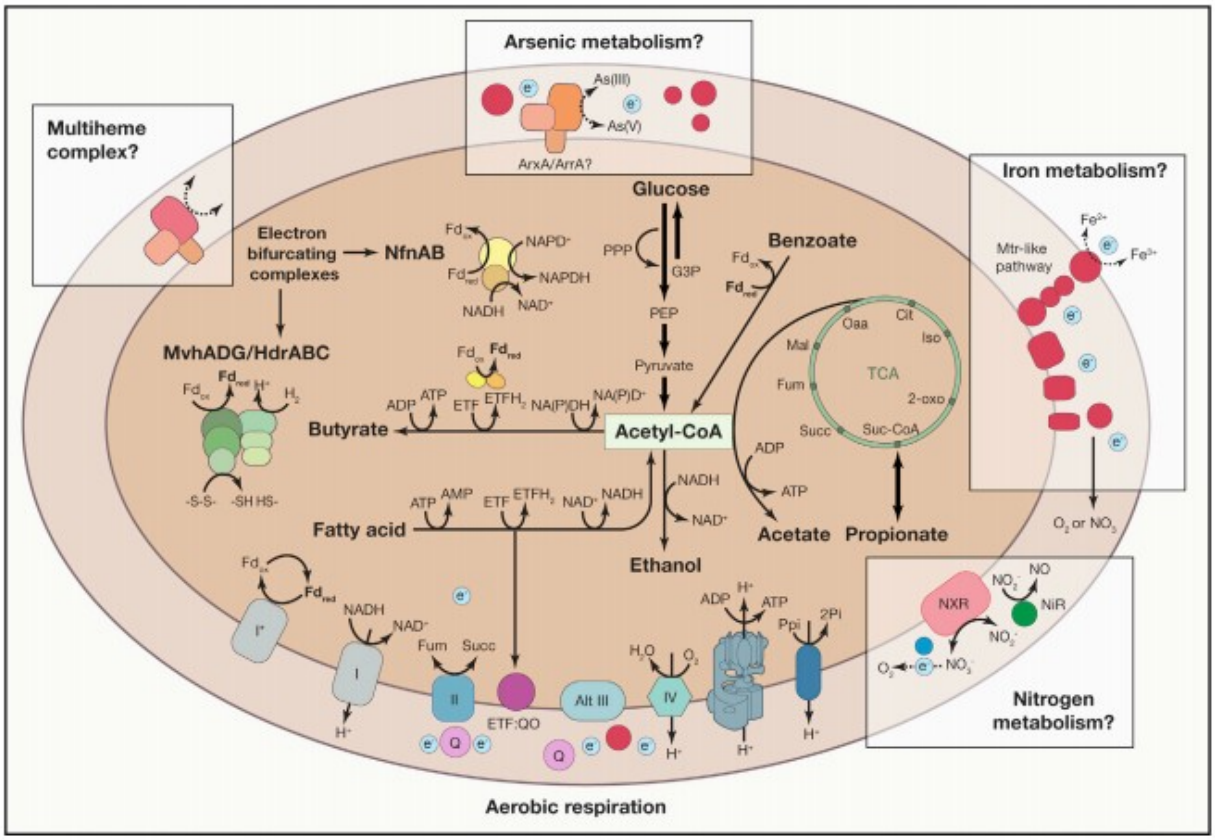

Figure 2. Metagenomics-Based Reconstruction of the Metabolism of Zixibacteria
This is a cell cartoon providing a simplified metabolic potential of Zixibacteria, a group of bacteria first reported from sediment by Castelle et al. (2013). Note the presence of pathways and complexes involved in aerobic growth (beta-oxidation of fatty acids and the terminal oxidase)—either oxidation or reduction of ferric/ferrous iron and arsenate/arenite and nitrate/nitrite, hydrogen metabolism, and anaerobic respiration via nitrite reduction, as well as fermentation to propionate, acetate, ethanol, and butyrate. Thus, overall, versatile metabolism enables Zixibacteria to thrive in a changing redox environment.
ArrA, arsenate reductase; ArxA, arsenite oxidase; NXR, nitrite/nitrate oxidoreductase; PPP, pentose phosphate pathway, PEP, phosphoenoylpyruvate; I, complex I or NADH dehydrogenase; I*, 11-subunit NADH dehydrogenase; II, complex II or succinate dehydrogenase; Alt III, alternative complex III; IV, complex IV or heme-copper oxygen reductase; ETQ:QO, electron transferring quinone oxidoreductase; NiR, nitrite reductase; Mtr, extracellular respiratory pathway that is essential for the reduction or oxidation of iron via multihemes cytochromes; MvhADG/HdrABC, cytoplasmic complex composed of the [NiFe]-hydrogenase MvhADG and the heterodisulfide reductase HdrABC, which is an iron-sulfur flavoprotein; Fd, ferredoxin. NfnAB complex is an iron-sulphur flavoprotein complex. All the red symbols represent c-type cytochromes.

A long-standing mystery is related to the apparent detection of Cyanobacteria in human fecal samples, suggesting the unexpected existence of these bacteria in gut microbiomes. Adult fecal samples that

were relatively enriched in these organisms (based on 16S rRNA gene sequencing) were targeted by genome-resolved metagenomics methods, and a complete (closed) genome was reconstructed (Di Rienzi et al., 2013). The study included an additional genome for a distinct group that was recovered from an acetate-amended sediment metagenome (Wrighton et al., 2012). Metabolic predictions for these genomes indicated the absence of photosynthetic machinery, leading to the conclusion that these bacteria have a fermentation-based metabolism. Based on phylogenetic analyses, the bacteria were assigned to the candidate phylum Melainabacteria (Di Rienzi et al., 2013). Genomes from a distinct but related lineage, Sericytochromatia, were subsequently detected in a coal-bed methane well, a laboratory bioreactor biofilm, and acetate-amended sediments (Soo et al., 2017). Soo et al. propose both Melainabacteria and Sericytochromatia as classes of Cyanobacteria, a disagreement that reflects the common challenge of achieving consensus regarding taxonomic designations. Due to the absence of photosynthetic machinery in lineages sibling to Cyanobacteria, Soo et al. confirm and extend the earlier conclusion (Di Rienzi et al., 2013) that photosynthesis evolved after the divergence of Cyanobacteria. Further, their analyses suggest that aerobic respiration arose after the evolution of photosynthesis.

Another distinct bacterial group is the candidate phylum Rokubacteria (Hug et al., 2015), first genomically described via metagenomic analysis of sediments collected at the same site as Zixibacteria and some Melainabacteria. These bacteria had been previously detected in soil 16S rRNA gene surveys. Based on metabolic predictions, it was suggested that the sediment-associated Rokubacteria are acetoclastic heterotrophs that likely use beta-oxidation of fatty acids for energy generation and produce butyrate that would be consumed by other community members. These Rokubacteria were also predicted to contribute to sulfur cycling (via oxidation of thiosulfate to sulfide and its reduction to hydrogen sulfide) and nitrogen cycling (via nitrite oxidation) and can probably oxidize carbon monoxide. Subsequently, it was proposed that Rokubacteria, which are relatively abundant in grassland soil, play a key role in carbon turnover via methanol oxidation (Butterfield et al., 2016). Rokubacteria were recently described as “genomic giants,” detected in the rhizosphere, volcanic mud, oil wells, aquifers, and the deep subsurface (Becraft et al., 2017).

Another important discovery involves bacteria of the candidate phylum Tectobacteria. The first genomically described members belong to the candidate genus *Entotheonella* and were described using metagenomic and single-cell sequencing methods targeted at microbial communities in the marine sponge *Theonella swinhoei* (Wilson et al., 2014). *Entotheonella* are predicted to produce a huge variety of bioactive compounds that may mediate ecological interactions. As sponges are well known as rich sources of diverse natural products, the research likely addressed the question of which organisms are the source of these compounds. The results of this

study underline the potential of genome-resolved metagenomics targeted to candidate phyla groups to uncover biosynthetic pathways for a vast treasure trove of secondary metabolites that could address the pressing need for new antimicrobial compounds and other pharmaceuticals.

Bacteria of the CPR and Their Similarities to DPANN Archaea

Although the first draft genomes for CPR were first reported only in 2012 (Wrighton et al., 2012), there are now thousands of sequences on hand. Based on the first ~800 genomes, Brown et al. (2015) proposed the existence of at least 35 candidate phyla within the CPR. Parallel cryogenic transmission electron microscope images that targeted post-0.2 μm filtrates collected from the same samples verified small cell volumes for these bacteria (see below; Luef et al., 2015). Interestingly, some candidate phyla within the radiation (Gracilibacteria, BD1-5 and Absconditabacteria, SR1) use an alternative genetic code (Campbell et al., 2013, Kantor et al., 2013, Rinke et al., 2013, Wrighton et al., 2012). The repurposing of the UGA stop codon to code for glycine was confirmed for an enrichment that contained Gracilibacteria by metaproteomics (Hanke et al., 2014). Building upon prior work by Wrighton et al., 2012, Kantor et al., 2013, Brown et al., 2015, Nelson and Stegen, 2015, and others, nantharaman et al. (2016 extended the early observation that CPR genomes are small and that most lack numerous biosynthetic pathways (Figure 3). Many are predicted to be unable to produce nucleotides *de novo* and have minimal amino acid and cofactor biosynthetic capacity. No CPR genomes analyzed to date contain the components necessary to synthesize membrane lipids required for the cell envelope, so further research is needed to determine the nature and sources of these components. The CPR bacteria have unusual ribosome compositions, and whole lineages (groups of putative candidate phyla) are missing what were considered to be universal ribosomal proteins (Brown et al., 2015). Parcubacteria have been reported to lack ribosomal small subunit methyltransferase G and ribosome-silencing factor (Nelson and Stegen, 2015).

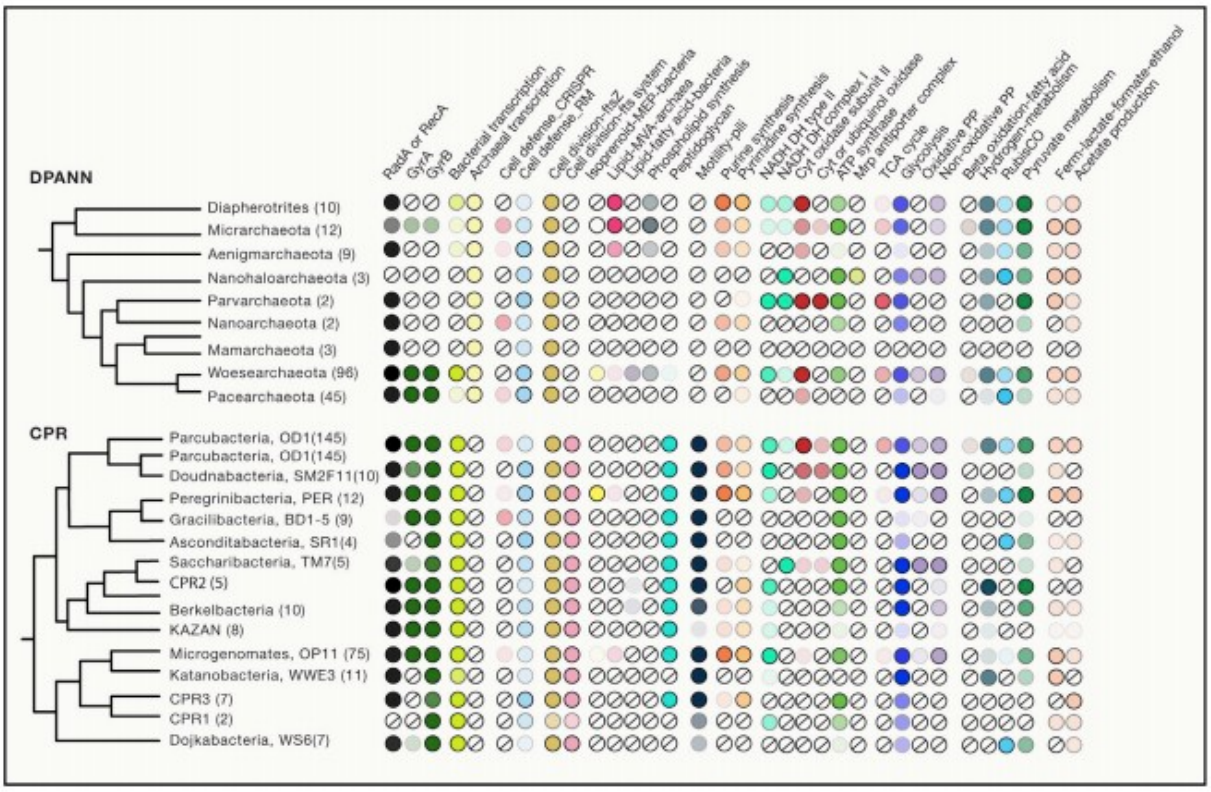

Figure 3. Biological traits across DPANN and CPR

A schematic tree of DPANN (left, top) and CPR (left, bottom) and corresponding overview of presence or absence (ø) of certain biological traits across both radiations. Numbers shown next to the candidate phylum name indicate the number of genomes used in the analysis. It should be noted that although certain capacities are shown as "present," they may not be found in the genomes of all members of the listed candidate phylum. Variation in shading in each column indicates the overall frequency of each capacity, with stronger colors signifying that the trait is widespread in the group.

In a new analysis reported here, we used a set of high-quality draft genomes and some complete genomes to provide a first glance into the metabolic variation within the CPR and DPANN simultaneously. As is evident from the descriptions above, limited metabolic capacities are common across both radiations and gaps in metabolic features are often shared (Figure 3). As noted previously, typically bacterial genes (e.g., bacterial transcription factors) occur in some DPANN archaea (Rinke et al., 2013). The distribution of isoprenoid-related genes is discussed further below.

Episymbiosis: An Unusual Lifestyle Predicted to Be Common for CPR and DPANN

A few CPR and DPANN species are associated with eukaryotic hosts (Gong et al., 2014). However, most are likely symbionts of bacteria or archaea, given their abundance and diversity in samples that have few, if any, eukaryotes. Based on strong enrichment in post-0.2 μm filtrates, it was predicted that many may be episymbionts—i.e., symbionts that associate with the surfaces of host cells rather than being contained within them. Cryogenic transmission electron microscopy (cryo-TEM) data show pili-like structures that extend from CPR cell surfaces (Figure 4A) and, in some cases, contact other microbial cells (Figures 4B and 4C; Luef et al., 2015). These

might provide access to nucleic acids or other metabolites. A few CPR and DPANN organisms have been directly shown to be episymbionts. For ARMAN (either Micrarchaetoa or Parvarchaeota), cryoelectron tomographic data revealed penetration of their cell interiors via cytoplasmic extensions from larger cells without a cell wall (Figure 4C; Baker et al., 2010, Comolli and Banfield, 2014). Based on the overall microbial composition, the larger cells were identified as Thermoplasmatales archaea. In other cases, cells are attached via short bridges to Thermoplasmatales cells (Figure 4D)—probably Gplasma (Baker et al., 2010)—which were renamed *Cuniculiplasma divulgatum* by Golyshina et al., (2016). These points of contact resemble those associated with *N. equitans* cells attached to host *Ignicoccus hospitalis* archaea in a co-culture of these organisms (Huber et al., 2002). Similarly, a co-culture of a parasitic Saccharibacterium (TM7) and its *Actinomyces odontolyticus* (Actinobacteria) host has been obtained from the human mouth (He et al., 2015). Transmission electron microscope images indicate host-episymbiont interaction via a region of cell-cell contact similar in form to those seen for ARMAN archaea and host Thermoplasmatales.

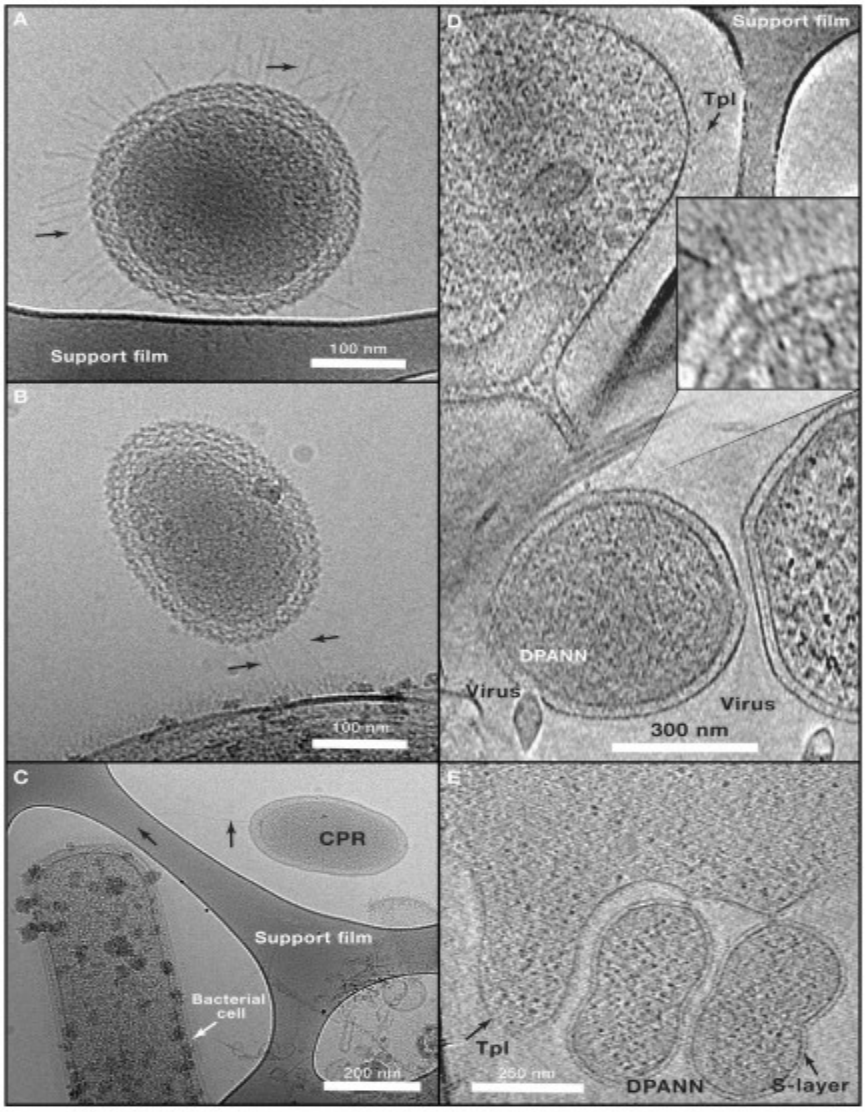

Figure 4. Cryo-TEM Images of CPR and DPANN Illustrating Close Association with Other Cells

Images of CPR modified from Luef et al. (2015) (A–C) and of DPANN from Baker et al. (2010) and Comolli and Banfield (2014) (D and E).

(A) Pili-like structures cover the surface of a CPR cell.

(B) Pili-like structures apparently connecting to the surface of an adjacent bacterium.

(C) Ultra-small cell with a pilus-like structure that extends to the surface of another large cell that is covered with nanoparticle aggregates.

(D) Interaction between DPANN (ARMAN) archaea and a Thermoplasmatales archaeon. Also shown is the association of viruses with an ARMAN cell (lower left and right).

(E) A Thermoplasmatales archaeon with two DPANN cells undergoing cell division.

Isoprenoid Lipid Biosynthesis Pathways Blur a Historical Distinction of Archaea and Bacteria

Isoprenoids are metabolites that are essential in all living organisms in the three domains of life. Isoprenoids have diverse metabolic functions, including as quinones, chlorophylls, bacteriochlorophylls, rhodopsins, and carotenoids. Archaeal membranes are composed of isoprenoid-based lipids, the isopentenyl pyrophosphate and dimethylallyl diphosphate precursors of which are typically synthesized via the mevalonate (MVA) pathway. Bacteria use a nonhomologous pathway—the methylerythritol phosphate (MEP) pathway—to synthesize the precursors. Previously, only few bacteria (mostly Gram positive) were known to possess the MVA pathway (Pasternak et al., 2013). Interestingly, most of these bacteria predate other bacteria. In new analyses, we revisited the question of the distribution of the MVA pathway in bacteria, making use of hundreds of new genomes from the CPR and DPANN radiation. The results show that some CPR bacteria (mostly Microgenomates and a few Peregrinibacteria) possess the MVA pathway rather than the bacterial MEP pathway (Figure 3). Intriguingly, the MVA pathway of the CPR is of the type found in eukaryotes. The eukaryote MVA pathway includes three enzymes that do not normally occur in in the archaeal MVA pathway (phosphomevalonate kinase [PMK], diphosphomevalonate carboxylase [MDC], and isopentenyl diphosphate isomerase [IDI1]).

As expected for Archaea, the MVA pathway occurs in organisms from several DPANN phyla, including Diapherotrites, Micrarchaeota, and Aenigmarchaeota. Similar to the CPR and distinct from other archaea, the DPANN MVA pathway includes the full set of enzymes found in eukaryotes. The MVA pathway is rarely found in Woesearchaeota. Instead, they have the bacterial MEP pathway. This is important because the MEP pathway has not been reported in archaea.

Previously, the presence of the MVA pathway in a few bacteria was explained by potential acquisition via horizontal gene transfer from archaeal or eukaryotic donors (Boucher and Doolittle, 2000, Wilding et al., 2000). Later, in-depth phylogenomic analyses indicated that the MVA pathway may have been ancestral in all three domains of life (Lombard and Moreira, 2011). We conducted phylogenetic analyses using two key enzymes (MVA kinase and PMK) from the MVA pathways to evaluate whether the new CPR and DPANN sequences throw light on the evolutionary origin of isoprenoid biosynthesis (Figure 5). The resulting phylogenetic tree resolves a novel intermediate clade for members of the Microgenomates superphylum that is placed at the base of both families (Figure 5). Of note, the Microgenomates that harbor the unusual new form of MVA kinase also possess the regular MVA kinase (Figure 5). The tree topology rules out recent horizontal gene transfer(s) from other archaea or eukaryotes as the explanation for the MVA pathway in CPR bacteria, with a few important exceptions (transfers from DPANN to Dependentiae [TM6] and from Firmicutes to Woesearchaeota; Figure 5).

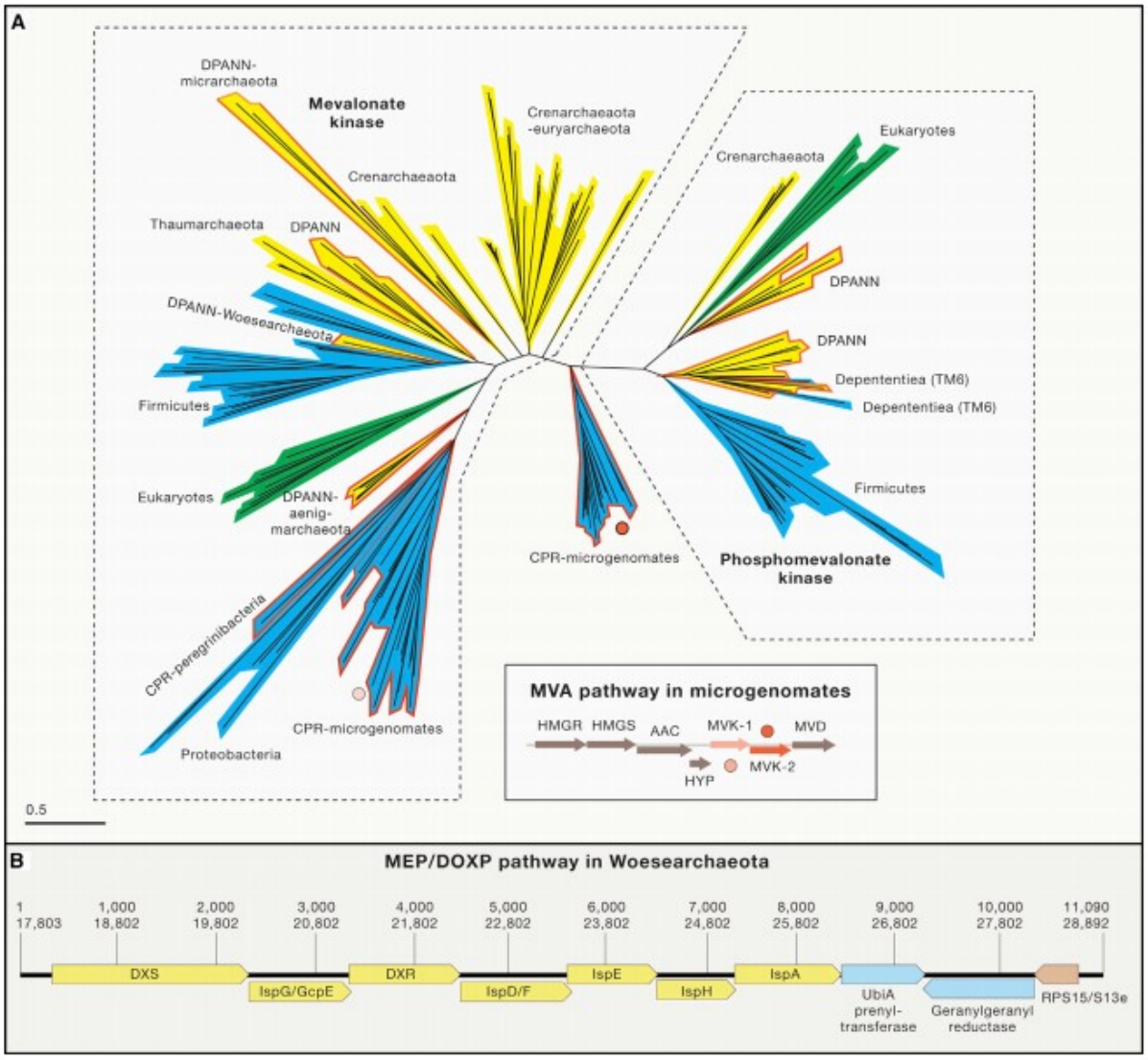

Figure 5. Isoprenoid Lipid Biosynthesis Pathways in CPR and DPANN Organisms

(A) Maximum likelihood tree of two key enzymes of the MVA pathway (with the MVA kinase on the left and the PMK on the right, both from the same GHMP kinase superfamily) involved in isoprenoid precursors production and present in some CPR and DPANN genomes. Bacteria are highlighted in blue and archaea are in yellow. Also included is the gene cluster organization of the MVA pathway found in some CPR organisms from the Microgenomates superphylum.

(B) Gene cluster organization of the MEP pathway identified in some Woesearchaeota genomes.

Abbreviations: HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; ACC, acetoacetyl-CoA thiolase; HYP, hypothetical; MVK-1, MVA kinase; MVK-2, new type of MVA kinase; DXS, 1-deoxy-dxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; IspG/GcpE, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; IspD, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; IspF, 2-Cmethyl-D-erythritol 2,4-cyclodiphosphate synthase; IspE, 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase; IspH, 4-hydroxy-3- methylbut-2-enyl diphosphate reductase; IspA, geranylgeranyl diphosphate synthase, type I.

The “lipid divide” has been a central feature used to establish the distinction between bacteria and archaea. The newly reported distribution of the MVA and the MEP pathways in both domains reopens the question of their evolutionary origin and alters the prior conclusion that the MEP pathway is restricted to bacteria.

How the CPR and DPANN May Change Our View of Evolution

As noted above, the availability of genomes for a much more comprehensive set of organisms revealed major new features of the tree of life (Figure 1). The question of whether the tree topology involves two major subgroups within bacteria and archaea—the CPR and DPANN, respectively—is unresolved. Are the apparent groupings of CPR and DPANN artifacts of rapid evolution or reflections of early origins and a long history of diversification?

A phenomenon that could drive fast evolution is genome reduction, which is an obvious consideration for CPR and DPANN, given their small cell and genome sizes. Largely symbiotic lifestyles are linked to rapid gene loss and fast accumulation of mutations in well-studied symbioses (e.g., those that involve partnerships of bacteria and eukaryotic hosts), although these may not be appropriate models for associations in which the host is bacterial or archaeal. The early stages of genome reduction in symbionts of eukaryotes are characterized by proliferation of mobile elements, formation of pseudogenes, multiple genomic rearrangements, and deletion of chromosome fragments. To date, such phenomena have not been described as prominent features of CPR and DPANN genomes (Nelson and Stegen, 2015). In more anciently evolved symbionts, such as *Buchnera aphidicola*, mobile elements and most pseudogenes have been eliminated (van Ham et al., 2003). Thus, if the CPR and DPANN experienced radiation-wide genome reduction, it may have occurred long ago.

It is important to note that the small genomes of bacterial symbionts of insects and those of human-associated pathogens, such as *Chlamydia*, do not cluster within the CPR. This suggests that the phylogenetic placement of the CPR is not an artifact of genome streamlining. That said, metabolisms predicted for CPR and DPANN vary dramatically, and while some have sufficient metabolic capacities to suggest that they may be free living, others lack numerous biosynthetic capacities. Thus, we predict that genome reduction may be an important phenomenon in some lineages.

The alternative explanation for the existence of CPR and DPANN as distinct major radiations is that both arose from very early-evolving organisms (possibly with small genomes), and the long branches of the CPR and DPANN are due mostly to undersampling (rather than rapid evolution). Although their apparent deep branching phylogenetic placement remains in question, other observations make an early origin of both groups from a common ancestral pool worth considering. Most important, perhaps, is the similarity in the suite of biosynthetic capacities that both groups possess (extensive genes of the information system) and in missing pathways (e.g., lack of the electron transport chain and tricarboxylic acid [TCA] cycles that are widely distributed in other bacteria and archaea). If, as one might predict, extensive use of oxygen-based respiration arose relatively late (following the advent of O_2-generating photosynthesis around 2 billion years ago), it would make sense that groups with ancient metabolic platforms would be anaerobes. The metabolisms of CPR and DPANN are consistent with such a world, as these organisms appear to be almost exclusively anaerobes, lacking a full TCA

cycle and electron transport chain required for aerobic growth. Similarly, almost all are incapable of dissimilatory nitrate reduction, and none are predicted to have metabolisms based on dissimilatory sulfate reduction. Both oxidized nitrogen and sulfur-based compounds would have been at low abundance prior to an oxidized atmosphere. Obviously, it is speculation, but we imagine that nucleotide and amino acid biosynthesis capacities may have been present in ancestral populations but lost during genome reduction as (most) CPR and DPANN adopted lifestyles that depended on organisms with more recently evolved capacities (e.g., aerobic growth, photosynthesis, and other forms or chemoautotrophy). The alternative explanation that cannot be ruled out at this time is that the many similarities in the genomic features of CPR and DPANN may have arisen by convergent evolution, possibly involving late radiation-wide adaption to anaerobic habitats.

Although what we see now in CPR and DPANN genomes must be a faint echo of what once was, the presence of specific genes in widely divergent groups may hint at features of ancestral organisms that were lost from many modern groups. Included in this list are capacities such as modified glycolysis (reliant on pentose phosphate pathway enzymes), hydrogen metabolism (as suggested for the ancestor of Archaea; Williams et al., 2017), and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). In fact, one of the most striking cases of widely but sparsely distributed enzymes found in both CPR and DPANN organisms are the form II/III- and III-like RuBisCOs that apparently function in a nucleotide-based pathway that feeds into lower glycolysis and fermentation (Wrighton et al., 2016). This enzyme occurs in some extremely minimal genomes (Dojkabacteria [WS6] and Pacearchaeota), consistent with its central role in the metabolism of these organisms. Thus, RuBisCO that functions in nucleotide metabolism and central carbon metabolism may have been key to the physiology of ancestral CPR and DPANN. A bacterial or archaeal cell is, on average, 20% RNA, and RNA is 40% ribose by weight (i.e., a cell is about 8% ribose; Schönheit et al., 2016). Therefore, it has been suggested that ribose was likely an abundant sugar available on early Earth for fermentation. Hence, type III- and II/III-based RuBisCO pathway of nucleoside monophosphate conversion to 3-phosphoglycerate may be a relic of ancient heterotrophy.

The alternative explanation for the very striking patterns of similarity in the presence and absence of genes is a combination of gene loss and horizontal gene transfer. Horizontal gene transfer certainly has shaped the evolution of CPR and DPANN (Jaffe et al., 2016), but this remains relatively unstudied. So far, where more complex metabolic capacities occur, they are present sporadically, and the enzymes are highly divergent relative to all currently known sequences. Likely, the evolutionary trajectories of CPR and DPANN have been shaped by combination of gene loss, lateral transfer, and convergent evolution. The balance in importance of these processes across the radiations is very much an open question.

Asgard Archaea Reinforce the Idea that Archaea Were Ancestors of Eukaryotic Cells

The Asgard superphylum, comprising Lokiarchaeota (Spang et al., 2015), Thorarchaeota (Seitz et al., 2016), Odinarchaeota (aremba-Niedzwiedzka et al., 201), and Heimdallarchaeota (aremba-Niedzwiedzka et al., 201) represents an important recent discovery from metagenomics. Based on phylogenetic analyses that included sequences from members of the Asgard superphylum, it was proposed that these archaea share a relatively close evolutionary relationship with Eukaryotes (Spang et al., 2015). There have been many competing hypotheses regarding the origin of eukaryotic cells, and the idea that they arose from the fusion of bacterial and archaeal cells has a long history. Jim Lake and colleagues suggested this in the 1980s (the “eocyte” hypothesis; e.g, Lake, 2015, Lake et al., 1984, Williams et al., 2013). A mixed origin for the eukaryotic cell involving both bacteria and archaea was supported by the fact that genes for DNA replication, transcription, and translation systems of eukaryotes resemble archaeal genes, whereas genes involved in energy and carbon metabolism are similar to genes of bacteria. Martin and Müller (1998) proposed that Eukaryotes arose through symbiotic association of an anaerobic, autotrophic, hydrogen-dependent archaeal host with bacterial symbiont. Others suggested that they arose via symbiosis between methanogenic archaea and an ancestral sulfate-reducing myxobacterium (Moreira and Lopez-Garcia, 1998). The claim regarding the phylogenetic placement of Asgard archaea with Eukaryotes (Spang et al., 2015) was an important new twist and considered controversial by some (Da Cunha et al., 2017).

The phylogenetic argument of Spang et al. (2015) is supported by recent and more comprehensive studies (aremba-Niedzwiedzka et al., 201). Important discoveries in the original and new studies relate to the presence of numerous genes and pathways in Asgard archaea that are normally considered to be eukaryotic signatures (eukaryotic signature proteins, ESPs), thus mostly absent in other archaea. These include some components related to membrane remodeling and trafficking, cytoskeletal features, ubiquitin-modifier systems, tubulins, and vesicle-formation processes in eukaryotes (actin, small Ras GTPases, ESCRT complex; Spang et al., 2015, Zaremba-Niedzwiedzka et al., 2017). Although genes encoding some of these proteins had been noted previously in archaeal genomes (Ettema et al., 2011, Hartman and Fedorov, 2002, Lindås et al., 2008, Yutin and Koonin, 2012), the presence of multiple components adds significant weight to claims based on phylogenetic reconstructions. Thus, Asgard archaea likely provide important clues regarding the nature of the precursors of the first eukaryotic cells. Although it must be acknowledged that the biochemical functions of the intriguing eukaryotic features of Asgard archaea remain to be determined, the presence of such a repertoire may imply that if the ancestor of eukaryotes was archaeal, the organisms were far more complex than otherwise predicted.

Metabolic Handoffs, Microbial Communities

A strength of genome-resolved metagenomics methods is that they provide information about the metabolic potential of coexisting community members. Thus, in addition to resolving which organism can do what (Figure 6A), it is possible to examine the distributions of traits over coexisting organisms within an ecosystem (Anantharaman et al., 2016, Dombrowski et al., 2017, Wrighton et al., 2014). The Anantharaman et al. analysis leveraged over 2,500 genomes from a single aquifer environment (representing around ⅓ of all organisms detected there) and revealed evidence for extensive interconnection among the metabolisms of coexisting community members. These connections involve the supply and consumption of compounds of the nitrogen, sulfur, carbon, hydrogen, and other biogeochemical cycles. For example, an organism may be able to conduct only a subset of the redox transformations in a biogeochemical pathway. Thus, they release a partially oxidized (e.g., thiocyanite that can be oxidized to sulfite) or partially reduced (e.g., nitrite that can be reduced to nitrous oxide) intermediate product for use by other organisms. This ecosystem structure is counter to the expectation that, for example, a single bacterium is responsible for all steps in nitrate reduction. The phenomenon, involving cooperative biogeochemical cycling, has been referred to as “metabolic handoffs” (Anantharaman et al., 2016, Hug et al., 2015) and is diagrammed in Figure 6B. Through metabolic handoffs, the ecosystem can adopt a large number of configurations by shifting the mixture of organisms involved in a process as environmental conditions change. Complex interdependencies, and especially obligate symbioses, may explain why some organisms are difficult to grow in pure cultures (as isolates).

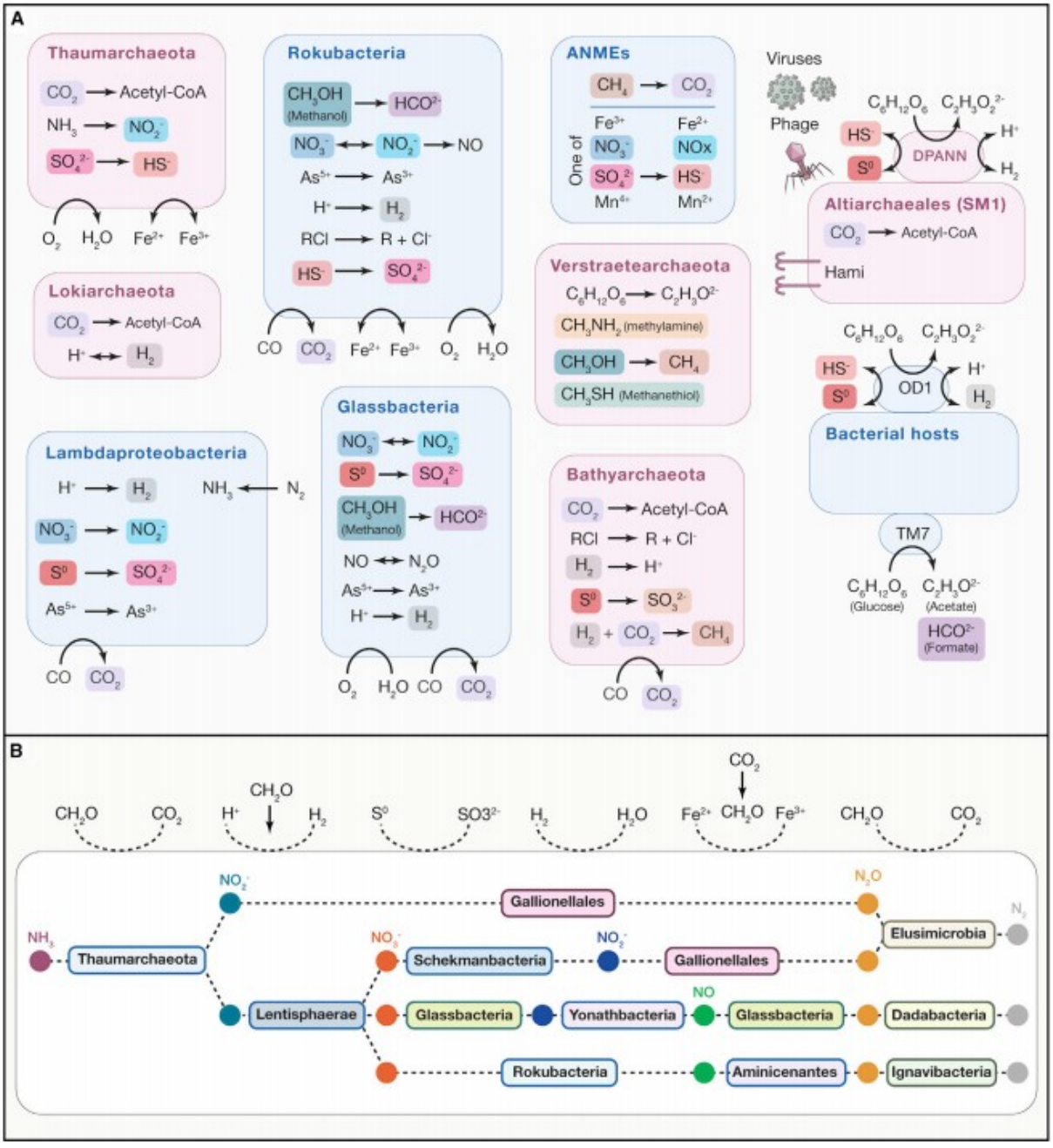

Figure 6: Information from genome-resolved analysis of natural microbial communities. (A) Predicted metabolic features for some newly described bacteria and archaea from a range of different sites and studies (for references, please see the text). The simplified cell cartoons highlight how genomics-based analyses of organisms from candidate phyla have revealed their biogeochemically important roles. Organisms are represented in the context of sediment, an environment that hosts many newly described bacteria and archaea, but they occur in other environments. Genomes of bacteriophage (phage), viruses (of archaea), and plasmids can also be reconstructed, enabling direct analysis of the roles of these entities in augmenting host metabolism and lateral gene transfer. (B) Diagram illustrating the concept of coexisting microorganisms working together to complete a single biogeochemical pathway (ammonia, NH3, to N2). This is based on Figure 6 and traits data in Supplementary Table 9 of nantharaman et al. (2016. Examples of coupling of reduction of nitrogen-based compounds to the iron, carbon, hydrogen, and sulfur cycles are included. The phenomena

illustrated are referred to as “metabolic handoffs” because the product of one organism is made available to another organism with the capacity to use it. Importantly, Anantharaman et al. demonstrated that few organisms have the ability to generate energy from all of the steps in most pathways analyzed. Note that essentially all of the organisms studied are only distantly related to isolated organisms.For taxonomic affiliations, see Table S1.

Concluding Thoughts and Some Frontiers for Future Research

The tsunami of new genome sequences for bacteria and archaea has dramatically reshaped our understanding of life’s diversity and expanded knowledge regarding microbial roles in biogeochemical processes that impact atmospheric chemistry, soil fertility, water composition, and human health. The genomes provide context for metaproteomic and metatranscriptomic data, which can identify the pathways that are operational at any time on an organism-by-organism basis (e.g., Frias-Lopez et al., 2008, Ram et al., 2005). The availability of new genomic information for candidate phyla organisms, arguably the majority of all bacteria and archaea, has demanded that we rethink concepts related to microbial community structure and functioning, the nature of symbioses, and the origins of evolutionary innovations. Huge and fundamental questions remain regarding early evolution that can only be addressed by better and more detailed phylogenetic analyses that leverage more comprehensive sets of genomes.

The scale of microbial diversity in Domain Eukarya is likely underestimated. A major advance anticipated in the coming years is the discovery, metabolic exploration, and analysis of the habitat distribution of microbial eukaryotes. A few studies have already employed cultivation-independent genomic methods for *de novo* reconstruction of genomes of microbial and even macro-eukaryotes from short read datasets (Kantor et al., 2017, Mosier et al., 2016, Sharon et al., 2013). Although approaches are improving rapidly (West et al., 2017), genome-resolved metagenomics methods have not yet found broad application for this purpose. The combination of cultivation-independent genomic information with transcriptomic data may resolve how the metabolism of microbial eukaryotes depends on the types and activities of associated bacteria and archaea. For example, laboratory studies using a single genotype of the human pathogen *Candida albicans* (Eukaryota, Opisthokonta, Fungi) showed that it underwent a phenotypic switch, with dramatic alteration of its gene expression patterns, when it was grown in the presence of different bacterial isolates (Fox et al., 2014). Such phenomena could be directly tested by genomically defining *Candida* in real human gut microbial communities via genome-resolved metagenomics and then evaluating its transcriptome (via metatranscriptomics) as a function of the flanking microbial community composition.

Medical research has been substantially impacted by the introduction of cultivation-independent microbiome studies. However, genomically resolved analyses remain in the minority, partially due to the high costs and slow speed relative to rRNA sequence-based analyses. We anticipate that this will

change with the advent of new, real-time, and low-cost sequencing methodologies. Although rRNA gene surveys and existing metagenomic data suggest that it is unlikely that many major new microbial groups (e.g., phyla) will be discovered in the human microbiome, new species-, genus- and higher-taxonomic-level groups certainly will be uncovered. Importantly, genome sequences for even closely related strains can differ significantly, and variation may involve genes that contribute to pathogenicity and antibiotic resistance. Thus, metagenomic methods have the potential to provide new insights into the medical importance and drug susceptibility of human-associated microbes and ultimately could guide medical treatments.

An understudied question relates the medical significance of CPR and DPANN. Overall, they do not appear to be major players in most human microbiomes. Saccharibacteria (TM7) and Absconditabacteria (SR1) have been found in the mouth (Dewhirst et al., 2010). TM7 have been linked to periodontal disease, and one subtype is found primarily at diseased sites (Brinig et al., 2003). Members of this phylum are also associated with dermatitis and cystic fibrosis (He et al., 2015), and they were detected in other mucus-rich sites associated with vaginosis and inflammatory bowel disease (Fredricks et al., 2005, Kuehbacher et al., 2008). Importantly, experiments involving co-culture of a TM7 with its host showed that TM7 could modulate the immune response (He et al., 2015). Intriguingly, bacteria from these groups, along with a variety of DPANN archaea, are relatively common in dolphin mouths (Dudek et al., 2017), suggesting their broader importance in animal health. Woesearchaeota were recently detected in the lungs of humans (Koskinen et al., 2017), and cell-free DNA of certain CPR bacteria was found in the blood of many human subjects (Kowarsky et al., 2017). Both CPR and DPANN are abundant in groundwater and thus may be ingested regularly by humans and animals. It is our view that PCR-independent methods applied to human microbiome samples from diverse body sites, disease states, and populations could provide new information relating to the potential medical significance of CPR and DPANN.

Throughout this Perspective, our focus has been on genomics, i.e., what can be learned from DNA sequence information. Ultimately, however, experiments are needed to address the biological relevance of the predictions informed by sequence analysis. The landmark study involving the TM7 co-culture with its actinobacterial host provides a template for future work. Metagenomics studies provide information about appropriate environmental sample sources that could be used to seed enrichments. Co-abundance patterns for CPR or DPANN and other bacteria or archaea could be used to identify possible symbiont-host associations. Based on host gene content, it should be possible to design enrichment studies to bring other CPR and DPANN into co-cultures. Going beyond the physiological and morphological studies that co-cultures enable, specific organisms in enrichments could be targeted using modern genome-editing methods to

remove or upregulate genes in either the symbiont or host to discover the mechanisms of their interaction.

With sufficient genomic and linked functional data, it will be possible to really begin to understand how natural ecosystems function. Comprehensive “omics” information is already being collected from extensive suites of time series and spatial samples, as well as from laboratory bioreactors and full field-scale manipulation experiments. Ultimately, across ecosystem types, it should be possible to link organisms to processes and to understand the conditions that cause organisms of interest to proliferate. Thus, genomically informed ecosystem (including human microbiome) manipulations should soon come within reach.

Acknowledgments

Support was provided by grants from the Lawrence Berkeley National Laboratory’s Genomes-to-Watershed Scientific Focus Area . The U.S. Department of Energy (DOE), Office of Science, and Office of Biological and Environmental Research funded the work under contract DE-AC02-05CH11231 and the DOE carbon cycling program DOE-SC10010566 , the Innovative Genomics Institute at Berkeley and the Chan Zuckerberg Biohub . We thank the editor and two anonymous reviewers for excellent input to this manuscript.

References

Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J. 11, 2407–2425.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219.

Baker, B.J., and Dick, G.J. (2013). Omic Approaches in Microbial Ecology: Charting the Unknown. Microbe 8, 353–360.

Baker, B.J., Tyson, G.W., Webb, R.I., Flanagan, J., Hugenholtz, P., Allen, E.E., and Banfield, J.F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. Science 314, 1933–1935.

Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., Verberkmoes, N.C., Hettich, R.L., and Banfield, J.F. (2010). Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl. Acad. Sci. USA 107, 8806–8811.

Baker, B.J., Saw, J.H., Lind, A.E., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., and Ettema, T.J.G. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. Nat. Microbiol. 1, 16002.

Beal, E.J., House, C.H., and Orphan, V.J. (2009). Manganese- and iron-dependent marine methane oxidation. Science 325, 184–187.

Becraft, E.D., Woyke, T., Jarett, J., Ivanova, N., Godoy-Vitorino, F., Poulton, N., Brown, J.M., Brown, J., Lau, M.C.Y., Onstott, T., et al. (2017). Rokubacteria: Genomic Giants among the Uncultured Bacterial Phyla. Front. Microbiol. 8, 2264.

Bird, J.T., Baker, B.J., Probst, A.J., Podar, M., and Lloyd, K.G. (2016). Culture Independent Genomic Comparisons Reveal Environmental Adaptations for Altiarchaeales. Front. Microbiol. 7, 1221.

Borrel, G., Parisot, N., Harris, H.M.B., Peyretaillade, E., Gaci, N., Tottey, W., Bardot, O., Raymann, K., Gribaldo, S., Peyret, P., et al. (2014). Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics 15, 679.

Boucher, Y., and Doolittle, W.F. (2000). The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Mol. Microbiol. 37, 703–716.

Brinig, M.M., Lepp, P.W., Ouverney, C.C., Armitage, G.C., and Relman, D.A. (2003). Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. Appl. Environ. Microbiol. 69, 1687–1694.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat. Rev. Microbiol. 6, 245–252.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523, 208–211.

Butterfield, C.N., Li, Z., Andeer, P.F., Spaulding, S., Thomas, B.C., Singh, A., Hettich, R.L., Suttle, K.B., Probst, A.J., Tringe, S.G., et al. (2016). Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ 4, e2687.

Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T., So¨ ll, D., and Podar, M. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc. Natl. Acad. Sci. USA 110, 5540–5545.

Castelle, C.J., Hug, L.A., Wrighton, K.C., Thomas, B.C., Williams, K.H., Wu, D., Tringe, S.G., Singer, S.W., Eisen, J.A., and Banfield, J.F. (2013). Extraordinary

phylogenetic diversity and metabolic versatility in aquifer sediment. Nat. Commun. 4, 2120.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr. Biol. 25, 690–701.

Chen, L.-X., Me´ ndez-Garcı´a, C., Dombrowski, N., Servı´n-Garciduen˜ as, L.E., Eloe-Fadrosh, E.A., Fang, B.-Z., Luo, Z.-H., Tan, S., Zhi, X.-Y., Hua, Z.-S., et al. (2017). Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. ISME J. Published online December 8, 2017. https://doi.org/10. 1038/s41396-017-0002-z.

Comolli, L.R., and Banfield, J.F. (2014). Inter-species interconnections in acid mine drainage microbial communities. Front. Microbiol. 5, 367.

Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS Genet. 13, e1006810.

Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C.R., Yu, W.-H., Lakshmanan, A., and Wade, W.G. (2010). The human oral microbiome. J. Bacteriol. 192, 5002–5017.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., and Ley, R.E. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. eLife 2, e01102.

Dojka, M.A., Harris, J.K., and Pace, N.R. (2000). Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of bacteria. Appl. Environ. Microbiol. 66, 1617–1621.

Dombrowski, N., Seitz, K.W., Teske, A.P., and Baker, B.J. (2017). Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiome 5, 106.

Dudek, N.K., Sun, C.L., Burstein, D., Kantor, R.S., Aliaga Goltsman, D.S., Bik, E.M., Thomas, B.C., Banfield, J.F., and Relman, D.A. (2017). Novel Microbial Diversity and Functional Potential in the Marine Mammal Oral Microbiome. Curr. Biol. 27, 3752–3762.

Ettema, T.J.G., Linda˚ s, A.-C., and Bernander, R. (2011). An actin-based cytoskeleton in archaea. Mol. Microbiol. 80, 1052–1061.

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science 350, 434–438.

Fox, E.P., Cowley, E.S., Nobile, C.J., Hartooni, N., Newman, D.K., and Johnson, A.D. (2014). Anaerobic bacteria grow within Candida albicans biofilms and induce biofilm formation in suspension cultures. Curr. Biol. 24, 2411–2416.

Fredricks, D.N., Fiedler, T.L., and Marrazzo, J.M. (2005). Molecular identification of bacteria associated with bacterial vaginosis. N. Engl. J. Med. 353, 1899–1911.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008). Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. USA 105, 3805–3810.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175–188.

Golyshina, O.V., Kublanov, I.V., Tran, H., Korzhenkov, A.A., Lu¨ nsdorf, H., Nechitaylo, T.Y., Gavrilov, S.N., Toshchakov, S.V., and Golyshin, P.N. (2016). Biology of archaea from a novel family Cuniculiplasmataceae (Thermoplasmata) ubiquitous in hyperacidic environments. Sci. Rep. 6, 39034.

Gong, J., Qing, Y., Guo, X., and Warren, A. (2014). ‘‘Candidatus Sonnebornia yantaiensis’’, a member of candidate division OD1, as intracellular bacteria of the ciliated protist Paramecium bursaria (Ciliophora, Oligohymenophorea). Syst. Appl. Microbiol. 37, 35–41.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. Science 305, 1457–1462.

Hanke, A., Hamann, E., Sharma, R., Geelhoed, J.S., Hargesheimer, T., Kraft, B., Meyer, V., Lenk, S., Osmers, H., Wu, R., et al. (2014). Recoding of the stop codon UGA to glycine by a BD1-5/SN-2 bacterium and niche partitioning between Alpha- and Gammaproteobacteria in a tidal sediment microbial community naturally selected in a laboratory chemostat. Front. Microbiol. 5, 231.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., and Tyson, G.W. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature 500, 567–570.

Harris, J.K., Kelley, S.T., and Pace, N.R. (2004). New perspective on uncultured bacterial phylogenetic division OP11. Appl. Environ. Microbiol. 70, 845–849.

Hartman, H., and Fedorov, A. (2002). The origin of the eukaryotic cell: a genomic investigation. Proc. Natl. Acad. Sci. USA 99, 1420–1425.

He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.-Y., Dorrestein, P.C., Esquenazi, E., Hunter, R.C., Cheng, G., et al. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc. Natl. Acad. Sci. USA 112, 244–249.

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., Sievert, S.M., and Wang, F. (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. Nat. Microbiol. 1, 16035.

Hernsdorf, A.W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K., Probst, A., Burstein, D., Thomas, B.C., and Banfield, J.F. (2017). Potential for microbial H2and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. ISME J. 11, 1915–1929.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417, 63–67.

Hug, L.A., Thomas, B.C., Sharon, I., Brown, C.T., Sharma, R., Hettich, R.L., Wilkins, M.J., Williams, K.H., Singh, A., and Banfield, J.F. (2015). Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ. Microbiol.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. Nat. Microbiol. 1, 16048.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., and Pace, N.R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. 180, 366–376.

Ino, K., Hernsdorf, A.W., Konno, U., Kouduka, M., Yanagawa, K., Kato, S., Sunamura, M., Hirota, A., Togo, Y.S., Ito, K., et al. (2018). Ecological and genomic profiling of anaerobic methane-oxidizing archaea in a deep granitic environment. ISME J. 12, 31–47.

Jaffe, A.L., Corel, E., Pathmanathan, J.S., Lopez, P., and Bapteste, E. (2016). Bipartite graph analyses reveal interdomain LGT involving ultrasmall prokaryotes and their divergent, membrane-related proteins. Environ. Microbiol. 18, 5072–5081.

Kantor, R.S., Wrighton, K.C., Handley, K.M., Sharon, I., Hug, L.A., Castelle, C.J., Thomas, B.C., and Banfield, J.F. (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio 4, e00708–e00713.

Kantor, R.S., Huddy, R.J., Iyer, R., Thomas, B.C., Brown, C.T., Anantharaman, K., Tringe, S., Hettich, R.L., Harrison, S.T.L., and Banfield, J.F. (2017). Genome-Resolved Meta-Omics Ties Microbial Dynamics to Process Performance in Biotechnology for Thiocyanate Degradation. Environ. Sci. Technol. 51, 2944–2953.

Koskinen, K., Pausan, M.R., Perras, A.K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., and Moissl-Eichinger, C. (2017). First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. MBio 8, e00824–17.

Kowarsky, M., Camunas-Soler, J., Kertesz, M., De Vlaminck, I., Koh, W., Pan, W., Martin, L., Neff, N.F., Okamoto, J., Wong, R.J., et al. (2017). Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. Proc. Natl. Acad. Sci. USA 114, 9623–9628.

Kuehbacher, T., Rehman, A., Lepage, P., Hellmig, S., Fo¨ lsch, U.R., Schreiber, S., and Ott, S.J. (2008). Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease. J. Med. Microbiol. 57, 1569–1576.

Lake, J.A. (2015). Eukaryotic origins. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370, 20140321. Lake, J.A., Henderson, E., Oakes, M., and Clark, M.W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc. Natl. Acad. Sci. USA 81, 3786–3790.

Lazar, C.S., Baker, B.J., Seitz, K., Hyde, A.S., Dick, G.J., Hinrichs, K.-U., and Teske, A.P. (2016). Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. Environ. Microbiol. 18, 1200–1211.

Lindas, A.-C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J.G., and Bernander, R. (2008). A unique cell division machinery in the Archaea. Proc. Natl. Acad. Sci. USA 105, 18942–18946.

Lombard, J., and Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. Mol. Biol. Evol. 28, 87–99.

Luef, B., Frischkorn, K.R., Wrighton, K.C., Holman, H.-Y.N., Birarda, G., Thomas, B.C., Singh, A., Williams, K.H., Siegerist, C.E., Tringe, S.G., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. Nat. Commun. 6, 6372.

Martin, W., and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. Nature 392, 37–41. Moreira, D., and Lopez-Garcia, P. (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. 47, 517–530.

Mosier, A.C., Miller, C.S., Frischkorn, K.R., Ohm, R.A., Li, Z., LaButti, K., Lapidus, A., Lipzen, A., Chen, C., Johnson, J., et al. (2016). Fungi Contribute Critical but Spatially Varying Roles in Nitrogen and Carbon Cycling in Acid Mine Drainage. Front. Microbiol. 7, 238.

Mwirichia, R., Alam, I., Rashid, M., Vinu, M., Ba-Alawi, W., Anthony Kamau, A., Kamanda Ngugi, D., Go¨ ker, M., Klenk, H.-P., Bajic, V., and Stingl, U. (2016). Metabolic traits of an uncultured archaeal lineage–MSBL1–from brine pools of the Red Sea. Sci. Rep. 6, 19181.

Narasingarao, P., Podell, S., Ugalde, J.A., Brochier-Armanet, C., Emerson, J.B., Brocks, J.J., Heidelberg, K.B., Banfield, J.F., and Allen, E.E. (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J. 6, 81–93.

Nelson, W.C., and Stegen, J.C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front. Microbiol. 6, 713.

Nobu, M.K., Narihiro, T., Kuroda, K., Mei, R., and Liu, W.-T. (2016). Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME J. 10, 2478–2487.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276, 734–740.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., and Kyrpides, N.C. (2016). Uncovering Earth's virome. Nature 536, 425–430.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542.

Pasternak, Z., Pietrokovski, S., Rotem, O., Gophna, U., Lurie-Weinberger, M.N., and Jurkevitch, E. (2013). By their genes ye shall know them: genomic signatures of predatory bacteria. ISME J. 7, 756–769.

Pester, M., Schleper, C., and Wagner, M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr. Opin. Microbiol. 14, 300–306.

Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., and Keller, M. (2007). Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. 73, 3205–3214.

Probst, A.J., Weinmaier, T., Raymann, K., Perras, A., Emerson, J.B., Rattei, T., Wanner, G., Klingl, A., Berg, I.A., Yoshinaga, M., et al. (2014). Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nat. Commun. 5, 5497.

Probst, A.J., Ladd, B., Jarett, J.K., Geller-McGrath, D.E., Sieber, C.M.K., Emerson, J.B., Anantharaman, K., Thomas, B.C., Malmstrom, R.R., Stieglmeier, M., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat. Microbiol. https://doi.org/10.1038/s41564-017-0098-y.

Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., 2nd, Shah, M., Hettich, R.L., and Banfield, J.F. (2005). Community proteomics of a natural microbial biofilm. Science 308, 1915–1920.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431–437.

Schonheit, P., Buckel, W., and Martin, W.F. (2016). On the Origin of Heterotrophy. Trends Microbiol. 24, 12–25.

Schulz, F., Eloe-Fadrosh, E.A., Bowers, R.M., Jarett, J., Nielsen, T., Ivanova, N.N., Kyrpides, N.C., and Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. Microbiome 5, 140.

Seitz, K.W., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., and Baker, B.J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME J. 10, 1696–1705.

Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., and Banfield, J.F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res. 23, 111–120.

Simmons, S.L., Dibartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. (2008). Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. PLoS Biol. 6, e177.

Soo, R.M., Hemp, J., Parks, D.H., Fischer, W.W., and Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science 355, 1436–1440.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179.

Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357.

Stahl, D.A., and de la Torre, J.R. (2012). Physiology and diversity of ammoniaoxidizing archaea. Annu. Rev. Microbiol. 66, 83–101.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.

van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Ferna´ ndez, J.M., Jime´ nez, L., Postigo, M., Silva, F.J., et al. (2003). Reductive genome evolution in Buchnera aphidicola. Proc. Natl. Acad. Sci. USA 100, 581–586.

Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., and Tyson, G.W. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. Nat. Microbiol. 1, 16170.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

West, P.T., Probst, A.J., Grigoriev, I.V., Thomas, B.C., and Banfield, J.F. (2017). Genome-reconstruction for eukaryotes from complex natural microbial communities. bioRxiv. https://doi.org/10.1101/171355.

Wilding, E.I., Brown, J.R., Bryant, A.P., Chalker, A.F., Holmes, D.J., Ingraham, K.A., Iordanescu, S., So, C.Y., Rosenberg, M., and Gwynn, M.N. (2000). Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. J. Bacteriol. 182, 4319–4327.

Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature 504, 231–236.

Williams, T.A., Szollosi, G.J., Spang, A., Foster, P.G., Heaps, S.E., Boussau, B., Ettema, T.J.G., and Embley, T.M. (2017). Integrative modeling of gene and genome evolution roots the archaeal tree of life. Proc. Natl. Acad. Sci. USA 114, E4602–E4611.

Wilson, M.C., Mori, T., Ru¨ ckert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., et al. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74, 5088–5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA 87, 4576–4579.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337, 1661–1665.

Wrighton, K.C., Castelle, C.J., Wilkins, M.J., Hug, L.A., Sharon, I., Thomas, B.C., Handley, K.M., Mullin, S.W., Nicora, C.D., Singh, A., et al. (2014). Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. ISME J. 8, 1452–1463.

Wrighton, K.C., Castelle, C.J., Varaljay, V.A., Satagopan, S., Brown, C.T., Wilkins, M.J., Thomas, B.C., Sharon, I., Williams, K.H., Tabita, F.R., and Banfield, J.F. (2016). RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. ISME J. 10, 2702–2714.

Yutin, N., and Koonin, E.V. (2012). Archaeal origin of tubulin. Biol. Direct 7, 10. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Ba¨ ckstro¨ m, D.,

Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358.