Extremophiles
DOI 10.1007/s00792-014-0664-7

SPECIAL ISSUE: REVIEW

10th International Congress on Extremophiles

Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter"

Brian P. Hedlund · Jeremy A. Dodsworth · Senthil K. Murugapiran · Christian Rinke · Tanja Woyke

Received: 2 April 2014 / Accepted: 5 June 2014
© Springer Japan 2014

Abstract Despite >130 years of microbial cultivation studies, many microorganisms remain resistant to traditional cultivation approaches, including numerous candidate phyla of bacteria and archaea. Unraveling the mysteries of these candidate phyla is a grand challenge in microbiology and is especially important in habitats where they are abundant, including some extreme environments and low-energy ecosystems. Over the past decade, parallel advances in DNA amplification, DNA sequencing and computing have enabled rapid progress on this problem, particularly through metagenomics and single-cell genomics. Although each approach suffers limitations, metagenomics and single-cell genomics are particularly powerful when combined synergistically. Studies focused on extreme environments have revealed the first substantial genomic information for several candidate phyla, encompassing putative acidophiles (Parvarchaeota), halophiles (Nanohaloarchaeota), thermophiles (Acetothermia, Aigarchaeota, Atribacteria, Calescamantes, Korarchaeota, and Fervidibacteria), and piezophiles (Gracilibacteria). These data have enabled insights into the biology of these organisms, including catabolic and anabolic potential, molecular adaptations to life in extreme environments, unique genomic features such as stop codon reassignments, and predictions about cell ultrastructure. In addition, the rapid expansion of genomic coverage enabled by these studies continues to yield insights into the early diversification of microbial lineages and the relationships within and between the phyla of Bacteria and Archaea. In the next 5 years, the genomic foliage within the tree of life will continue to grow and the study of yet-uncultivated candidate phyla will firmly transition into the post-genomic era.

Keywords Single-cell genomics · Metagenomics · Candidate phyla · Genomic encyclopedia of bacteria and archaea (GEBA) · "Microbial dark matter"

Abbreviations

AMD	Acid mine drainage
FACS	Fluorescence-activated cell sorting
GEBA	Genomic encyclopedia of bacteria and archaea
MDA	Multiple displacement amplification
MDM	"Microbial dark matter"
SAG	Single amplified genome
SSU rRNA	Small subunit ribosomal RNA

This article is part of a special issue based on the 10th International Congress on Extremophiles held in Saint Petersburg, Russia, September 7-11, 2014.

Communicated by H. Santos.

B. P. Hedlund (✉) · J. A. Dodsworth · S. K. Murugapiran
School of Life Sciences, University of Nevada Las Vegas, 4004, 4505 Maryland Parkway, Las Vegas, NV 89154-4004, USA
e-mail: brian.hedlund@unlv.edu

C. Rinke · T. Woyke
DOE Joint Genome Institute, Walnut Creek, CA 94598, USA

Introduction

> "It has long been known that the standard methods of bacteriology–pure culture isolation and observation upon artificial media–often yield only an incomplete knowledge of a particular microbial flora"–Arthur Henrici, 1933.

Published online: 12 August 2014

The seemingly modern concept of the "uncultured microbial majority" (Rappé and Giovannoni 2003) has deep roots. Henrici recognized abundant and morphologically unusual morphotypes in freshwater lakes that had not been described in the literature (Henrici and Johnson 1935), some of which were isolated and described decades later and became founding members of the phyla Planctomycetes and Verrucomicrobia (de Bont et al. 1970; Staley 1973; Stackebrandt et al. 1984; Hedlund et al. 1997). But it was not until the development of a phylogenetic framework in microbiology (Woese and Fox 1977), and a system to place microorganisms from natural microbial communities into that framework (Olsen et al. 1986), that the nature and scale of the limitations of microbial cultivation began to clarify.

Extreme environments featured prominently in early studies employing rRNA surveys of natural environments, particularly terrestrial geothermal springs in Yellowstone National Park (Stahl et al. 1984; Reysenbach et al. 1994). Despite rapid progress on the cultivation of thermophiles at that time (Stetter et al. 1983; Zillig et al. 1983; Stetter et al. 1987; Stetter 2013), the limited diversity of cultured thermophiles compared with those in the natural environment was immediately obvious. This problem was exposed rather dramatically in a few reports from the Pace lab focusing on a single geothermal spring, Obsidian Pool in Yellowstone National Park, which led to the prediction of the Korarchaeota as a candidate phylum of Archaea (Barns et al. 1994, 1996) and eleven candidate phyla of Bacteria, which were designated OP1–OP11 (Hugenholtz et al. 1998a, b). The rRNA approach has continued to reveal yet-uncultivated microbial lineages in both extreme and non-extreme environments and it has recently been estimated that microbial isolates represent less than 20 % of the phylogenetic diversity of archaea and bacteria (Wu et al. 2009). This "uncultured microbial majority" includes forty to fifty yet-uncultivated candidate phyla of bacteria and a similar number of yet-uncultivated major lineages of archaea (McDonald et al. 2012; Baker and Dick 2013). In recognition of the scope and scale of this problem, several researchers have compared the "uncultured microbial majority" problem to the dark matter problem in astrophysics using terms such as "biological dark matter", "dark matter", and "microbial dark matter" (Marcy et al. 2007; Wu et al. 2009; Dodsworth et al. 2013; Rinke et al. 2013). We feel that this analogy has value to draw attention to the problem and because the major descriptors of dark matter in astrophysics such as "ubiquitous", "abundant", and "observable only by indirect techniques" (Zioutas et al. 2004) can be easily applied to the cultivation problem in microbiology.

If the past has been dominated by studies defining the nature of the problem, recent studies have made significant advancements through both microbial cultivation (Stott et al. 2008; Mori et al. 2009; Podosokorskaya et al. 2013; Dodsworth et al. 2014) and by accessing genomes of uncultivated organisms through metagenomics and single-cell genomics approaches, which are enabled by recent parallel advancements in DNA amplification, DNA sequencing, and computing. This publication briefly reviews current approaches to access and construct genomes of candidate phyla, some landmark studies on extremophiles, and the future of the study of candidate phyla using both cultivation-dependent and -independent approaches.

Accessing genomes of candidate phyla by metagenomic and single-cell genomics approaches

Because of their small genomes, unique physiology, and deeply branching phylogeny, extremophiles (particularly thermophiles) were among the first microbes to have their genomes sequenced (Fraser et al. 2000). Cultivation-independent genomics methods such as metagenomics (Handelsman 2004; Scholz et al. 2012), and more recently single-cell genomics (Stepanauskas 2012; Blainey 2013), have extended the 'genomics revolution' to include many candidate phyla found in extreme environments. The challenges and caveats of construction and interpretation of composite genomes from complex environmental samples or incomplete single-cell datasets are significant. However, careful application of these techniques and rigorous quality control and analysis can yield valuable insights into the phylogeny of these organisms and tremendously expand our understanding of their physiological potential beyond that inferred by small subunit (SSU) rRNA gene fragments.

Metagenomics

Metagenomics, sometimes also referred to as community genomics or environmental genomics, involves the analysis of bulk DNA from a given environment and can be driven functionally (i.e. by screening clone libraries for functions of interest) or bioinformatically after shotgun sequencing (Handelsman 2004; Scholz et al. 2012). One of the strengths of metagenomics is its versatility and relative simplicity in sample preparation; it can be applied to any community from which sufficient DNA can be extracted and can employ extraction procedures that are often not compatible with single-cell genomics approaches. In principle, this offers access to all community members provided enough sequencing is performed, including cells tightly adhered to solid surfaces or complex, aggregated microbial assemblages that are not readily amenable to single-cell analysis. However, because the resulting sequence data represent a mixture of genomic fragments from different organisms at varying abundances, the interpretation of the data in an organismal or phylogenetic context is not trivial. This problem is especially pronounced for candidate phyla for which reference sequences

beyond SSU rRNA gene fragments are typically not available. Nonetheless, "binning" or compartmentalization of metagenome data can be accomplished by clustering contigs by nucleotide composition (e.g. %G+C, codon usage, tetranucleotide frequency, homology) using a variety of techniques (Dick et al. 2009; Dröge and McHardy 2012; Mande et al. 2012; Scholz et al. 2012), with some approaches also considering read depth and a comparative framework (e.g. time series; Strous et al. 2012). Taxonomic assignment of the resulting bins can be done by homology searches and identification of phylogenetic anchor genes within a given bin. Careful application of such techniques to metagenomes has yielded the first complete or nearly complete genomes of several candidate phyla from extreme environments without any prior cultivation (Baker et al. 2010; Narasingarao et al. 2012; Nunoura et al. 2011, Takami et al. 2012). These genomic datasets necessarily represent a mosaic of genome fragments and are best analyzed in the context of representing the pangenome of closely related organisms (e.g. strains or species) from their host environments.

Single-cell genomics

In contrast to the bulk approach of metagenomics, single-cell genomics accesses genomes one cell at a time, allowing study of these organisms at the most fundamental biological unit (Stepanauskas 2012; Blainey and Quake 2014). Key aspects of this technique include the separation of individual cells from a complex mixture, followed by cell lysis and amplification of genomic DNA. Isolation of single cells can be done using a variety of techniques, including fluorescence-activated cell sorting (FACS) or optofluidics (optical trapping in a microfluidic device; Landry et al. 2013), which have various benefits depending on the application (Blainey 2013; Rinke et al. 2014). After lysis of individual cells, which is typically inefficient (Stepanauskas 2012), the femtogram levels of DNA in an individual cell ($\sim$ 1 fg per Mbp) are amplified using multiple displacement amplification (MDA) (Lasken 2012) or other genome amplification approaches (Zong et al. 2012) to nanogram- to microgram-levels required for sequencing. The resulting single amplified genomes (SAGs) are then typically screened by PCR amplification and sequencing of SSU rRNA genes to identify those belonging to candidate phyla or other taxa of interest. SAGs of interest are shotgun sequenced, assembled, and analyzed. The number of single-copy conserved markers (SCMs) in the assembly can give an estimate of how well a given SAG covers the target organism's genome (Baker et al. 2010; Rinke et al. 2013).

Many of the challenges in single-cell genomics are consequences of the very low amounts of starting DNA and the high degree of amplification required. Because of the sensitivity to contamination, careful preparation and handling of samples, reagents, and equipment is necessary (Stepanauskas 2012; Woyke et al. 2011). Post-sequencing quality control measures such as analysis of nucleotide frequency (Dodsworth et al. 2013), comparison to databases of likely contaminants (e.g. human, *Pseudomonas*, *Delftia*; Woyke et al. 2011), and screening for SCMs present at greater than one copy (Rinke et al. 2013) can all help identify potentially contaminating contigs for removal. MDA introduces chimeric artifacts and a severe bias in genomic coverage. As a result, individual SAG datasets are typically fragmented and do not represent complete genomes (Rinke et al. 2013). While amplification bias can result in problems with standard assembly algorithms, specific methods have been developed to deal specifically with SAG data (e.g. Nurk et al. 2013). Problems with bias and chimeras can be partially overcome by combining data from multiple, closely related single cells, e.g. those with average nucleotide identity (ANI) >95 % that likely represent members of a single species (Konstantinidis et al. 2006; Rinke et al. 2013). Jackknifed assembly procedures designed to remove chimeras present in one SAG, but not others, have been successfully employed to increase assembly continuity (Dodsworth et al. 2013; Marshall et al. 2012). The resulting composite assemblies can often represent nearly complete pangenomes for a given strain or species, enabling physiological interpretation based on the absence as well as presence of genes and pathways.

The combined application of single-cell genomics and metagenomics offers great opportunities for synergy because the advantages of these techniques are complementary (Fig. 1; Lasken 2012); metagenomics is not plagued by problems associated with MDA or separation of individual cells from a complex mixture, while single-cell genomics offers direct and unambiguous association of phylogeny and function (Walker 2014). For example, SAG data can greatly enhance the efficacy of metagenome binning procedures by providing key links between phylogeny, nucleotide frequency composition, and gene content. These links can be used both to better assign taxonomy to individual metagenome contigs (Rinke et al. 2013) and, in some cases, define nearly complete genomes of candidate phyla from metagenome datasets (Dodsworth et al. 2013). Conversely, metagenome reads or contigs can either be mapped to or be used as scaffolds for closely related SAG datasets, significantly improving continuity of SAG assemblies (Blainey 2013; Dodsworth et al. 2013).

Landmark studies accessing genomes of yet-uncultivated extremophile phyla

Extreme environments have featured prominently in metagenomics and single-cell genomics studies focusing on

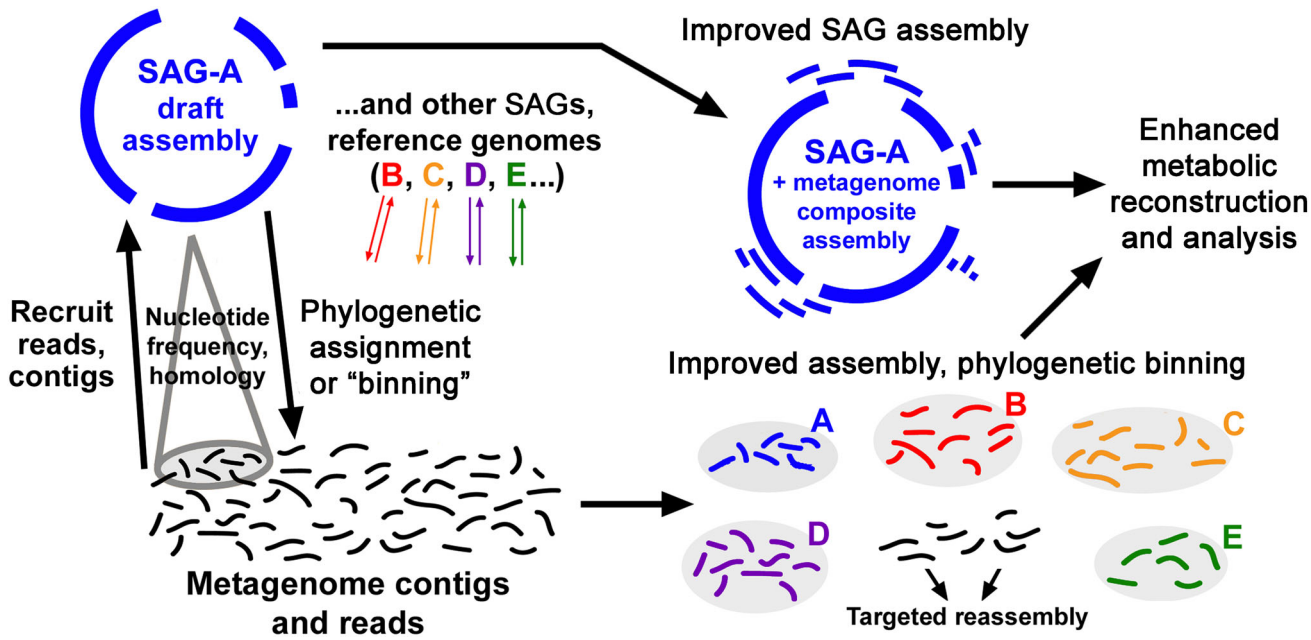

Fig. 1 Synergistic analysis of single amplified genomes (SAGs) and metagenomes improves assembly, analysis, and metabolic reconstruction. Relationships between SAG assemblies and metagenome sequences based on nucleotide frequency and homology (*left*) allow unambiguous assignment of metagenomic bins lacking SSU rRNA genes (*bottom right*), enhance SAG scaffolding (*top right*), and enable enhanced contaminant filtering of both datasets

novel lineages due to a variety of factors, including (1) the low to intermediate complexity of microbial communities in extreme environments, (2) the low abundance of bacterial phyla that dominate most non-extreme environments, (3) the well-known difficulties cultivating many extremophilic microorganisms, and (4) the intrinsic interest in exploring life in extreme environments. This section reviews some landmark studies applying metagenomics and/or single-cell genomics approaches to access genomes from novel lineages of bacteria and archaea (Figs. 2, 3).

Acidophiles

Tremendous progress has been made by the Banfield group in the development and application of approaches to study natural microbial communities, including metagenomics efforts focusing on biofilms in Richmond Mine, California. Richmond Mine hosts active acid mine drainage (AMD) due to the microbially catalyzed dissolution of pyrite and ranges in pH from ~0.5 to 1.5 and temperature from ~30 to 59 °C and has millimolar to molar concentrations of heavy metals (Druschel et al. 2004). Although initial Sanger-based metagenomics efforts focused on dominant community members (Tyson et al. 2004; Baker et al. 2006), subsequent efforts focused on novel, low-abundance archaeal lineages named archaeal richmond mine acidophilic nanoorganisms (ARMAN). Initially, only a few small biofilm metagenomic contigs containing novel SSU rRNA gene sequences were recovered (Baker et al. 2006), although deeper Sanger sequencing and binning based on nucleotide word frequency allowed construction of a near-complete composite genome of the ARMAN-2 lineage, which was named '*Candidatus* Micrarchaeum acidiphilum' (Baker et al. 2010). A discovery that the ARMAN groups dominated small size fractions of filtered biofilm suspensions (<0.45 µm; Baker et al. 2006) enabled genomic exploration of two distantly related lineages, ARMAN-4 and ARMAN-5, through sequencing of DNA from small cell fractions amplified by MDA (Baker et al. 2010). Nearly complete ARMAN-4 and ARMAN-5 genotypes were named '*Candidatus* Parvarchaeum acidophilum' and '*Candidatus* Parvarchaeum acidiphilus', respectively. All three composite genomes are very small (~1 Mbp) and have abnormally high coding density, typical of obligate symbionts, and transmission electron tomography images demonstrated interactions between a minority of ARMAN cells with Thermoplasmatales morphotypes (Baker et al. 2010); however, the nature of the presumed symbiosis is not clear. The '*Ca.* Micrarchaeum acidiphilum' genome is predicted to encode genes enabling beta-oxidation of organic acids and the '*Ca.* Parvarchaeum' genomes encode genes for glycolysis. All three ARMAN genomes encode complete or near-complete TCA cycles and are predicted to be capable of aerobic respiration, which is supported by the

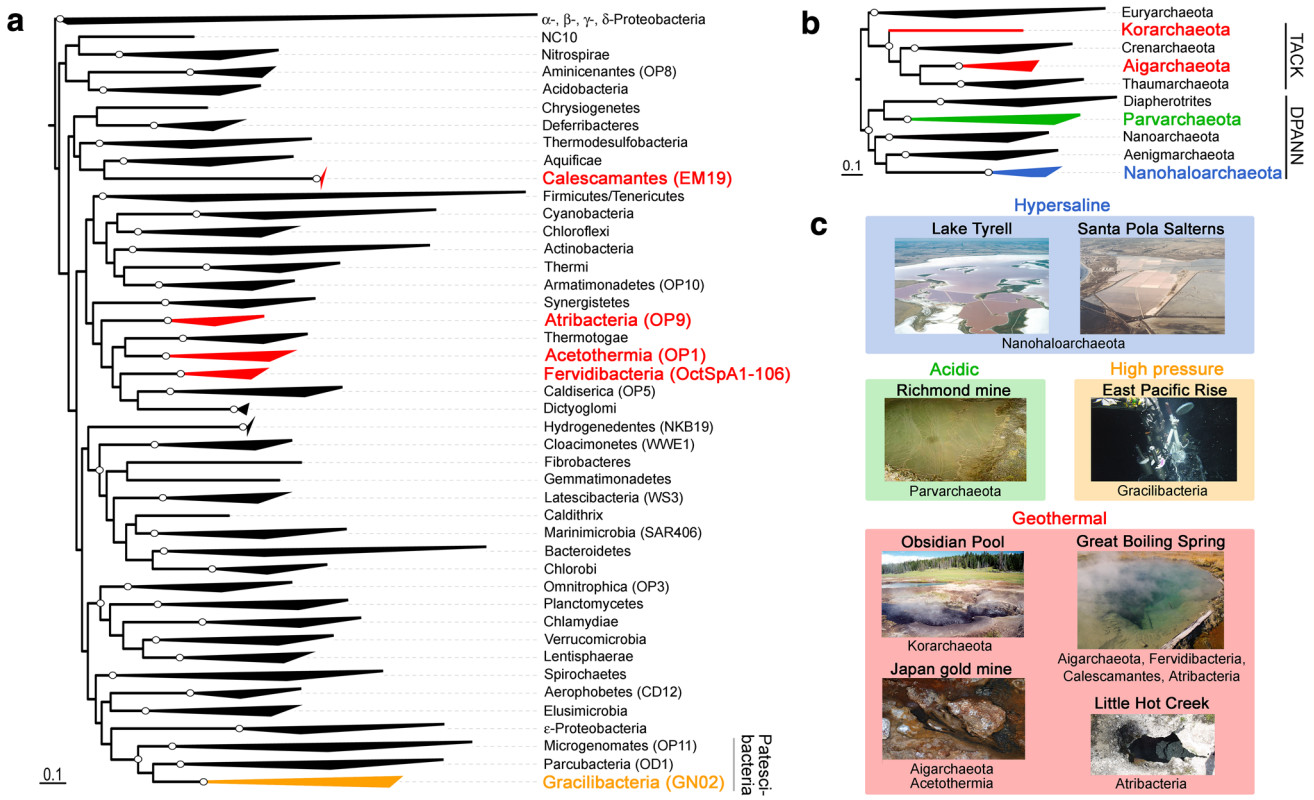

Fig. 2 Maximum-likelihood consensus trees of Bacteria (**a**) and Archaea (**b**) highlighting yet-uncultivated candidate phyla with complete or near-complete composite genomes determined by metagenomics or single-cell genomics and habitats they were derived from (**c**). Trees are based on concatenated alignments of conserved single-copy markers (SCMs), modified from Rinke et al. 2013 (*white circles*, >90 % consensus). Colors highlight extremophiles and their habitats: *green* acidophiles/acid mine drainage; *blue* halophiles/hypersaline lake; *red* thermophiles/terrestrial geothermal springs; *yellow* piezophiles/piezotolerant/hydrothermal vents. TACK, DPANN, and Patescibacteria are proposed superphyla (Guy and Ettema 2011; Rinke et al. 2013). Photo credits: Brian Hedlund (Obsidian Pool); Jeremy Dodsworth (Great Boiling Spring, Little Hot Creek); Ken Takai and Takuro Nunora (Japanese Gold Mine); Stefan Sievert, chief scientist, WHOI; Funding agency: NSF; ©Woods Hole Oceanographic Institution ("Crab Spa", East Pacific Rise)

recovery of abundant ARMAN respiratory proteins in natural biofilms (Baker et al. 2010). The ARMAN phylotypes have recently been ascribed to candidate phylum Parvarchaeota, which groups with other archaea with reduced genomes and small cell size in the 'DPANN superphylum' (Rinke et al. 2013).

Halophiles

Size fractionation was also used to enrich two novel archaeal phylotypes from surface waters of hypersaline Lake Tyrell, Australia (27–29 % salinity), which were prominent in Sanger and pyrosequenced metagenomes (Narasingarao et al. 2012). Following iterative phylogenetic binning, near-complete composite genomes were constructed for two related phylotypes, which were named '*Candidatus* Nanosalina sp.' and '*Candidatus* Nanosalinarum sp.'. In a separate study, a distantly related genome, named '*Candidatus* Haloredivivus sp,' was recovered through combined assembly of a metagenome bin from a 19 % salinity sample from the Santa Pola salterns near Alicante, Spain, and a SAG from a FACS-sorted sample from a 37 % salinity site in the same system (Ghai et al. 2011). All three genomes shared features that are unusual among known halophilic archaea, including small size (~1.2–1.3 Mbp) and very low G+C content (42–43.5 %). These genomes all encode a rhodopsin and genes suggesting a photoheterotrophic lifestyle, similar to other halophilic archaea, with complete glycolytic pathways predicted in all three genomes. The '*Ca.* Nanosalina sp.' and '*Ca.* Nanosalinarum sp.' genomes encode both oxidative and reductive pentose phosphate pathways enabled by a glucose-6-phosphate dehydrogenase distantly related to that of the abundant and ubiquitous halophilic bacterium *Salinibacter*. All three genomes also have unique amino acid content characterized by an abundance of acidic amino acids and a paucity of bulky aromatic amino acids, strongly suggesting a 'salt in' strategy consistent with life in hypersaline habitats. These three phylotypes were recently proposed to represent the candidate phylum Nanohaloarchaeota, also within the 'DPANN superphylum' (Rinke et al. 2013).

Candidate phylum	Candidate organisms	Composite SAGs	Metagnome bins	Autotrophy	Phototrophy	Aerobic resp.	Anaerobic resp.	Fermentation only	Genome <1.3 Mb	References
Parvarchaeota (ARMAN)	'Parvarchaeum', 'Micrarchaeum'					✓			✓	Baker et al. 2010
Nanohaloarchaeota	'Haloredivivus', 'Nanosalina', 'Nanosalinarum'				✓	✓				Ghai et al. 2011; Narasingarao et al. 2012
Korarchaeota	'Korarchaeum cryptofilum'							✓		Elkins et al. 2008
Aigarchaeota (pSL4, HWCG I)	'Caldiarchaeum subterraneum', >4 other spp.			✓		✓	✓			Nunoura et al. 2011 Rinke et al. 2013
Acetothermia (OP1)	'Acetothermum autotrophicum'			✓		✓	✓			Takami et al. 2012
Calescamantes (EM19)	'Calescibacterium nevadense'					✓	✓			Rinke et al. 2013
Fervidibacteria (OctSpA1-106)	'Fervidibacter sacchari'					✓	✓			Rinke et al. 2013
Atribacteria (OP9)	'Caldatribacterium californiense', 'C. saccharolyticum'							✓		Dodsworth et al. 2013
Gracilibacteria (GN02)	'Altimarinus pacificus'							✓		Rinke et al. 2013

Fig. 3 Summary of genomic coverage and predicted features of extremophilic "microbial dark matter" (MDM) based on metagenome and/or single-cell genome approaches. Reduced genomes (<1.3 Mbp) are typically found in obligate symbionts (Baker et al. 2006)

Thermophiles

Following up on studies predicting the phylum Korarchaeota (Barns et al. 1996), the dominant Korarchaeota phylotype in Obsidian Pool was established in 85 °C mixed culture chemostats in the Stetter laboratory (Burggraf et al. 1997). After years of attempts to obtain axenic cultures, a chemical/physical purification technique was developed to enrich Korarchaeota by treating samples with 0.2 % SDS and collecting cells in the 0.45-µm filtrate (Elkins et al. 2008). The highly purified preparation was Sanger sequenced, resulting in a single 1.59 Mbp contig. The organism, dubbed '*Candidatus* Korarchaeum cryptofilum', was predicted to couple peptide fermentation to hydrogen production and possibly encodes a mechanism to couple carbon monoxide oxidation to ATP synthesis through a [NiFe] carbon monoxide dehydrogenase. The genome indicated an inability to synthesize purines, CoA, and other coenzymes, which suggests a dependency on other members of the natural community.

Additional progress on novel thermophiles was made by Takai's group on thermophilic biofilms (~70 °C) from a subsurface gold mine in Japan. Initially, a single fosmid clone (41.2 kbp) with a novel SSU rRNA gene sequence from a group named Hot Water Crenarchaeotal Group I (HWCG I) was Sanger sequenced (Nunoura et al. 2005). Subsequently, a single genomic contig was assembled following Sanger and 454 pyrosequencing of fosmid clones and assembly and gap-filling by PCR (Nunoura et al. 2011). The organism was taxonomically assigned "*Candidatus* Caldiarchaeum subterraneum" in the candidate phylum Aigarchaeota and may couple hydrogen or carbon monoxide oxidation to aerobic or anaerobic respiration using nitrate or nitrite as electron acceptors. "*Ca.* Caldiarchaeum subterraneum" may be autotrophic via the dicarboxylate/4-hydroxybutyrate pathway, but lacks a canonical 4-hydroxybutyryl-CoA dehydratase. Phylogenetic, phylogenomic, and comparative genomic studies have consistently revealed a deep relationship between Thaumarchaeaota, Aigarchaeota, Crenarchaeota, and Korarchaeota in the 'TACK superphylum' (Guy and Ettema 2011; Rinke et al. 2013), yet there is some uncertainty about whether Aigarchaeota is an independent phylum or a deep branch within the Thaumarchaeota (Guy and Ettema 2011; Spang et al. 2013). This question will be resolved with deeper genomic coverage of these groups.

The same metagenomic library was used to assemble four large contigs representing a single phylotype belonging to candidate bacterial phylum OP1, named '*Candidatus* Acetothermus autotrophicum' (Takami et al. 2012). '*Ca.* Acetothermus autotrophicum' encodes a nearly complete acetyl-CoA pathway for carbon fixation and acetogenesis and a branched, partial TCA cycle proposed to feed anabolic pathways. Acetogenesis is proposed to be coupled to generation of an ion-motive force through a membrane-associated ferredoxin:NAD^+-oxidoreductase complex

(Rnf). The authors proposed the name Acetothermia for the OP1 candidate phylum and calculated a maximum growth temperature of 84.7 °C based on SSU rRNA G+C content.

Most recently, cultivation-independent genomic exploration has focused on novel phylotypes in Great Boiling Spring, Nevada (Costa et al. 2009; Cole et al. 2013). A single-cell genomics effort in Great Boiling Spring, including >30 cells from three sediment sites (78–85 °C), was included in the Genomic Encyclopedia of Bacteria and Archaea-Microbial Dark Matter (GEBA-MDM) project led by the Joint Genome Institute. The study resulted in 14 new Aigarchaeota SAGs and SAGs representing candidate phyla OctSpA1-106 (5 SAGs) and EM19 (10 SAGs) (Rinke et al. 2013). The Aigarchaeota represented five different species-level groups based on average nucleotide identity and each is distinct from "*Ca.* Caldiarchaeum subterraneum". Genomic data suggest considerable metabolic diversity within the Aigarchaeota, including possible mechanisms for hydrogenotrophy, carbon monoxide oxidation, aerobic respiration, nitrogen oxide respirations, dissimilatory sulfate reduction, and aerobic catabolism of aromatic compounds. The five Oct-SpA1-106 SAGs represented two species-level groups and the ten EM19 SAGs represented one species. The major phylotypes for these groups were named '*Candidatus* Fervidibacter sacchari' in candidate phylum Fervidibacteria (Oct-SpA1-106) and '*Candidatus* Calescibacterium nevadense' in candidate phylum Calescamantes (EM19). Both groups may be capable of aerobic respiration of organic compounds, consistent with the ability to enrich Fervidibacteria on lignocellulose (Peacock et al. 2013), but further genomic analysis is necessary for more detailed metabolic predictions.

Great Boiling Spring, along with a microbial community in Little Hot Creek, California (~79 °C; Vick et al. 2010), was also one of two sites in focus for genomic exploration of candidate phylum OP9 (Dodsworth et al. 2013). A survey of the major morphotypes in Little Hot Creek using an optofluidic approach, followed by SSU rRNA gene PCR screening, revealed that most rod-shaped cells ~0.5 μm in diameter belong to a single phylotype of OP9. Although this morphotype was rare (~0.5 % of cells), morphology-based sorting resulted in 21 OP9 SAGs from Little Hot Creek. Subsequently, the Little Hot Creek OP9 SAGs were used to define a metagenomic bin corresponding to a distinct, but closely related, OP9 phylotype that was enriched by in situ incubation of corn stover at ~77 °C in Great Boiling Spring (Peacock et al. 2013). Reciprocal homology searches enhanced assessment of contamination in both datasets, and the metagenomic contig significantly enhanced SAG scaffolding, ultimately enabling construction of two nearly complete OP9 SAGs, named '*Candidatus* Caldiatribacterium saccharolyticum' and '*Candidatus* Caldiatribacterium californiense' in candidate phylum Atribacteria. Both Atribacteria genotypes are predicted to be obligate fermenters and capable of cellulose or hemicellulose depolymerization through secretion of an extracellular endo-1,4-β-glucanase and Emden-Meyerhof fermentation of sugars with production of ethanol, acetate, and hydrogen. Although a plurality of Atribacteria genes have highest sequence similarity to Firmicutes, phylogenomic analyses do not support an affiliation with this phylum or others within the "Terrabacteria" superphylum (Rinke et al. 2013), and the Atribacteria are predicted to synthesize a lipopolysaccharide-containing outer membrane (Dodsworth et al. 2013). In addition to geothermal systems, the Atribacteria also inhabit moderate- to low-temperature biomes, particularly environments that are anaerobic and organic-rich (Gittel et al. 2009; Riviére et al. Rivière et al. 2009). Substantial genomic coverage of several SAGs from a moderately thermophilic (45–50 °C) and anaerobic terephthalate-degrading bioreactor and the hypolimnion of an anoxic fjord in British Columbia (Sakinaw Lake) have recently been described (Rinke et al. 2013). Deeper analysis of these SAGs and other Atribacteria datasets promise a more comprehensive understanding of the phylum as a whole.

Piezophiles/Piezotolerant

The GEBA-MDM project also generated SAGs from a diffuse-flow venting system from the "Crab Spa" hydrothermal field on the East Pacific Rise (EPR; Sievert and Vetriani 2012). Although the sample temperature was ~25 °C (S. Sievert, Pers. Comm.) it contained a mixture of hydrothermal fluid and pelagic water and, therefore, could host psychrophiles, mesophiloes, or thermophiles. The vent field is located at ~2,500 m depth and is under ~25 MPa of hydrostatic pressure, likely mandating specific molecular adaptations for piezophily or piezotolerance (Oger and Jebbar 2010). Five SAGs of interest belonged to candidate phyla GN02, OP11, and OD1 (1–2 SAGs per group), which were subsequently ascribed to the phyla Gracilibacteria, Microgenomates, and Parcubacteria, respectively, in the proposed superphylum Patescibacteria (Rinke et al. 2013). Similar to previous reports of Microgenomates and Parcubacteria composite genomes from an uncontained aquifer (Wrighton et al. 2012; Kantor et al. 2013; Wrighton et al. 2014), the SAGs from EPR belonging to these two groups are estimated to be reduced in size, suggesting a possible symbiotic lifestyle (<1.1 Mbp). Although the genomic coverage of these EPR SAGs was relatively low, no strong genomic evidence of respiratory capacity exists, which is in agreement with genomic and proteomic data suggesting related organisms are fermenters (Wrighton et al. 2012; Kantor et al. 2013; Wrighton et al.

2014). Interestingly, the Gracilibacteria SAGs from the EPR, one of which was named "*Candidatus* Altimarinus pacificus", have recoded the opal stop codon (UGA) for glycine, which may be a genomic adaptation to cope with their extremely low DNA G+C content (<24 %).

The future of extremophile MDM biology

Turning the crank

Although substantial progress has been made, the time has never been better to apply existing metagenomics and single-cell genomics approaches to access genomes of novel microorganisms. Many major lineages of both bacteria and archaea have no sequenced representatives (Baker and Dick 2013) and many that have sequenced genomes have low genomic coverage (Rinke et al. 2013). Ever-growing metagenomic datasets, continued advancements in metagenomic binning, and rapidly falling DNA sequencing and computing costs will support this effort. JGI recently initiated GEBA-MDM phase 2, which seeks to ramp up efforts to survey genomic diversity of the 'uncultured microbial majority', by continuing to explore sites rich in under-sampled candidate microbial phyla by FACS-enabled single-cell genomics, including extreme environments such as terrestrial geothermal springs in the U.S. Great Basin (Dodsworth and Hedlund 2010), British Columbia, Canada (Grasby et al. 2013), and Yunnan Province, China (Hou et al. 2013), deep sea hydrothermal sediments from Guaymas basin (Biddle et al. 2012), and hypersaline mats from Guerro Negro (Harris et al. 2013). We anticipate that this effort and others will be driving forces for the continued improvement of the genomic coverage of both bacteria and archaea at the phylum level.

Taking MDM biology into the post-genome era

As the genomic gaps in the tree of life continue to fill, the focus will shift toward testing sequence-based predictions. Toward this end, metatranscriptomics and metaproteomics approaches have been applied to natural ecosystems to confirm the existence of predicted transcripts and proteins in natural samples, particularly in extreme environments with low complexity, such as the Richmond Mine AMD site (Baker et al. 2010; Ram et al. 2005). However, these approaches cannot assess the functions of the gene products or the organisms directly. A multitude of isotope labeling approaches can be used to address this problem. Assimilatory metabolism of specific taxa can be addressed by combining fluorescence in situ hybridization (FISH) with microautoradiography (MAR-FISH; Wagner et al. 2006), nano-scale secondary ion mass spectrometry (FISH-nanoSIMS; Behrens et al. 2008), or Raman spectroscopy (Raman-FISH; Neufeld and Murrell 2007) following labeling with either radioactive or stable isotopes. Larger scale efforts are able to assess the functions of many taxa simultaneously, such as the use of nanoSIMS to survey nucleic acids hybridized to phylogenetic microarrays following stable isotope labeling (Chip-SIP; Mayali et al. 2012, 2013) or by cell sorting based on distinctive Raman spectra and subsequently identifying cells or sequencing genomes. The latter approach can also be used without isotope labeling to identify natural Raman signatures (e.g. based on distinctive lipids or cell inclusions). Finally, heterologous expression of gene products of interest remains a promising approach to examine specific genes of interest from uncultivated taxa (Lloyd et al. 2013) and advancements in synthetic biology offer promise to recode whole operons for expression in model organisms (Temme et al. 2012).

Rapid and valuable advancements in cultivation-independent approaches notwithstanding, we believe that genome-enabled cultivation is one of the most valuable outcomes of genomic exploration of uncultivated lineages. In addition to traditional cultivation approaches, advancements in microbial cultivation that allow for metabolic interactions between species will continue to be important (Nichols et al. 2010). Ultimately, the combination of cultivation-independent approaches and axenic culture, both enabled by genomic exploration, promise an exciting future for the yet-uncultivated microbial lineages in nature.

Acknowledgments This work was supported by NASA Exobiology grant EXO-NNX11AR78G; U.S. National Science Foundation grant OISE 0968421; U.S. Department of Energy (DOE) grant DE-EE-0000716; and the Joint Genome Institute (CSP-182), supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231. B. P. H. acknowledges generous support from Greg Fullmer through the UNLV Foundation.

References

Baker BJ, Dick GJ (2013) Omic approaches in microbial ecology: charting the unknown. Microbe 8:353–360

Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, Banfield JF (2006) Lineages of acidophilic archaea revealed by community genomic analysis. Science 314:1933–1935

Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, Verberkmoes NC, Hettich RL, Banfield JF (2010) Enigmatic, ultrasmall, uncultivated Archaea. Proc Natl Acad Sci 107:8806–8811

Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc Natl Acad Sci 91:1609–1613

Barns SM, Delwiche CF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci 93:9188–9193

Behrens S, Loesekann T, Pett-Ridge J, Weber PK, Ng JW-O, Stevenson BS, Hutcheon ID, Relman DA, Spormann AM (2008) Linking phylogeny with metabolic activity of single microbial

cells using FISH-NanoSIMS. Appl Environ Microbiol 74:3143–3150

Biddle JF, Cardman Z, Mendlovitz H, Albert DB, Lloyd KG, Boetius A, Teske A (2012) Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. ISME J 6:1018–1031

Blainey PC (2013) The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37:407–427

Blainey PC, Quake SR (2014) Dissecting genomic diversity, one cell at a time. Nat Methods 11:19–21

Burggraf S, Heyder P, Eis N (1997) A pivotal Archaea group. Nature 385:780

Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, Dong H, Wu G, Hedlund BP (2013) Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. ISME J 7:718–729

Costa KC, Navarro JB, Shock EL, Zhang CL, Soukup D, Hedlund BP (2009) Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. Extremophiles 13:447–459

de Bont JA, Staley JT, Pankratz HS (1970) Isolation and description of a non-motile, fusiform, stalked bacterium, a representative of a new genus. Antonie Van Leeuwenhoek 36:397–407

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF (2009) Community-wide analysis of microbial genome sequence signatures. Genome Biol 10:R85

Dodsworth JA, Hedlund BP (2010) Microbiology and geochemistry of Smith Creek and Grass Valley hot springs: emerging evidence for wide distribution of novel thermophilic lineages in the US Great Basin. J. Earth Sci 21:315–318

Dodsworth JA, Blainey PC, Murugapiran SK, Swingley WD, Ross CA, Tringe SG, Chain PSG, Raymond J, Quake SR, Hedlund BP (2013) Single-cell and metagenomic analyses indicate a fermentative, saccharolytic lifestyle for members of the OP9 lineage. Nature Commun 4:1854

Dodsworth JA, Gevorkian J, Despujos F, Cole JK, Murugapiran SK, Ming H, Li WJ, Zhang G, Dohnalkova A, Hedlund BP (2014) *Thermoflexus hugenholtzii* gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the Chloroflexi, Thermoflexia classis nov., and description of Thermoflexaceae fam. nov. and Thermoflexales ord. nov. Int J Syst Evol Microbiol. doi:10.1099/ijs.0.055855-0

Dröge J, McHardy AC, 66 (2012) Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. Brief Bioinform 13(66):646–655

Druschel GK, Baker BJ, Gihring TM, Banfield JF (2004) Acid mine drainage biogeochemistry at Iron Mountain, California. Geochem Trans 5:13–32

Elkins JG, Kunin V, Anderson I, Barry K, Goltsman E, Lapidus A, Hedlund BP, Hugenholtz P, Kyrpides N, Graham D, Keller M, Wanner G, Richardson P, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of archaea. Proc Natl Acad Sci 105:8102–8107

Fraser CM, Eisen JA, Salzberg SL (2000) Microbial genome sequencing. Nature 406:799–803

Ghai R, Pašić L, Fernández AB, Martin-Cuadrado AB, Mizuno CM, McMahon KD, Papke RT, Stepanauskas R, Rodriguez-Brito B, Rohwer F, Sánchez-Porro C, Ventosa A, Rodríguez-Valera F (2011) New abundant microbial groups in aquatic hypersaline environments. Sci Rep 1:135

Gittel A, Sørensen KB, Skovhus TL, Ingvorsen K, Schramm A (2009) Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. Appl Environ Microbiol 75:7086–7096

Grasby SE, Richards BC, Sharp CE, Brady AL, Jones GM, Dunfield PF (2013) The Paint Pots, Kootenay National Park, Canada—a natural acid spring analogue for Mars. Can J Earth Sci 50:94–108

Guy L, Ettema TJ (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol 19:580–587

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Harris JK, Caporaso JG, Walker JJ, Spear JR, Gold NJ, Robertson CE, Hugenholtz P, Goodrich J, McDonald D, Knights D, Marshall P, Tufo H, Knight R, Pace NR (2013) Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. ISME J 7:50–60

Hedlund BP, Gosink JJ, Staley JT (1997) Verrucomicrobia div. nov., a new division of the bacteria containing three new species of *Prosthecobacter*. Antonie Van Leeuwenhoek 72:29–38

Henrici AT (1933) Studies of freshwater bacteria. I. A direct microscopic technique. J Bacteriol 25:277–286

Henrici AT, Johnson DE (1935) Studies of freshwater bacteria. II. Stalked bacteria, a new order of Schizomycetes. J Bacteriol 30:61–92

Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, Huang Q, Huang L, Wu G, Zhi X, Li W, Dodsworth JA, Hedlund BP, Zhang C, Hartnett HE, Dijkstra P, Hungate BA (2013) A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. PLoS One 8:e53350

Hugenholtz P, Goebel BM, Pace NR (1998a) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774

Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998b) Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180:366–376

Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF (2013) Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio 4:e00708–e00713

Konstantinidis KT, Ramette A, Tiedje JM (2006) The bacterial species definition in the genomic era. Phil Trans R Soc 361:1929–1940

Landry ZC, Giovanonni SJ, Quake SR, Blainey PC (2013) Optofluidic cell selection from complex microbial communities for single-genome analysis. Methods Enzymol 531:61–90

Lasken RS (2012) Genomic sequencing of uncultured microorganisms from single cells. Nat Rev Microbiol 10:631–640

Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S, Schramm A, Jørgensen BB (2013) Predominant archaea in marine sediments degrade detrital proteins. Nature 496:215–218

Mande SS, Mohammed MH, Ghosh TS (2012) Classification of metagenomic sequences: methods and challenges. Brief Bioinform 13:669–681

Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR (2007) Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci 104:11889–11894

Marshall IPG, Blainey PC, Spormann AM, Quake SR (2012) A single-cell genome for *Thiovulum* sp. Appl Environ Microbiol 78:8555–8563

Mayali X, Weber PK, Brodie EL, Mabery S, Hoeprich PD, Pett-Ridge J (2012) High-throughput isotopic analysis of RNA microarrays to quantify microbial resource use. ISME J 6:1210–1221

Mayali X, Weber PK, Pett-Ridge J (2013) Taxon-specific C/N relative use efficiency for amino acids in an estuarine community. FEMS Microbiol Ecol 83:402–412

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618

Mori K, Yamaguchi K, Sakiyama Y, Urabe T, Suzuki K (2009) *Caldisericum exile* gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. nov. and Caldisericia classis nov. Int J Syst Evol Microbiol 59:2894–2898

Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE (2012) De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J 6:81–93

Neufeld JD, Murrell JC (2007) Witnessing the last supper of uncultivated microbial cells with Raman-FISH. ISME J 1:269–270

Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS (2010) Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. Appl Environ Microbiol 76:2445–2450

Nunoura T, Hirayama H, Takami H, Oida H, Nishi S, Shimamura S, Suzuki Y, Inagaki F, Takai K, Nealson KH (2005) Genetic and functional properties of uncultivated thermophilic crenarchaeotes from a subsurface gold mine as revealed by analysis of genome fragments. Environ Microbiol 7:1967–1984

Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, Takai K, Takami H (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res 39:3204–3223

Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714–737

Oger PM, Jebbar M (2010) The many ways of coping with pressure. Res Microbiol 161:799–809

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA (1986) Microbial ecology and evolution: a ribosomal RNA approach. Annu Rev Microbiol 40:337–365

Peacock JP, Cole JK, Murugapiran SK, Dodsworth JA, Fisher JC, Moser DP, Hedlund BP (2013) Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after in situ lignocellulose enrichment. PLoS One 8:e59927

Podosokorskaya OA, Kadnikov VV, Gavrilov SN, Mardanov AV, Merkel AY, Karnachuk OV, Ravin NV, Bonch-Osmolovskaya EA, Kublanov IV (2013) Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. Environ Microbiol 15:1759–1771

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC 2nd, Shah M, Hettich RL, Banfield JF (2005) Community proteomics of a natural microbial biofilm. Science 308:1915–1920

Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394

Reysenbach AL, Wickham GS, Pace NR (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Appl Environ Microbiol 60:2113–2119

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437

Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, Dmitrieff E, Malmstrom R, Stepanauskas R, Woyke T (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. Nat Protoc. doi:10.1038/nprot.2014.067

Rivière D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach J, Li T, Camacho P, Sghir A (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. ISME J 3:700–714

Scholz MB, Lo CC, Chain PS (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol 23:9–15

Sievert SM, Vetriani C (2012) Chemoautotrophy at deep-sea vents: past, present, and future. Oceanography 25:218–233

Spang A, Martijn J, Saw JH, Lind AE, Guy L, Ettema TJ (2013) Close encounters of the third domain: the emerging genomic view of archaeal diversity and evolution. Archaea 2013:202358

Stackebrandt E, Ludwig W, Schubert W, Klink F, Schlesner H, Roggentin T, Hirsch P (1984) Molecular genetic evidence for early evolutionary origin of budding peptidoglycan-less eubacteria. Nature 307:735–737

Stahl DA, Lane DJ, Olsen GJ, Pace NR (1984) Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. Science 224:409–411

Staley JT (1973) Budding bacteria of the *Pasteuria-Blastobacter* group. Can J Microbiol 19:609–614

Stepanauskas R (2012) Single cell genomics: an individual look at microbes. Curr Opin Microbiol 15:613–620

Stetter KO (2013) A brief history of the discovery of hyperthermophilic life. Biochem Soc Trans 41:416–420

Stetter KO, König H, Stackebrandt E (1983) *Pyrodictium* gen. nov., a new genus of submarine disc-shaped sulphur reducing Archaebacteria growing optimally at 105°C. Syst Appl Microbiol 4:535–551

Stetter KO, Lauerer G, Thomm M, Neuner A (1987) Isolation of extremely thermophilic sulfate reducers: evidence for a novel branch of archaebacteria. Science 236:822–824

Stott MB, Crowe MA, Mountain BW, Smirnova AV, Hou S, Alam M, Dunfield PF (2008) Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. Environ Microbiol 10:2030–2041

Strous M, Kraft B, Bisdorf R, Tegetmeyer HE (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. Front Microbiol 3:410

Takami H, Noguchi H, Takaki Y, Uchiyama I, Toyoda A, Nishi S, Chee GJ, Arai W, Nunoura T, Itoh T, Hattori M, Takai K (2012) A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. PLoS One 7:e30559

Temme K, Zhao D, Voigt CA (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. Proc Natl Acad Sci 109:7085–7090

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43

Vick TJ, Dodsworth JA, Costa KC, Shock EL, Hedlund BP (2010) Microbiology and geochemistry of Little Hot Creek, a hot spring environment in the Long Valley Caldera. Geobiology 8:140–154

Wagner M, Nielsen PH, Loy A, Nielsen JL, Daims H (2006) Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays. Curr Opin Biotechnol 17:1–9

Walker A (2014) Adding genomic 'foliage' to the tree of life. Nat Rev Microbiol 12:78

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci 74:5088–5090

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S, Malmstrom R, Stepanauskas R, Cheng J-F (2011) Decontamination of MDA reagents for single cell whole genome amplification. PLoS One 6:e26161

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF (2012) Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661–1665

Wrighton KC, Castelle CJ, Wilkins MJ, Hug LA, Sharon I, Thomas BC, Handley KM, Mullin SW, Nicora CD, Singh A, Lipton MS, Long PE, Williams KH, Banfield JF (2014) Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. ISME J. doi:10.1038/ismej.2013.249

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462:1056–1060

Zillig W, Gierl A, Schreiber G, Wunderl S, Janekovic D, Stetter KO, Klenk HP (1983) The archaebacterium *Thermofilum pendens* represents, a novel genus of the thermophilic, anaerobic sulfur respiring Thermoproteales. Syst Appl Microbiol 4:79–87

Zioutas K, Hoffmann DH, Dennerl K, Papaevangelou T (2004) What is dark matter made of? Science 306:1485–1488

Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622–1626