Journal of Advanced Research (2015) 6, 269–282

Cairo University

Journal of Advanced Research

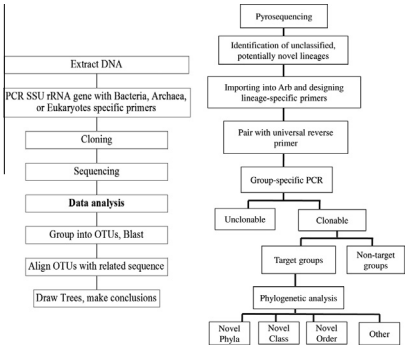
REVIEW

Assessing the global phylum level diversity within the bacterial domain: A review

Noha H. Youssef, M.B. Couger, Alexandra L. McCully, Andrés Eduardo Guerrero Criado, Mostafa S. Elshahed *

Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 August 2014
Received in revised form 6 October 2014
Accepted 23 October 2014
Available online 4 November 2014

ABSTRACT

Microbial ecology is the study of microbes in the natural environment and their interactions with each other. Investigating the nature of microorganisms residing within a specific habitat is an extremely important component of microbial ecology. Such microbial diversity surveys aim to determine the identity, physiological preferences, metabolic capabilities, and genomic features of microbial taxa within a specific ecosystem. A comprehensive review of various aspects of microbial diversity (phylogenetic, functional, and genomic diversities) in the microbial (bacterial, archaeal, and microeukaryotic) world is clearly a daunting task that could

* Corresponding author. Tel.: +1 (405) 744 1192; fax: +1 (405) 744 1112.
E-mail address: Mostafa@okstate.edu (M.S. Elshahed).
Peer review under responsibility of Cairo University.

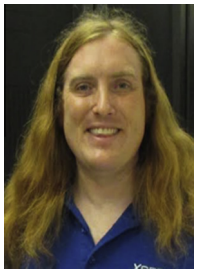
Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jare.2014.10.005

Keywords:
Phylogenetic diversity
Candidate phyla
16S rRNA gene
Culture-independant diversité surveys

not be aptly summarized in a single review. Here, we focus on one aspect of diversity (phylogenetic diversity) in one microbial domain (the Bacteria). We restrict our analysis to the highest taxonomic rank (phylum) and attempt to investigate the extent of global phylum level diversity within the Bacteria. We present a brief historical perspective on the subject and highlight how the adaptation of molecular biological and phylogenetic approaches has greatly expanded our view of global bacterial diversity. We also summarize recent progress toward the discovery of novel bacterial phyla, present evidences that the scope of phylum level diversity in nature has hardly been exhausted, and propose novel approaches that could greatly facilitate the discovery process of novel bacterial phyla within various ecosystems.

Noha Youssef is an Assistant Professor in the Department of Microbiology and Molecular Genetics at Oklahoma State University, Stillwater, OK, USA. She graduated with a Bachelor degree in Pharmacy from Ain Shams University, Cairo, Egypt. She obtained her PhD from the department of Botany and Microbiology at the University of Oklahoma, Norma, OK, USA. Her PHD research was in the area of petroleum microbiology and microbially enhanced oil recovery. Her post-graduate research was conducted in Dr. Elshahed laboratory, with a research focus on molecular microbial ecology and environmental genomics. Currently, research in her laboratory is focused on single cell genomics and the ecology and evolution of anaerobic fungi.



Matthew Brian Couger is a doctoral student in the Microbiology and Molecular Genetics program at Oklahoma State University. He currently is serving as the Extreme Science and Engineering (XSEDE) Bioinformatics Domain Champion, a position for consulting on large-scale bioinformatics projects. His current and ongoing interest is quantitative molecular biology, molecular evolution, synthetic biology, high performance computing, and bioinformatics.

Alexandra L. McCully graduated summa cum laude in 2013 from Oklahoma State University with a Bachelor of Science degree in Microbiology and Molecular Genetics. She trained as an undergraduate researcher to investigate salt adaptation strategies in halophilic microorganisms and analyze phylogenetic assignments within the domain Bacteria. Currently, she is working towards her PhD at Indiana University studying microbial metabolic interactions and biofuel production.

Andres Eduardo Guerrero Criado has studied in Venezuela, the United States, and Spain. Currently as an undergraduate of Microbiology, Cell and Molecular Biology and Genetics/Biochemistry at Oklahoma State University he is involved in research in bioinformatics, phylogeny, microbial ecology and protein structure elucidation under Dr. Noha Youssef. In addition, at Washington University in St. Louis Medical School he collaborated with Dr. Jean Schaffer and her Diabetic Cardiovascular Disease Center studying the role of RNASET2 in oxidative stress. The duality of these programs offers the perfect combination of research and practice to pursue a degree in Medical Research.

Mostafa Elshahed graduated from Cairo University faculty of Pharmacy in 1993. He obtained his Ph.D. from the University of Oklahoma in 2001. His PhD studies focused on elucidating the pathways for benzoate degradation under anaerobic conditions. His post-doctoral studies, also at the university of Oklahoma focused on the microbial ecology of terrestrial sulfidic springs. Dr. Elshahed joined Oklahoma State University as an Assistant Professor in 2007 and was promoted to an associate professor in 2011. Currently Dr. Elshahed has multiple research interests including Environmental genomics, petroleum microbiology, and the biology and metabolism of the anaerobic fungi.

Historical background

Microbial ecology is the scientific discipline where scientists examine microbes in their environment, their impact and adaptation to their habitat and their interactions with each other. Microbial diversity surveys, which aim to identify the types of microorganisms within a specific habitat are an integral part of microbial ecology. The discovery of "animalcules" (single celled microscopic microorganisms), by Antony van Leeuwenhoek in various samples e.g. rain drops, water samples from wells and lakes, oral and stool samples from humans

is, in essence, microbial diversity surveys [1]. Following Leeuwenhoek's discoveries, a relative hiatus in microbiology research ensued in the 18th and the earlier parts of the 19th century. The revival of microbiology research during the mid 19th–early 20th century was characterized by a marked shift in research philosophy. Holistic observation of microorganisms in their natural habitats was replaced with a reductionist research philosophy, with emphasis on the identification of etiological agents of microbially mediated phenomena such as fermentation and pathogenesis. Research during this era, deservedly referred to as the "golden age of microbiology" has lead to multiple seminal advances e.g. development of solid media for culturing bacteria, germ theory of disease, staining techniques, and vaccination procedures [2]. However, such spectacular advances have shifted the research focus of microbiologists from an ecosystem-oriented, holistic philosophy to a reductionist, pure-culture centric focus.

The Russian/Ukrainian scientist Sergei Winogradsky, whose biography is almost as interesting as his research accomplishments, advocated a research approach that emphasizes the study of microorganisms in their natural habitats in mixed cultures or in isolates recently recovered from the ecosystem of interest. Winogradsky correctly reasoned that microorganisms in nature survive in conditions that are a far cry from the controlled, nutrient-rich conditions at which pure cultures are maintained in the laboratory. He reasoned that the behavior of a specific microorganism in its natural habitat is markedly different from its behavior in pure culture due to the differences in nutrient and resource availability between both conditions, as well as to the constant interactions with various microbial taxa coexisting within the same habitat [1]. His work on environmental samples, especially soil, has clearly led to a better appreciation of the metabolic and functional diversity of microorganisms in their natural habitats.

Winogradsky's research, and subsequent efforts by eminent microbiologists (Beijerinck, van Neal, Kluyver, and Hungate) has defined the goals of microbial ecology. These could be simplified for the non-specialist as the "who" (identity of microorganisms), "what" (their metabolic capabilities), "where" (their spatiotemporal distribution within an ecosystem as well as in a global scale), and "why" (functions in a specific ecosystem and role in geochemical cycling). The "who" is, obviously, the most basic question in microbial ecology (add references). After 340 years postanimalcules discovery and almost a century since the revival of microbial ecology by Winogradsky, one would imagine that this seemingly straightforward question has satisfactory been answered, and that the science of microbial discovery and description of new taxa would be as dead as the science of discovering new organs in the human body. This could not be any further from the truth. A global census of all microbial species on earth is now recognized as a truly impossible task [3]. Even with a single sample from a highly diverse ecosystem (e.g. soil), such census still represents a daunting challenge [4,5,6].

In this review, we examine the scope of bacterial diversity within the domain Bacteria. We limit our assessment of phylogenetic diversity to the highest taxonomic rank (phylum) and attempt to address seemingly straightforward questions: How many bacterial phyla exist in nature? Have all such phyla already been described? And what approaches could be implemented to more effectively document novel, yet undescribed phylum level diversity within the Bacteria?

From the great plate count anomaly to the uncultured bacterial majority

The great plate count anomaly and the "missing" cells

It has been observed, as early as 1932, that within freshwater samples, only an extremely small fraction of microscopically observed microbial cells is recoverable as pure cultures in microbial growth media [7]. This observation (initially seen in freshwater) has since been validated in a wide array of environmental samples (e.g. marine, soils, and freshwater habitats, see [8] and references within). Typically, the absolute majority (99–99.9%) of cells within an environmental sample are not recoverable in pure culture using plating or most probable number (MPN) enumeration procedure. Specific measures have been shown to slightly improve the proportion of cultured cells within select environmental samples. These include the utilization of multiple media targeting various metabolic capabilities and physiological preferences, longer incubation time [9], novel isolation contraptions [10,11], use of dilute media to mimic resource scarcity in nature and/or media mimicking natural settings [12], and the implementation of more sensitive growth detection methods [11,13]. Nevertheless, even with improved methodologies, the majority of cells within highly complex habitats remain uncultured. The term "The great plate count anomaly" has been aptly coined to describe this phenomenon in 1988 [8].

A logical inquiry stemming from the recognition of this phenomenon is the identity of microorganisms escaping enrichment and isolation procedures. Do these microorganisms represent novel, hitherto unknown bacterial taxa, or do they represent close relatives of bacterial taxa available in pure culture that possess attenuated growth capabilities, multiple unidentified auxotrophies, and/or yet-unclear physiological and growth requirements? The presence of unique cellular morphologies in environmental samples that have never been recovered in pure cultures has often hinted at the putative novelty of at least a fraction of these uncultured cells [14]. However, prior to the advent of molecular taxonomic approaches and their wide utilization in diversity surveys this question was mostly philosophical in nature [15].

Use of molecular phylogeny in culture-independent diversity surveys

The late American microbiologist Carl Woese pioneered the use of 16S rRNA gene as a phylogenetic marker to provide an evolutionary-based taxonomic outline for living organisms. Using comparative 16S rRNA gene sequence analysis, he proposed a three kingdom classification scheme [16], where all living creatures are grouped into three domains (Bacteria, Archaea, and Eukaryotes). His further investigation of cultured taxa within the bacterial domain has produced the first high rank taxonomic outline for Bacteria, with all known bacterial taxa grouped into 12 different phyla or divisions (Fig. 1) [17].

Building on these efforts, the American microbiologist Norman Pace has pioneered the use of 16S rRNA gene-based sequencing and analysis procedures as a tool for direct identification of microbial populations in environmental samples. This approach was originally dubbed "phylotyping" but is

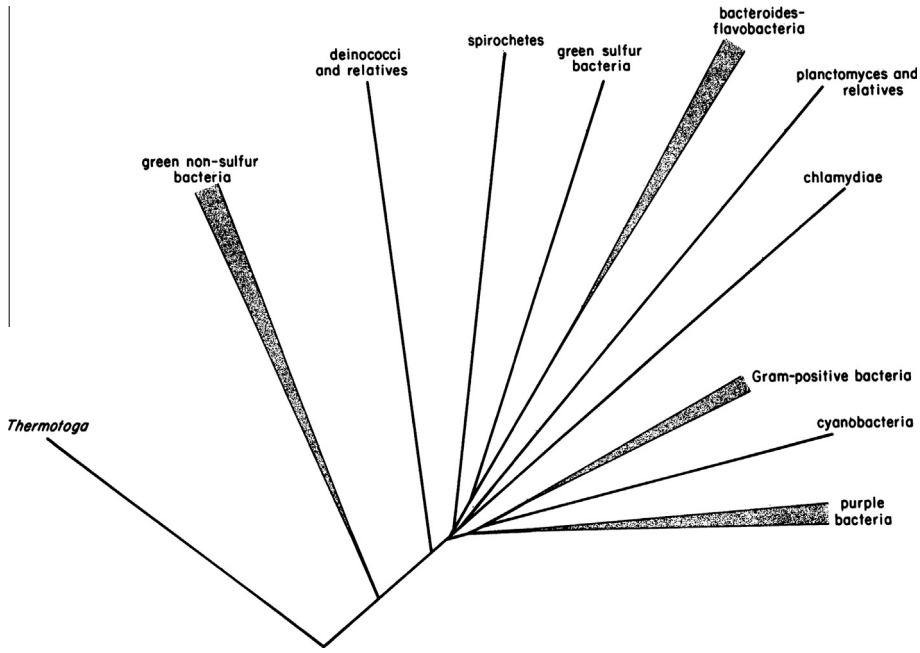


Fig. 1 Phylogenetic tree depicting the twelve "original" bacterial phyla proposed by Carl Woese in his seminal review on bacterial evolution. Adapted from Ref. [9]. These phyla are Thermotogae, Chloroflexi (Green non-sulfur Bacteria), Deinococcus, Spirochaetes, Chlorobia (Green sulfur bacteria), Bacteroidetes, Planctomycetes, Chlamydia, Cyanobacteria, Gram-positive Bacteria (comprising the high GC Actinobacteria, and the low GC Firmicutes), Proteobacteria (Purple bacteria).

more commonly referred to now as "16S rRNA gene-based culture-independent diversity survey", or simply "16S rRNA analysis" (Fig. 2) [18]. It involves direct isolation of bulk DNA from an environmental sample followed by PCR amplification of a fragment of the 16S rRNA gene using primers targeting conserved regions within the molecule. The amplicon, representing a mix of 16S rRNA genes originating from different cells within the environmental sample of interest is then cloned and sequenced (or directly sequenced when using newer high throughput sequencing procedures, see below) [15,19]. The obtained sequences are analyzed and their phylogenetic affiliation is assessed using various phylogenetic and bioinformatics procedures. This approach has the monumental advantage of being culture-independent i.e. capable of identifying microorganisms within a specific environmental samples regardless of their amenability or refractiveness to isolation [18]. As such, it is well suited to address questions posed above regarding the identity and taxonomy of uncultured microorganisms routinely escaping detection in enrichment and isolation-based procedures.

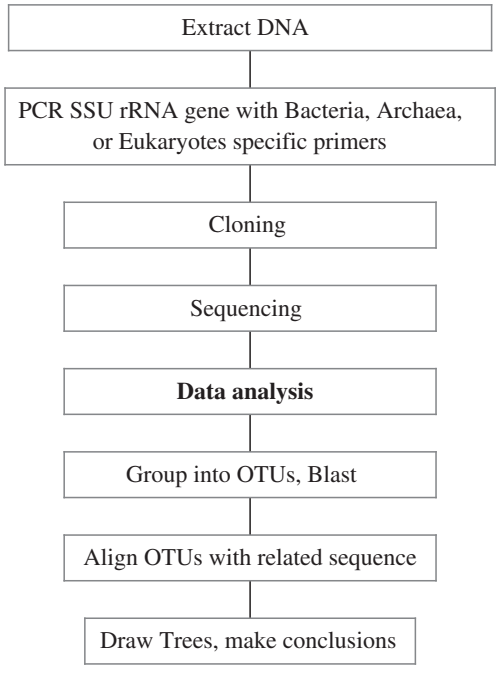


Fig. 2 Flowchart depicting the "16S rRNA analysis" protocol. The protocol starts by DNA extraction, followed by amplifying the small subunit rRNA gene using universal or domain-specific primers. PCR products are then cloned and sequenced. Obtained small subunit rRNA gene sequences are then analyzed, binned into operational taxonomic units (OTUs), and used for phylogenetic inferences.

The uncultured bacterial majority revealed

The 16S rRNA gene-based approach has been readily adopted in the past three decades by the absolute majority of the scientific community, and extensively utilized to study the microbial diversity in ecosystems ranging from large global habitats, e.g. oceans [20–40], and soil [41–60], to hardly accessible extreme environments such as deep sea hydrothermal vents [61–76], Antarctic lakes [32,62,77–82], and Antarctic soils

Table 1 Bacteria phyla names according to Greengenes [91] and SILVA [33] databases (August 2014).[a]

Greengenes	SILVA
AC1	
Acidobacteria	*Acidobacteria*
Actinobacteria	**Actinobacteria**
AD3	
AncK6	
	aquifer1
	aquifer2
Aquificae	*Aquificae*
Armatimonadetes	*Armatimonadetes*
Bacteroidetes	**Bacteroidetes**
	BD1-5
BHI80-139	BHI80-139
BRC1	BRC1
Caldiserica	*Caldiserica*
Caldithrix	
CD12	
Chlamydiae	**Chlamydiae**
Chlorobi	**Chlorobi**
Chloroflexi	**Chloroflexi**
Chrysiogenetes	*Chrysiogenetes*
	CKC4
Cyanobacteria	**Cyanobacteria**
Deferribacteres	*Deferribacteres*
Thermi	**Deinococcus-Thermus**
Dictyoglomi	*Dictyoglomi*
Elusimicrobia	*Elusimicrobia*
EM3	
EM19	
FBP	
FCPU426	
Fibrobacteres	*Fibrobacteres*
Firmicutes	**Firmicutes**
Fusobacteria	*Fusobacteria*
	GAL08
GAL15	
Gemmatimonadetes	*Gemmatimonadetes*
GN01	
GN02	
GN04	
GOUTA4	GOUTA4
H-178	
Hyd24-12	Hyd24-12
Kazan-3B-28	
	KB1
KSB3	
LCP-89	
	JL-ETNP-Z39
	JS1
LD1	LD1-PA38
Lentisphaerae	*Lentisphaerae*
MAT-CR-M4-B07	
MVP-21	
MVS-104	
NC10	
Nitrospirae	*Nitrospirae*
NKB19	
NPL-UPA2	NPL-UPA2
OC31	OC31
OctSpA1-106	
OD1	OD1
OP1	
OP3	OP3

(*continued on next page*)

Table 1 (*continued*)

Greengenes	SILVA
OP8	OP8
OP9	OP9
OP11	OP11
PAUC34f	
Planctomycetes	**Planctomycetes**
Poribacteria	
Proteobacteria	**Proteobacteria**
	RsaHF231
	S2R-29
SAR406	
SBR1093	
	SBYG-2791
SC4	
	SHA-109
	SM2F11
Spirochaetes	**Spirochaetae**
SR1	SR1
Synergistetes	*Synergistetes*
TA06	TA06
Tenericutes	*Tenericutes*
	Thermodesulfobacteria
Thermotogae	**Thermotogae**
TM6	TM6
TM7	TM7
TPD-58	
Verrucomicrobia	*Verrucomicrobia*
VHS-B3-43	
	WCHB1-60
	WD272
WPS-2	
WS1	
WS2	
WS3	WS3
WS4	
WS5	
WS6	WS6
WWE1	
ZB3	

[a] Phyla shown in Boldface are those already known with cultured representatives prior to the advent of 16S rRNA gene diversity surveys. Phyla in italics are those with cultured representatives originally identified using 16S rRNA sequencing as uncultured bacterial phyla, with representative isolates subsequently obtained. The rest of the phyla currently have no cultured representatives.

[33,62,83–90]. Collectively, these studies have demonstrated that the scope of phylogenetic diversity is much broader than previously implied from culture-based studies. Multiple novel microbial lineages have been identified, many of which appear to be deeply branching within the bacterial tree and unaffiliated with any of the known bacterial phyla. The discovery of these lineages necessitated coining the term candidate phylum (or candidate division) to accommodate these bacterial phyla where only 16S rRNA sequences but no isolates are available. Indeed, examination of taxonomic outlines provided by curated 16S rRNA gene databases e.g. Greengenes [91] and SILVA [33] suggests that, currently, the majority of currently recognized bacterial phyla are candidate phyla (Table 1). Therefore, the application of 16S rRNA gene based diversity surveys has resulted in the discovery of multiple novel bacterial lineages at the highest taxonomic rank and have revolutionized

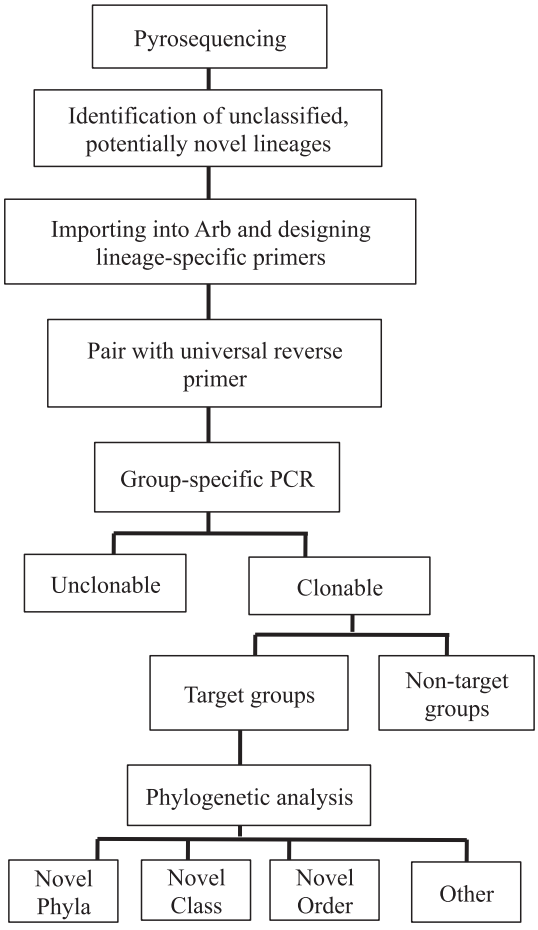


Fig. 3 Flowchart depicting a targeted approach developed for the identification of novel bacterial phyla within the rare biosphere. The approach combines the sequence read length and accuracy of the Sanger sequencing approach with the high throughput capability of next generation (Pyrosequencing or Illumina) sequencing approaches. Pyrosequencing or Illumina sequencing output are first used to identify potentially novel members within rare members of the community. The short sequences are then used to design custom primers. The newly designed primers are then used in conjunction with a forward, or reverse bacterial primer for amplification of near-complete 16S rRNA gene sequences. Obtained PCR products are cloned and Sanger-sequenced, and the sequences obtained are used for detailed phylogenetic inferences.

our understanding of the scope of phylum level diversity in nature. More importantly, such analysis clearly demonstrated that a fraction of microbial cells consistently missed in enumeration and isolation approaches clearly belong to novel, hitherto unrecognized bacterial lineages.

Global phylum level diversity in bacteria

These new discoveries of novel bacterial phyla and candidate phyla have added multiple new deep branches (phyla) to the bacterial trees of life, but are we done with this exercise? Has the phylum level diversity within the Bacteria been exhausted, or are there multiple, yet-undescribed novel bacterial phyla (or even domains) in nature? One would imagine that, after three decades of research, thousands of published 16S rRNA gene-based diversity surveys, 5.4 million Sanger-generated 16S rRNA gene sequences in GenBank and >1.7 billion sequences in high throughput sequencing archives e.g. SRA [92], CAMERA [93], and MG-RAST [94], and the discovery and documentation of tens of novel bacterial candidate phyla, that the global scope of diversity of bacteria on earth has been documented, at least at the highest taxonomic (phylum) level. However, based on our research experience in the last decade, the authors are now firm believers that the scope of global phylum level bacterial diversity is much greater than currently recognized in curated 16S rRNA gene databases such as Greengenes [91] and SILVA [33] (Table 1). Below, we present three different reasons why we believe that this is the case, as well as procedures that could putatively facilitate the discovery of these novel phyla.

Novel bacterial phyla as constituents of the rare biosphere

Within highly diverse microbial ecosystems, several distribution models can be used to fit the frequency data, e.g. ordinary Poisson distribution, gamma-mixed Poisson, inverse Gaussian-mixed Poisson, lognormal-mixed Poisson, Pareto-mixed Poisson, and mixture of 2 exponentials-mixed Poisson [58,95–99]. Regardless of the distribution pattern, the community structure in diverse habitats typically exhibits a taxon rank distribution curve with a long tail corresponding to bacterial species present in low abundance. This fraction constituting the majority of species is referred to as the "rare" biosphere [20]. The reason why these lineages are present and maintained at low abundances, as well as their global distribution patterns and putative ecological roles (or lack thereof), is an active area of interest to microbial ecologists and evolutionary microbiologists.

Access to the rare members of the community has been greatly augmented by the advent of high throughput sequencing technologies and their adaptation to amplicon-based 16S rRNA gene-based diversity surveys e.g. pyrosequencing [20], and Illumina sequences [100]. Such adaptation has allowed for the generation of hundreds of thousands (pyrosequencing) to millions (Illumina) of sequencing reads in a single run and hence provided unprecedented access to the rare biosphere. Collectively, these studies have documented the extremely high level of species richness within the rare biosphere. More interestingly, within such studies, a significant fraction of the obtained sequences (10–74% [101–105] are considered unclassified beyond a preset sequence similarity threshold, e.g., 80%, to the closest classifiable relative in databases.

However, it is important to note that, while pyrosequencing-, and Illumina-based studies are excellent tools for suggesting the occurrence of novel bacterial diversities within a sample, they are very poor in accurately documenting and describing such diversity. Accurate determination of the phylogenetic affiliation of such pyrosequencing-, and Illumina-generated sequences is unfeasible, mainly due to the short-read-length output of currently available high throughput technologies, and the error rate associated with them, which preclude the direct deposition of obtained short sequences into public databases e.g. GenBank. Hopes on the development of a high throughput, long-read sequencing approach have been high,

but the newer systems that offer that (e.g. PacBio SMRT) have a dreadfully high error rate (~14% indels for PacBio SMRT sequencing) that preclude their utilization for high throughput phylogenetic studies.

Therefore, Sanger-generated near full-length 16S rRNA gene sequences remain the only viable way for the accurate description and documentation of novel bacterial lineages. In spite of the fact that an extremely large number of Sanger-generated 16S rRNA gene sequences (> 5 M, as of August 2014) are currently available through the GenBank database, the absolute majority of these sequences have been obtained during the course of small-scale diversity surveys (e.g. < 200 sequences generated per study). Accordingly, these studies, and consequently the entire database have an extremely poor representation of the rare biosphere within the ecosystems studied.

Two strategies have been developed as a means to obtain near full-length 16S rRNA gene sequences from the rare biosphere. The first is a brute force approach in which a large number of clones are sequenced from a single sample, and the other depends on the development of a more targeted approach to specifically access putatively novel members within the rare biosphere. Due to cost issues, relatively few studies have utilized a brute force approach for this process. For example, [106] examined the bacterial diversity in grassland soil by analyzing 13,001 sequences from a single sample. This study demonstrated that rare members of the microbial community have, on average, more novelty (i.e. less sequence similarity to their closest relative in the database) compared to more abundant members of the samples. More importantly, the authors identified multiple novel lineages at various taxonomic levels, with the identification of 6 putative new phyla. Another more impressive more recent effort [107] focused on analyzing ~119,000 Sanger-generated sequences obtained from 10 equivalent sections pooled from 4 core samples of a 5 cm thick Guerrero Negro microbial mat, and resulted in the identification of 43 putatively novel phyla. Collectively, both studies, as well as other deep sequencing Sanger-based studies conducted on a smaller scale, e.g. [108,109] consistently demonstrate that novel bacterial phyla are still to be encountered in the rare biosphere.

A more targeted approach to zoom in on putatively novel members of the rare biosphere has been independently developed by three different research laboratories and used to target putatively novel and rare members of the microbial community in a sulfide and sulfur-rich spring in southwestern Oklahoma (Zodletone spring) [110], freshwater microbial communities [111], and marine sponges [103]. This approach (Fig. 3) is based on using sequences generated in high throughput sequencing surveys to identify sequences with low sequence similarity (e.g. < 80%) to closest relatives in GenBank

Table 2 Common 16S rRNA bacterial primers used for culture-independent analysis.[a]

Primer name	Primer sequence[b]
8F	AGAGTTTGATCCTGGCTCAG
27F	AGAGTTTGATCMTGGCTCAG
338R	GCCTTGCCAGCCCGCTCAG
338F	ACTCCTACGGGAGGCWGCAGC
518R	GTATTACCGCGGCTGCTGG
530F	ACGCTTGCACCCTCCGTATT
805R	GGATTAGATACCCTGGTAGTC
967F	CAACGCGAAGAACCTTACC
1238R	GTAGCRCGTGTGTMGCCC
1100F	YAACGAGCGCAACCC
1492R	CGGTTACCTTGTTACGACTT

[a] F indicates a forward primer and R indicates a reverse primer. Number in the primer name indicates the starting position of the primer sequence within the *E. coli* 16S rRNA gene sequence.
[b] Data from references [57,126].

Table 3 List of degenerate primers for 27F and 1492R designed as specified in the text. The sequences of the non-degenerate 27F and 1492R are given in the table heading.

AGAGUUUGAUCAUGGCUCAG	AAGUCGUAACAAGGUAACC
BGAGUUUGAUCAUGGCUCAG	**G**AGUCGUAACAAGGUAACC
A**H**AGUUUGAUCAUGGCU**D**AG	A**G**GUCGUAACAAGGUAACC
AG**B**GUUUGAUCAUGGC**V**CAG	AA**H**UCGUAACAAGGUAACC
AGA**H**UUUGAUCAUGH**D**UCAG	AAG**C**CGUAACAAGGUAACC
AGAG**G**UUGAUCAUG**H**CUCAG	AAGU**D**GUAACAAGGUAACC
AGAG**A**UUGAUCAUG**H**CUCAG	AAGUC**H**UAACAAGGUAACC
AGAG**C**UUGAUCAUG**H**CUCAG	AAGUCG**C**AACAAGGUAACC
AGAGU**G**UGAUCAUG**H**CUCAG	AAGUCGU**G**ACAAGGUAACC
AGAGU**A**UGAUCAUG**H**CUCAG	AAGUCGUA**B**CAAGGUAACC
AGAGU**C**UGAUCAUG**H**CUCAG	AAGUCGUAA**D**AAGGUAACC
AGAGUU**C**GAUCAUGGCUCAG	AAGUCGUAAC**G**AGGUAACC
AGAGUUU**A**AUCAUGGCUCAG	AAGUCGUAACA**G**GGUAACC
AGAGUUUG**G**UCAUGGCUCAG	AAGUCGUAACAAA**G**UAACC
AGAGUUUGA**V**CAUGGCUCAG	AAGUCGUAACAAGA**U**AACC
AGAGUUUGAU**D**AUGGCUCAG	AAGUCGUAACAAGG**V**AACC
AGAGUUUGAUC**B**UGGCUCAG	AAGUCGUAACAAGGU**B**ACC
AGAGUUUGAUCA**C**GGCUCAG	AAGUCGUAACAAGGUA**B**CC
AGAGUUUGAUCAU**U**GCUCAG	AAGUCGUAACAAGGUAA**D**C
AGAGUUUGAUCAUG**U**CUCAG	AAGUCGUAACAAGGUAAC**D**
AGAGUUUGAUCAUGGCUC**G**G	
AGAGUUUGAUCAUGGCUCA**H**	

database. Primers specific to these putatively novel sequences are then designed and used in conjunction with universal bacterial primers to obtain near full length 16S rRNA amplicons which could be cloned, sequenced using Sanger sequencing, and subjected to detailed phylogenetic analysis. Using this approach, five novel bacterial phyla were identified within the rare members of the microbial community in Zodletone spring in Southwestern Oklahoma [110]. Therefore, regardless of the approach utilized, it is clear that all dedicated efforts expended on identifying novelty within the rare biosphere in various ecosystems almost invariably yielded novel bacterial phyla. We hence conclude that a sustained and dedicated effort to investigate phylum level diversity in the rare biosphere in multiple complex habitats could hence have a profound effect on our understanding of the global scope of phylum level diversity within the domain Bacteria.

Novel bacterial phyla in the shadow biosphere

All 16S rRNA gene-based diversity surveys are initiated by amplification of 16S rRNA genes using primers that target conserved regions within the 16S rRNA molecule. A list of universal bacterial primers used in diversity surveys is shown in Table 2. It has often been argued that these "universal" primers could not theoretically amplify every single microbial strain within a single complex environmental sample, and that a fraction of microbial diversity is routinely missed in PCR-based diversity studies. However, the proportion of missed diversity, or the "shadow biosphere" as a fraction of the total number of cells is currently unclear. Indeed, 16S rRNA gene sequences within genomic fragments obtained via PCR-independent techniques, e.g. cloned in fosmids [112], have mismatches to the sequences of commonly used universal 16S rRNA primers [113]. Further, a detailed *in silico* analysis of 16S rRNA gene sequences identified in PCR-independent metagenomic survey in NCBI environmental survey repository also identified multiple 16S rRNA gene sequences that harbor mismatches to common universal bacterial 16S rRNA primers [114].

In addition, several studies provide empirical evidence that the shadow biosphere harbors a disproportionally large fraction of bacterial cells belonging to novel bacterial phyla. For example, the discovery of candidate divisions AD3, NC10, and mesophilic Thermotoga as integral constituents within soil ecosystems has long been hampered by the common mismatches exhibited in their 16S rRNA gene sequences to universal bacterial primers, resulting in their chronically common misrepresentation and outright absence in soil clone libraries [115]. More importantly, recent studies from the Banfield laboratory at UC-Berkley have constituted multiple genome assemblies from metagenomic datasets derived from a variety of habitats [113,116–120]. Many of these reconstituted genomes represent completely novel bacterial phyla that have never been observed before, in pure cultures, or in 16S PCR-based diversity surveys. All such novel biosphere-derived phyla exhibit multiple mismatches within their 16S rRNA gene sequences to various "universal" bacterial primers currently in use, and hence were always missed in diversity surveys. A similar situation has been encountered within the domain Archaea, where culture-independent single cell genomic analysis recovered genomes belonging to completely novel archaeal

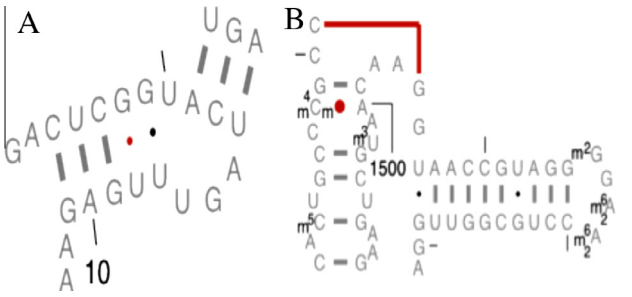


Fig. 4 Secondary structure of regions (A) 8–27, and (B) 1492–1510 of the 16S rRNA molecule. Canonical base pairing (shown as lines) is targeted for designing degenerate primers such that a change in one base is associated with a complementary change in the pairing position. Noncanonical base pairings (A-A, C-C, G-G, C-A, U-G, G-A, U-U), and wobble base pairing (G-U), often a consequence of canonical pairings, are theoretically less necessary for maintaining ribosomal integrity, and so are not targeted for primer design. The sequences of all possible degenerate 27F and 1492R primers are shown in Table 3.

phyla with 16S rRNA gene sequences exhibiting marked mismatches, and even indels (insertions and deletions), which render them recalcitrant to amplification using current PCR primers and protocols [121,122].

Utilization of PCR independent metagenomic approaches as a routine procedure for assessing diversity might be possible in the future, but currently, PCR-based approaches represent the most feasible way to assess diversity. Therefore, to assess diversity within the shadow biosphere using PCR-based approaches, newer strategies are needed. One approach to potentially limit or decrease the proportion of cells missed due to primer mismatches is to utilize miniprimers (10 bp primers) instead of the standard 18–20 bp primers currently in use, and to employ engineered S-Tbr DNA polymerase instead of Taq polymerase to allow such amplification procedure [114]. Theoretically, mismatches are less probable to occur in a shorter 10 bp primer when compared to a standard 18–20 bp primer. Isenbarger et al. [114] used this approach to examine bacterial diversity in soil, as well as a microbial mat sample from Cabo Rojo, PR using a shorter version of the standard 27F and 1505R primers (Table 2) [27F-10 (5′ TTCCGGTTGA) 1505R-10 (5 CCTTGTTACG)], and engineered S-Tbr DNA polymerase. The authors compared clone libraries observed using both approaches and clearly demonstrated that a higher proportion of putatively novel sequences were obtained with the miniprimer approach when compared to standard primer approach.

We further propose an additional approach based on designing multiple degenerate primers to account for mismatches to the universal 16S rRNA gene. Since base pairing is necessary to maintain 16S rRNA secondary structure, degenerate primers will be designed to theoretically maintain canonical base pairings in 16S rRNA secondary structure (Fig 3), i.e. any base change at one position will be compensated by a complementary base change at the pairing position (Fig. 3). Applications of such an exercise to two primers (27f, and 1492r) would generate a list of 21 degenerate forward, and 19 degenerate reverse primers (Table 3). Each of these degenerate primers can theoretically be paired with the universal forward or reverse primer and used for 16S rRNA sequence

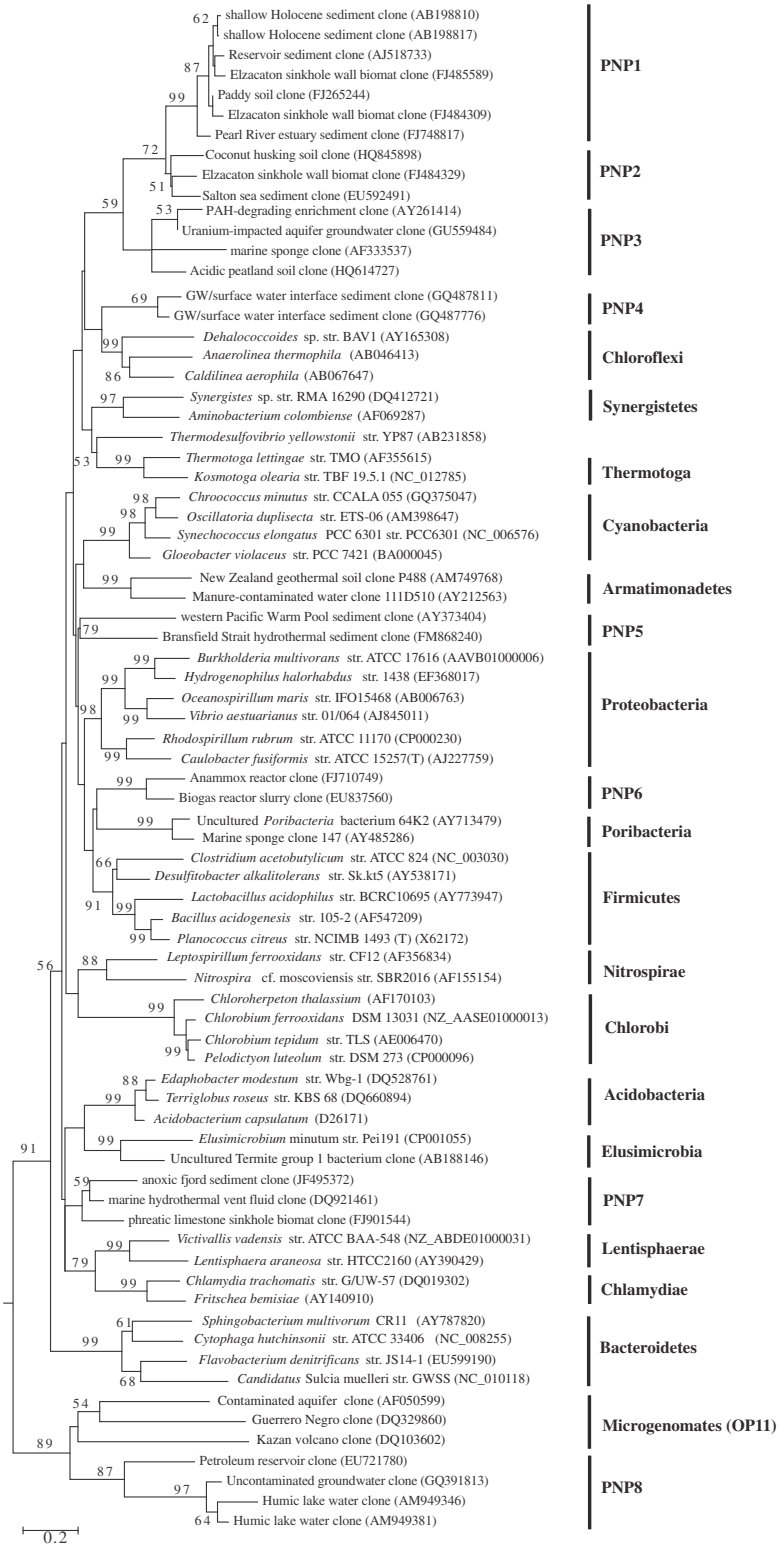


Fig. 5 Maximum likelihood dendogram based on the 16S rRNA gene sequences affiliated with representatives of the putatively novel phyla (PNP1-PNP8). Bootstrap values (in percentages) are based on 1000 replicates and are shown for branches with more than 50% bootstrap support. Sequences obtained from the ENA database (n = 3,178,046) were classified in MOTHUR using classify.seqs command with the Greengenes taxonomy outline and Wang method. Sequences that failed to classify into a known phylum with at least 50% bootstrap support (n = 664,621) were considered potentially novel and were subjected to extensive phylogenetic analysis using a combination of Mega [124], RaxML [123], and Arb [125]. Seventy-nine sequences formed 8 independent, deep-branching, reproducibly monophyletic, bootstrap-supported clusters, upon applying various tree-building algorithms as well as upon varying the composition and size of the data set used for phylogenetic analysis. Representatives of these 8 novel phyla are shown in the tree along with their source.

amplification in a multiplexed high throughput PCR approach to identify novel sequences. Such approach has been mulled before but has never been utilized to our knowledge to identify diversity (see Fig. 4).

Inadequate documentation of phylum level diversity within existing databases

In addition to the failure to detect novel bacterial phyla due to their rarity in environmental samples or to their possession of mismatches to most commonly used 16S rRNA gene primers, we argue that current inadequate curation of deposited 16S rRNA gene sequences is leading to failure in recognizing novel bacterial phyla for which 16S rRNA gene sequence has already been reported. All published studies of 16S rRNA gene surveys deposit sequences obtained in a public database, most commonly GenBank database (available at ftp://ftp.ncbi.nih.gov/blast/db/nt. and EMBL database). Many of the studies are focused on various ecological questions and do not conduct a detailed assessment of the phylogenetic affiliation of every obtained 16S rRNA gene sequence. Therefore, 16S rRNA gene sequences representing novel phyla could be deposited unnoticed to GenBank database. Curated 16S rRNA gene databases (e.g. Greengenes [91], and SILVA [33]) routinely upload recently deposited 16S rRNA gene sequences in GenBank and add such sequences to their taxonomic outlines. However, proposing novel bacterial phyla based on newly obtained sequences represent but one of the interests and responsibilities of database curators, and many novel 16S rRNA sequences that putatively represent novel bacterial phyla are simply refer to as "unclassified" in such databases.

We hypothesized that 16S rRNA sequences representing multiple novel bacterial phyla have already been obtained and deposited in public databases but has so far escaped detection and documentation due to reasons highlighted above. As a proof of principle, we queried one of such database depositories, the European Nucleotide Archive (ENA) [92], for novel 16S rRNA sequences. At the time of download (September, 2013), 3,178,046 16S rRNA gene sequences were obtained. The sequences were trimmed for length to remove all sequences shorter than 800 bp and were classified using Greengenes taxonomy and Wang method employed in Mothur. Most of the sequences (~80%) were classified into a known phylum or candidate division with >50% bootstrap support. The remaining 20% of sequences were subjected to an extensive phylogenetic analysis using maximum likelihood approaches (implemented in RaxML [123] and Mega [124]). As a result, 79 different sequences were judged to represent 8 novel bacterial phyla. These 79 sequences formed eight different independent, deep branching, reproducibly monophyletic, bootstrap-supported clusters, upon applying various tree-building algorithms as well as upon varying the composition and size of the data set used for phylogenetic analysis (Fig. 5). Sequences representing potentially novel classes and orders belonging to known phyla were also identified (data not shown). Therefore, such analysis, conducted sequences from the relatively smaller ENA database, clearly demonstrates that novel bacterial phyla are routinely detected in diversity surveys but often escapes documentation. Similar analysis using sequences in larger databases e.g. GenBank, as well as continuous evaluation of recently deposited sequences could clearly result in the identification of additional novel phyla.

Conclusions

We hope to convey that, in spite of the spectacular technological advances in DNA sequences, and intense research in the area of microbial diversity, that to-date, a complete census of the phylum level diversity within the domain bacteria has not yet been realized. A similar statement could be made regarding the domain Archaea and, to some extent, the microeukaryotes. Our review summarizes progress toward such goal, and outlines potential strategies and procedures that could facilitate the discovery process. It is interesting to note that many of such novel bacterial phyla appear to have a limited distribution and often represent a minor fraction of the microbial community within a specific habitat. The reason for their retention of such cells in highly diverse habitats, and their potential role within a specific ecosystem (or lack thereof) is an issue that is currently unclear. Access to the genome of such microorganisms through single cell genomics or metagenomics, or success in obtaining representative pure cultures would be required to address such questions.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Acknowledgments

Funding in our laboratory to support work on bacterial diversity and environmental genomics is supported by the National Science Foundation Microbial Observatories Program (Grant EF0801858).

References

[1] Bibel DJ. Microbial musings. Belmont, CA: Star Pub Co.; 2000.
[2] Blevins SM, Bronze MS. Robert koch and the 'golden age' of bacteriology. Int J Infect Dis 2010;14:744–51.
[3] Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown T, Brown CT, et al. The terabase metagenomics workshop and the vision of an earth microbiome project. Stand Genomic Sci 2010;3:243–8.
[4] Schloss PD, Handelsman J. Towards a census of bacteria in soil. PLoS Comp Biol 2006;2:e92.
[5] Baveye PC. To sequence or not to sequence the whole-soil metagenome? Nat Rev Microbiol 2009;7:756.
[6] Vogel TM, Simonet P, Jansson JK, Hirsch PR, Tiedje JM, Elsas JDv, et al. Terragenome: a consortium for the sequencing of a soil metagenome. Nat Rev Microbiol 2009;7:252.
[7] Razumov AS. The direct method of calculation of bacteria in water. Comparison with the koch method. Mikrobiologiya 1932;1:131–46.

[8] Staley JT, Konopka A. Measurement of in-situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Ann Rev Microbiol 1985;39:321–46.
[9] Joseph SJ, Hugenholtz P, Sangwan P, Osborne CA, Janssen PH. Laboratory cultivation of widespread and previously uncultured soil bacteria. Appl Environ Microbiol 2003;69:7210–5.
[10] Kaeberlein T, Lewis K, Epstein SS. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 2002;296:1127–9.
[11] Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, et al. Cultivating the uncultured. Proc Natl Acad Sci USA 2002;99:15681–6.
[12] Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol 2002;2002(68).
[13] Rappe M, Connon SA, Vergin KL, Giovannoni SJ. Cultivation of the ubiquitous sar11 marine bacterioplankton clade. Nature 2002;418:630–3.
[14] Walsby AE. A square bacterium. Nature 1980;283:69–71.
[15] Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16s ribosomal rna sequences for phylogenetic analyses. Proc Natl Acad Sci USA 1985;82:6955–9.
[16] Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 1977;74:5088–90.
[17] Woese CR. Bacterial evolution. Microbiol Rev 1987;51:221–71.
[18] Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16s ribosomal rna sequences for phylogenetic analyses. Proc Natl Acad Sci USA 1985;82:6955–9.
[19] Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 2006;103:12115–20.
[20] Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 2006;103(32):12115–20.
[21] DeLong EF. Archaea in coastal marine environments. Proc Natl Acad Sci USA 1992;89(12):5685–9.
[22] Field KG, Gordon D, Wright T, Rappe M, Urback E, Vergin K, et al. Diversity and depth-specific distribution of sar11 cluster rrna genes from marine planktonic bacteria. Appl Environ Microbiol 1997;63(1):63–70.
[23] Fuhrman JA, McCallum K, Davis AA. Phylogenetic diversity of subsurface marine microbial communities from the atlantic and pacific oceans. Appl Environ Microbiol 1993;59(5):1294–302.
[24] Giovannoni SJ, Rappe MS, Vergin KL, Adair NL. 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. Proc Natl Acad Sci USA 1996;93(15):7979–84.
[25] Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 1991;173(14):4371–8.
[26] Kirchman DL, Cottrell MT, Lovejoy C. The structure of bacterial communities in the western arctic ocean as revealed by pyrosequencing of 16S rRNA genes. Environ Microbiol 2010;12(5):1132–43.
[27] Malmstrom RR, Straza TR, Cottrell MT, Kirchman DL. Diversity, abundance, and biomass production of bacterial groups in the western arctic ocean. Aquat Microb Ecol 2007;47:45–55.
[28] Inagaki F, Nunoura T, Nakagawa S, Teske A, Lever M, Lauer A, et al. Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the pacific ocean margin. Proc Natl Acad Sci USA 2006;103(8):2815–20.
[29] Lauro FM, Chastain RA, Blankenship LE, Yayanos AA, Bartlett DH. The unique 16 rRNA genes of piezophiles reflect both phylogeny and adaptation. Appl Environ Microbiol 2007;73(3):838–45.
[30] Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML, et al. Abundance and diversity of microbial life in ocean crust. Nature 2008;453(7195):653–6.
[31] Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. Activity of abundant and rare bacteria in a coastal ocean. Proc Natl Acad Sci USA 2011;108(31):12776–81.
[32] Tian F, Yu Y, Chen B, Li H, Yao Y-F, Guo X-K. Bacterial, archaeal and eukaryotic diversity in arctic sediment as revealed by 16S rRNA and 18S rRNA gene clone libraries analysis. Polar Biol 2009;32(1):93–103.
[33] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The silva ribosomal rna gene database project: Improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590–596.
[34] Biers EJ, Sun S, Howard EC. Prokaryotic genomes and diversity in surface ocean waters: Interrogating the global ocean sampling metagenome. Appl Environ Microbiol 2009;75(7):2221–9.
[35] Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the arctic ocean. Proc Natl Acad Sci USA 2009;106(52):22427–32.
[36] Hongxiang X, Min W, Xiaogu W, Junyi Y, Chunsheng W. Bacterial diversity in deep-sea sediment from northeastern pacific ocean. Acta Ecol Sinica 2008;28(2):479–85.
[37] Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. Evidence for a persistent microbial seed bank throughout the global ocean. Proc Natl Acad Sci USA 2013;110(12):4651–5.
[38] Hunt DE, Lin Y, Church MJ, Karl DM, Tringe SG, Izzo LK, et al. Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl Environ Microbiol 2013;79(1):177–84.
[39] Whalan S, Webster NS. Sponge larval settlement cues: the role of microbial biofilms in a warming ocean. Sci Rep 2014:4.
[40] Mohit V, Archambault P, Toupoint N, Lovejoy C. Phylogenetic differences in attached and free-living bacterial communities in a temperate coastal lagoon during summer, revealed via high-throughput 16S rRNA gene sequencing. Appl Environ Microbiol 2014;80(7):2071–83.
[41] Kuffner M, Hai B, Rattei T, Melodelima C, Schloter M, Zechmeister-Boltenstern S, et al. Effects of season and experimental warming on the bacterial community in a temperate mountain forest soil assessed by 16S rRNA gene pyrosequencing. FEMS Microbiol Ecol 2012;82(3):551–62.
[42] Will C, Thürmer A, Wollherr A, Nacke H, Herold N, Schrumpf M, et al. Horizon-specific bacterial community composition of german grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol 2010;76(20):6751–9.
[43] Vasileiadis S, Puglisi E, Arena M, Cappa F, Cocconcelli PS, Trevisan M. Soil bacterial diversity screening using single 16S rRNA gene v regions coupled with multi-million read generating sequencing technologies. PLoS ONE 2012;7(8):e42671.
[44] Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K, et al. Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. Appl Environ Microbiol 2014;80(5):1777–86.
[45] Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, et al. Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci USA 2013;110(16):6548–53.

[46] Ferrenberg S, O'Neill SP, Knelman JE, Todd B, Duggan S, Bradley D, et al. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. ISME J 2013;7(6):1102–11.
[47] Wang J, Gu J-D. Dominance of candidatus scalindua species in anammox community revealed in soils with different duration of rice paddy cultivation in northeast china. Appl Microbiol Biotechnol 2013;97(4):1785–98.
[48] Desai C, Parikh RY, Vaishnav T, Shouche YS, Madamwar D. Tracking the influence of long-term chromium pollution on soil bacterial community structures by comparative analyses of 16S rRNA gene phylotypes. Res Microbiol 2009;160(1):1–9.
[49] Jechalke S, Focks A, Rosendahl I, Groeneweg J, Siemens J, Heuer H, et al. Structural and functional response of the soil bacterial community to application of manure from difloxacin-treated pigs. FEMS Microbiol Ecol 2014;87(1):78–88.
[50] Sprocati A, Alisi C, Tasso F, Fiore A, Marconi P, Langella F, et al. Bioprospecting at former mining sites across europe: microbial and functional diversity in soils. Environ Sci Poll Res 2014;21(11):6824–35.
[51] Rahman MM, Basaglia M, Vendramin E, Boz B, Fontana F, Gumiero B, et al. Bacterial diversity of a wooded riparian strip soil specifically designed for enhancing the denitrification process. Biol Fert Soils 2014;50(1):25–35.
[52] Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, et al. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 2007;1(4):283–90.
[53] Bintrim SB, Donohue TJ, Handelsman J, Roberts GP, Goodman RM. Molecular phylogeny of archaea from soil. Proc Natl Acad Sci USA 1997;94(1):277–82.
[54] Kuske CR, Barns SM, Busch JD. Diverse uncultivated bacterial groups from soils of the arid southwestern united states that are present in many geographic regions. Appl Environ Microbiol 1997;63(9):3614–21.
[55] Dunbar J, Barns SM, Ticknor LO, Kuske CR. Empirical and theoretical bacterial diversity in four arizona soils. Appl Environ Microbiol 2002;68(6):3035–45.
[56] Schloss PD, Handelsman J. Toward a census of bacteria in soil. PLoS Comput Biol 2006;2(7):e92.
[57] Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl Environ Microbiol 2009;75(16):5227–36.
[58] Youssef NH, Elshahed MS. Species richness in soil bacterial communities: a proposed approach to overcome sample size bias. J Microbiol Meth 2008;75(1):86–91.
[59] Youssef NH, Elshahed MS. Diversity rankings among bacterial lineages in soil. ISME J 2008;3(3):305–13.
[60] Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol 2006;72(3):1719–28.
[61] Kimura H, Higashide Y, Naganuma T. Endosymbiotic microflora of the vestimentiferan tubeworm (*lamellibrachia* sp.) from a bathyal cold seep. Mar Biotechnol (NY) 2003;5(6):593–603.
[62] Teske A, Hinrichs KU, Edgcomb V, de Vera Gomez A, Kysela D, Sylva SP, et al. Microbial diversity of hydrothermal sediments in the guaymas basin: evidence for anaerobic methanotrophic communities. Appl Environ Microbiol 2002;68(4):1994–2007.
[63] Amarouche-Yala S, Benouadah A, El Ouahab Bentabet A, López-García P. Morphological and phylogenetic diversity of thermophilic Cyanobacteria in algerian hot springs. Extremophiles 2014;18(6):1035–47.
[64] Anderson RE, Beltrán MT, Hallam SJ, Baross JA. Microbial community structure across fluid gradients in the juan de fuca ridge hydrothermal system. FEMS Microbiol Ecol 2013;83(2):324–39.
[65] Brazelton WJ, Ludwig KA, Sogin ML, Andreishcheva EN, Kelley DS, Shen C-C, et al. Archaea and bacteria with surprising microdiversity show shifts in dominance over 1000-year time scales in hydrothermal chimneys. Proc Natl Acad Sci USA 2010;107(4):1612–7.
[66] Byrne N, Strous M, Crepeau V, Kartal B, Birrien J-L, Schmid M, et al. Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. ISME J 2008;3(1):117–23.
[67] Dick GJ, Tebo BM. Microbial diversity and biogeochemistry of the guaymas basin deep-sea hydrothermal plume. Environ Microbiol 2010;12(5):1334–47.
[68] Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI, et al. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the mid-atlantic ridge. Environ Microbiol 2011;13(8):2158–71.
[69] Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. A comprehensive census of microbial diversity in hot springs of tengchong, yunnan province china using 16S rRNA gene pyrosequencing. PLoS ONE 2013;8(1):e53350.
[70] Lanzén A, Jørgensen SL, Bengtsson MM, Jonassen I, Øvreås L, Urich T. Exploring the composition and diversity of microbial communities at the jan mayen hydrothermal vent field using rna and DNA. FEMS Microbiol Ecol 2011;77(3):577–89.
[71] Rogers AD, Tyler PA, Connelly DP, Copley JT, James R, Larter RD, et al. The discovery of new deep-sea hydrothermal vent communities in the southern ocean and implications for biogeography. PLoS Biol 2012;10(1):e1001234.
[72] Sylvan JB, Toner BM, Edwards KJ. Life and death of deep-sea vents: bacterial diversity and ecosystem succession on inactive hydrothermal sulfides. mBio 2012;3(1).
[73] Voordeckers J, Do M, Hügler M, Ko V, Sievert S, Vetriani C. Culture dependent and independent analyses of 16S rRNA and atp citrate lyase genes: a comparison of microbial communities from different black smoker chimneys on the mid-atlantic ridge. Extremophiles 2008;12(5):627–40.
[74] Wang S, Xiao X, Jiang L, Peng X, Zhou H, Meng J, et al. Diversity and abundance of ammonia-oxidizing archaea in hydrothermal vent chimneys of the juan de fuca ridge. Appl Environ Microbiol 2009;75(12):4216–20.
[75] Yanagawa K, Kouduka M, Nakamura Y, Hachikubo A, Tomaru H, Suzuki Y. Distinct microbial communities thriving in gas hydrate-associated sediments from the eastern Japan sea. J Asian Earth Sci 2014;90:243–9.
[76] Zhou H, Li J, Peng X, Meng J, Wang F, Ai Y. Microbial diversity of a sulfide black smoker in main endeavour hydrothermal vent field, juan de fuca ridge. J Microbiol 2009;47(3):235–47.
[77] Shivaji S, Kumari K, Kishore KH, Pindi PK, Rao PS, Radha Srinivas TN, et al. Vertical distribution of bacteria in a lake sediment from antarctica by culture-independent and culture-dependent approaches. Res Microbiol 2011;162(2):191–203.
[78] Mikucki JA, Priscu JC. Bacterial diversity associated with blood falls, a subglacial outflow from the taylor glacier, antarctica. Appl Environ Microbiol 2007;73(12):4029–39.
[79] Tang C, Madigan MT, Lanoil B. Bacterial and archaeal diversity in sediments of west lake bonney, mcmurdo dry valleys, antarctica. Appl Environ Microbiol 2013;79(3):1034–8.
[80] Møller AK, Søborg DA, Al-Soud WA, Sørensen SJ, Kroer N. Bacterial community structure in high-arctic snow and freshwater as revealed by pyrosequencing of 16S rRNA genes and cultivation. Polar Res 2013;32:17390.
[81] Murray AE, Kenig F, Fritsen CH, McKay CP, Cawley KM, Edwards R, et al. Microbial life at −13 °C in the brine of an ice-sealed antarctic lake. Proc Natl Acad Sci 2012;109(50):20626–31.

[82] Nakai R, Abe T, Baba T, Imura S, Kagoshima H, Kanda H, et al. Microflorae of aquatic moss pillars in a freshwater lake, east antarctica, based on fatty acid and 16S rRNA gene analyses. Polar Biol 2012;35(3):425–33.
[83] Frank-Fahle BA, Yergeau É, Greer CW, Lantuit H, Wagner D. Microbial functional potential and community composition in permafrost-affected soils of the nw canadian arctic. PLoS ONE 2014;9(1):e84761.
[84] Ganzert L, Lipski A, Hubberten H-W, Wagner D. The impact of different soil parameters on the community structure of dominant bacteria from nine different soils located on livingston island, south shetland archipelago, antarctica. FEMS Microbiol Ecol 2011;76(3):476–91.
[85] Niederberger TD, McDonald IR, Hacker AL, Soo RM, Barrett JE, Wall DH, et al. Microbial community composition in soils of northern victoria land, antarctica. Environ Microbiol 2008;10(7):1713–24.
[86] Cary SC, McDonald IR, Barrett JE, Cowan DA. On the rocks: the microbiology of antarctic dry valley soils. Nat Rev Micro 2010;8(2):129–38.
[87] Aislabie JM, Jordan S, Barker GM. Relation between soil classification and bacterial diversity in soils of the ross sea region, antarctica. Geoderma 2008;144(1–2):9–20.
[88] de la Torre JR, Goebel BM, Friedmann EI, Pace NR. Microbial diversity of cryptoendolithic communities from the mcmurdo dry valleys, antarctica. Appl Environ Microbiol 2003;69(7):3858–67.
[89] Bajerski F, Wagner D. Bacterial succession in antarctic soils of two glacier forefields on larsemann hills, east antarctica. FEMS Microbiol Ecol 2013;85(1):128–42.
[90] Yergeau E, Newsham KK, Pearce DA, Kowalchuk GA. Patterns of bacterial diversity across a range of antarctic terrestrial habitats. Environ Microbiol 2007;9(11):2670–82.
[91] McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME J 2012;6:610–8.
[92] Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, et al. The european nucleotide archive. Nucleic Acids Res 2011;39(suppl 1):D28–31.
[93] Sun S, Chen J, Li W, Altintas I, Lin A, Peltier S, et al. Community cyberinfrastructure for advanced microbial ecology research and analysis: the camera resource. Nucleic Acids Res 2011;39(suppl 1):D546–51.
[94] Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics rast server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 2008;9(1):386.
[95] Bunge J. Estimating the number of species with catchall. Pac Symp Biocomput 2011:121–30.
[96] Behnke A, Bunge J, Barger K, Breiner HW, Alla V, Stoeck T. Microeukaryote community patterns along an O_2/H_2S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). Appl Environ Microbiol 2006;72(5):3626–36.
[97] Hong S-H, Bunge J, Jeon S-O, Epstein SS. Predicting microbial species richness. Proc Natl Acad Sci USA 2006;103(1):117–22.
[98] Stoeck T, Kasper J, Bunge J, Leslin C, Ilyin V, Epstein S. Protistan diversity in the arctic: a case of paleoclimate shaping modern biodiversity? PLoS ONE 2007;2(8):e728.
[99] Zuendorf A, Bunge J, Behnke A, Barger KJA, Stoeck T. Diversity estimates of microeukaryotes below the chemocline of the anoxic mariager fjord, denmark. FEMS Microbiol Ecol 2006;58(3):476–91.
[100] Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. Appl Environ Microbiol 2011;77(11):3846–52.
[101] Youssef NH, Couger MB, Elshahed MS. Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere. PLoS ONE 2010;5(8):e12414.
[102] Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil ph as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 2009;75(15):5111–20.
[103] Webster NS, Taylor MW, Behnam F, Lucker S, Rattei T, Whalan S, et al. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. Environ Microbiol 2010;12(8):2070–82.
[104] Hollister EB, Engledow AS, Hammett AJM, Provin TL, Wilkinson HH, Gentry TJ. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J 2010;4(6):829–38.
[105] SchÜtte UME, Abdo Z, Foster J, Ravel J, Bunge J, Solheim B, et al. Bacterial diversity in a glacier foreland of the high arctic. Mol Ecol 2010;19:54–66.
[106] Elshahed MS, Youssef NH, Spain AM, Sheik C, Najar FZ, Sukharnikov LO, et al. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl Environ Microbiol 2008;74(17):5422–8.
[107] Kirk Harris J, Gregory Caporaso J, Walker JJ, Spear JR, Gold NJ, Robertson CE, et al. Phylogenetic stratigraphy in the guerrero negro hypersaline microbial mat. ISME J 2013;7(1):50–60.
[108] Kelly J, Peterson E, Winkelman J, Walter T, Rier S, Tuchman N. Elevated atmospheric CO_2 impacts abundance and diversity of nitrogen cycling functional genes in soil. Microb Ecol 2013;65(2):394–404.
[109] Borrel G, Lehours AC, Bardot C, Bailly X, Fonty G. Members of candidate divisions OP11, OD1 and SR1 are widespread along the water column of the meromictic Lake Pavin (France). Arch Microbiol 2010;192(7):559–67.
[110] Youssef N, Steidley BL, Elshahed MS. Novel high-rank phylogenetic lineages within a sulfur spring (Zodletone spring, Oklahoma), revealed using a combined pyrosequencing-Sanger approach. Appl Environ Microbiol 2012;78(8):2677–88.
[111] Lynch MDJ, Bartram AK, Neufeld JD. Targeted recovery of novel phylogenetic diversity from next-generation sequence data. ISME J 2012;6(11):2067–77.
[112] Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol 1996;178(3):591–9.
[113] Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 2004;428(6978):37–43.
[114] Isenbarger TA, Finney M, Rios-Velazquez C, Handelsman J, Ruvkun G. Miniprimer pcr, a new lens for viewing the microbial world. Appl Environ Microbiol 2008;74(3):840–9.
[115] Ettwig KF, van Alen T, van de Pas-Schoonen KT, Jetten MS, Strous M. Enrichment and molecular detection of denitrifying methanotrophic bacteria of the NC10 phylum. Appl Environ Microbiol 2009;75(11):3656–62.
[116] Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun 2013;4:2120.
[117] Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC, et al. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. Elife 2013;2:e01102.
[118] Hug L, Castelle C, Wrighton K, Thomas B, Sharon I, Frischkorn K, et al. Community genomic analyses constrain the distribution of metabolic traits across the chloroflexi

phylum and indicate roles in sediment carbon cycling. Microbiome 2013;1(1):22.
[119] Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. mBio 2013;4(5):e00708–e007013.
[120] Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Rese 2013;23(1):111–20.
[121] Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature 2013;499(7459): 431–7.
[122] Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. ISME J 2015;9(2):447–60.
[123] Stamatakis A. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014.
[124] Hall BG. Building phylogenetic trees from molecular data with mega. Mol Biol Evol 2013;30(5):1229–35.
[125] Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. Arb: a software environment for sequence data. Nucleic Acids Res 2004;32(4):1363–71.
[126] Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. J Microbiol Meth 2003;55:541–55.