Annu. Rev. Microbiol. 2003. 57:369–94
doi: 10.1146/annurev.micro.57.030502.090759

First published online as a Review in Advance on July 30, 2003

THE UNCULTURED MICROBIAL MAJORITY

Michael S. Rappé and Stephen J. Giovannoni
Department of Microbiology, Oregon State University, Corvallis, Oregon 97331; email: michael.rappe@orst.edu; steve.giovannoni@orst.edu

Key Words molecular phylogeny, 16S rRNA, microbial diversity, candidate phylum

■ **Abstract** Since the delineation of 12 bacterial phyla by comparative phylogenetic analyses of 16S ribosomal RNA in 1987 knowledge of microbial diversity has expanded dramatically owing to the sequencing of ribosomal RNA genes cloned from environmental DNA. Currently, only 26 of the approximately 52 identifiable major lineages, or phyla, within the domain Bacteria have cultivated representatives. Evidence from field studies indicates that many of the uncultivated phyla are found in diverse habitats, and some are extraordinarily abundant. In some important environments, including seawater, freshwater, and soil, many biologically and geochemically important organisms are at best only remotely related to any strain that has been characterized by phenotype or by genome sequencing. Genome sequence information that would allow ribosomal RNA gene trees to be related to broader patterns in microbial genome evolution is scant, and therefore microbial diversity remains largely unexplored territory.

CONTENTS

INTRODUCTORY REMARKS 369
A CURRENT PICTURE OF THE BACTERIAL PHYLA 373
Bacterial Diversity 373
Phyla that are Underrepresented in Culture 376
Emergence of the Candidate Phyla of Uncultured Microorganisms 379
EVOLUTION AND THE SIGNIFICANCE OF GENE CLUSTERS 383
CLOSING REMARKS 387

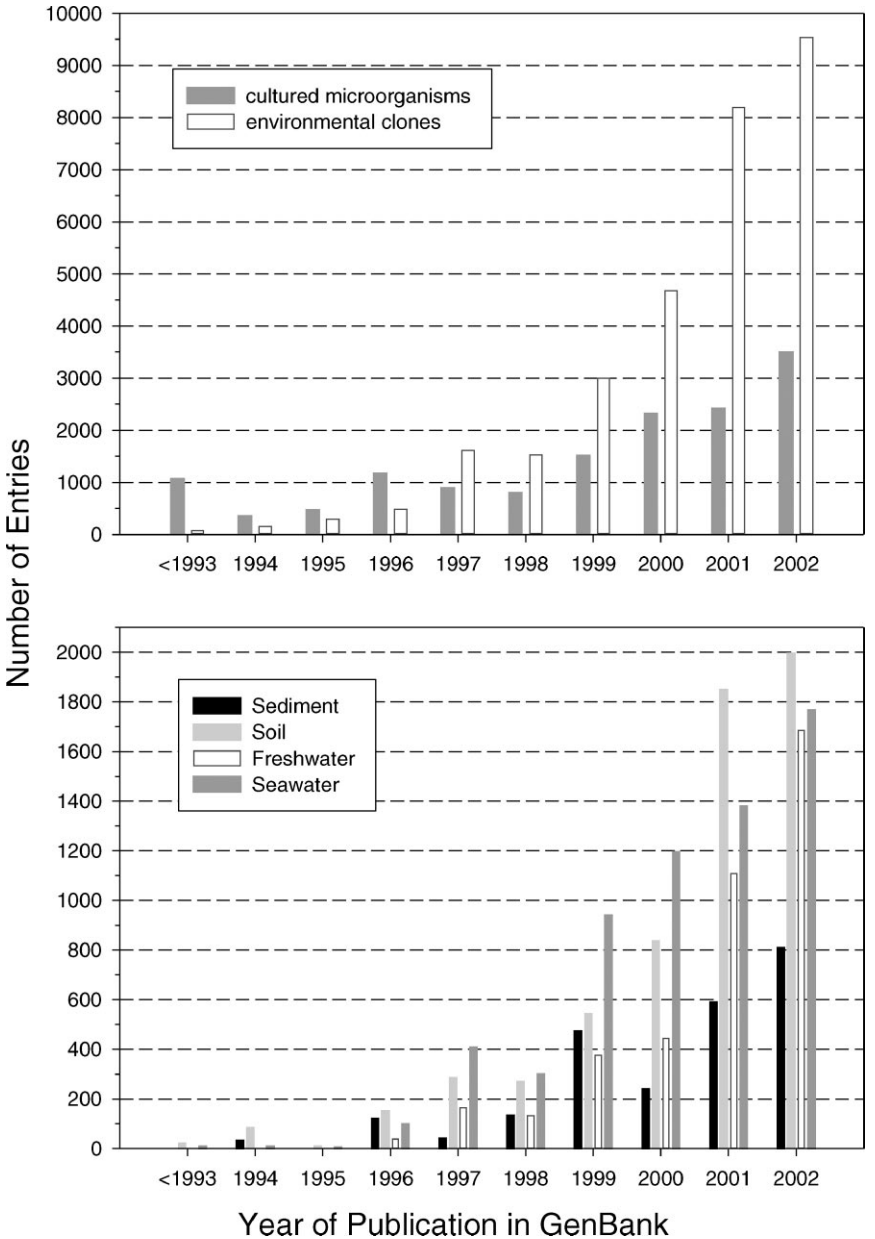
INTRODUCTORY REMARKS

Our title refers not to the sensibilities of our colleagues, and especially not to their political clout, but instead to new knowledge about microorganisms that have never been studied in culture. The discovery, cultivation, and study of novel organisms have been important features of microbiology since Koch. Nonetheless, uncultured microbial diversity was a relatively minor issue among microbiologists

in 1985, when a group led by Norman Pace began publishing papers founded on Pace's idea that microbial diversity could be explored by applications of molecular biology (76, 79). Early progress was slow, but the key ideas were established: Organisms could be identified without cultivation by retrieving and sequencing macromolecules from nature, and oligonucleotide probes could be used to manipulate, identify, and quantify molecules from different organisms (1, 107). Most of the new technology being developed for microbial ecology was based on the molecular phylogeny of ribosomal RNA (rRNA), particularly the small-subunit (SSU or 16S) rRNA, an approach for studying evolution that Carl Woese had pioneered specifically to solve the problem of resolving the earliest events in microbial evolution (112, 113). The ideas caught on, and today are widely employed by microbial ecologists.

However, the rise of molecular microbial ecology does not entirely explain the significance that the uncultured microbial majority has taken on. Other microbial ecologists counted cells and tallied the results to show that by sheer biomass microorganisms dominated many ecosystems [e.g., (111)]. Biogeochemists studied widespread chemical transformations that are consequences of microbial metabolism, and showed that they are both huge in magnitude and impact and complex in their details [e.g., (9)]. Collectively these observations created an emerging picture of vast populations of microorganisms that often had ancient evolutionary origins, but whose activities remained entirely enigmatic.

Figure 1 Graphs depicting the number of 16S ribosomal RNA gene sequences published in GenBank since 1993 (8). *Top*: total number of published 16S rRNA gene sequences from cultivated Bacteria and Archaea ($n = 14{,}434$) versus sequences derived from cultivation-independent studies ($n = 29{,}505$) as a function of year. *Bottom*: total number of published environmental gene clone sequences derived from sediment ($n = 2435$), soil ($n = 6037$), freshwater ($n = 3951$), and seawater ($n = 6104$) habitats as a function of year. All sequences published prior to 1993 are grouped in the first (<1993) column, while the "2002" column includes sequences published through November 19, 2002. As there is currently no standard convention in GenBank for naming or identifying rRNA gene sequences obtained from cultivation-independent studies, the values reported here should be considered indicative rather than absolute. The following search strategies were employed in Entrez. Cultured microorganisms: (SSU OR 16S OR small subunit) AND (rRNA OR rDNA OR ribosomal RNA) AND (bacteri* OR prokaryot* OR archae* OR eubacteri*) NOT (uncult* OR unidentified OR unknown). The same search was used for environmental gene clones, except the last term was changed to "AND (uncult* OR unidentified OR unknown)." Environmental clones derived from sediments and soils were obtained by adding the terms "AND sediment" or "AND soil," respectively. Seawater environmental clones were obtained by adding the search term "[AND (marine OR seawater OR sea water OR ocean)] NOT sediment," while freshwater environmental clones were obtained by adding the search term "[AND (freshwater OR lake OR river OR pond)] NOT sediment."


Number of Entries
cultured microorganisms
environmental clones
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
<1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
Sediment
Soil
Freshwater
Seawater
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Year of Publication in GenBank

The rise in knowledge about uncultured microbial diversity is depicted graphically in Figure 1, which shows GenBank entries as a function of year for 16S rRNA genes that were retrieved both from cultured isolates and directly from natural environments via cultivation-independent approaches. Environmental gene clone sequences began to appear in large numbers in 1996, by which time the technology for retrieving these had become routine. In 2002 alone, approximately 9500 environmental 16S rDNA clone sequences were published in GenBank (Figure 1). In early years, genes from marine systems dominated the entries (Figure 1). For technical reasons, seawater is easier to work with than soils, but by 2001 entries from soils led, probably because of their agricultural significance and also because soils are far more diverse. Therefore, the job of defining soil biodiversity ultimately requires a much greater sequencing effort.

The remainder of this review deals with interpreting the accumulated environmental rRNA gene sequences from GenBank, mainly by phylogenetic analysis. There are limits to this analysis. rRNA gene phylogenies will not reveal many things we want to know; for example, what unique innovations in cell function drove the evolution of microbial groups, and what are the important biochemical activities of these organisms in the environment? The horizontal exchange of genetic material has affected the evolution of cells so profoundly that their properties can only be known by directly studying the cells or their whole genomes (14, 74). However, notwithstanding the dithering of patterns of cellular evolution wrought by horizontal gene transfer, rRNA gene trees continue to be the critical source of information used to identify groups of microorganisms in nature and to organize relationships among them (64, 78). Scientists continue to utilize rRNA approaches extensively because they bear fruit, apparently because populations of cells in nature correlate well with rRNA sequence, so the sequences for the most part serve well as genetic markers. The other issue is phylogenetics. Do rRNA gene trees say anything about evolution? Genome sequence information is now pouring in so rapidly that it would be incautious to predict the limits of rRNA phylogenies. Nonetheless, early efforts to reconstruct phylogenies from whole genomes (90) seem to suggest that rRNA phylogenies suffer mainly from two problems: They reveal only a small part of the complex picture of cellular evolution, and they are subject to common phylogenetic errors, particularly erroneous associations between rapidly evolving lineages (83), problems with compositional bias (61), and incorrect arrangements of deeply branching lineages in cases where rRNA sequences simply lack the resolving power by virtue of their limited size (64, 83).

Key questions we address in this review are, How many microbial phyla are recognized at present, and what can we discern about microbial evolution from uncultured microbial diversity other than to regard them as rosters of species that await discovery? Recent reviews with similar objectives by Hugenholtz and coworkers (49, 50) concluded that there were about 36 bacterial phyla, a threefold increase from the 12 phyla described by Woese in 1987 (112). Of the 36 recognized by Hugenholtz, one third were “candidate” phyla, so-called because they had no

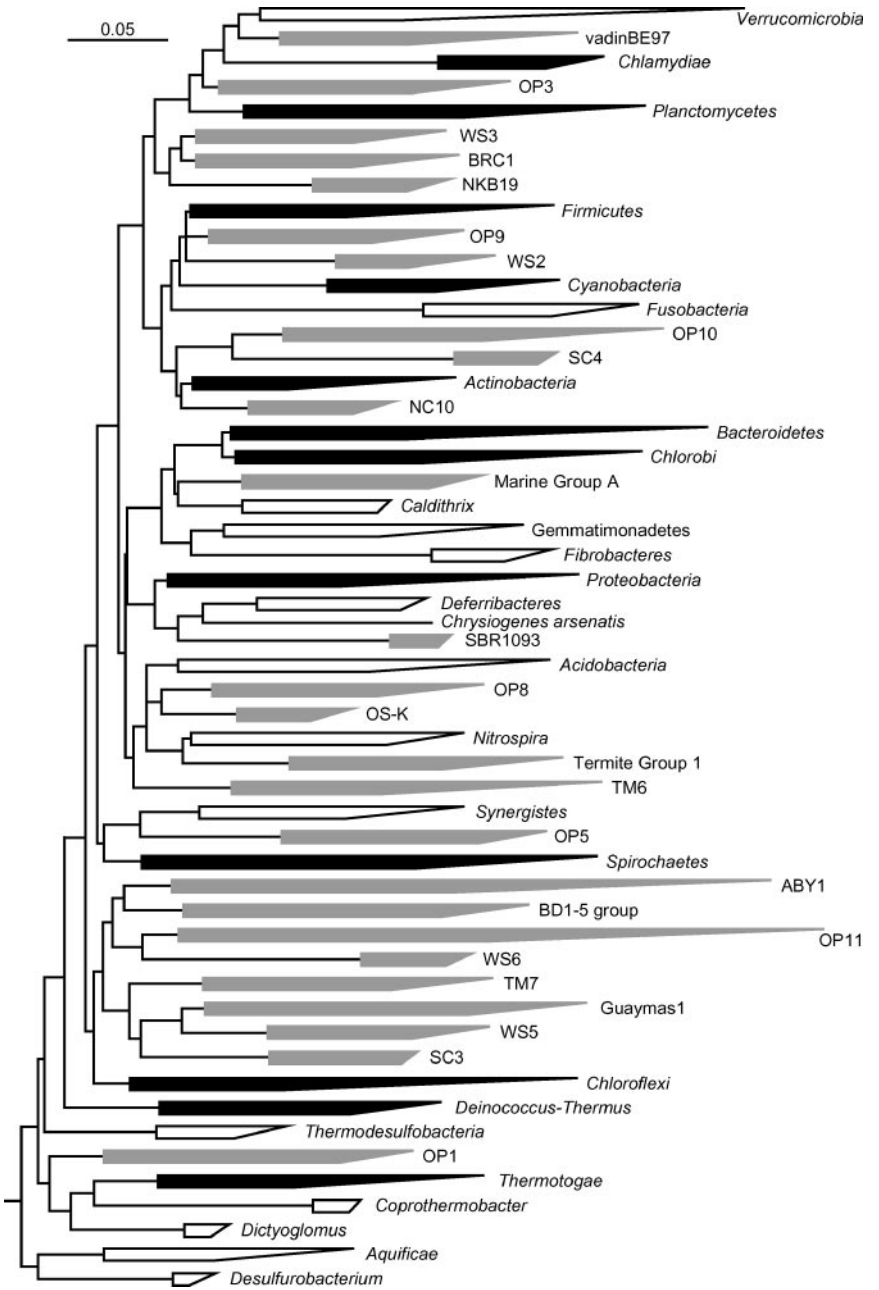
representatives in culture. The analyses we present here indicate that presently about 52 phyla can be discerned, of which 26 are candidate phyla, i.e., they are known only from environmental gene clone sequences. A second key question is, What do the data reveal about the broad patterns of evolution among some of these taxa, and how well can we expect knowledge gained from examining representative cultures to represent general trends among these phyla? While an exhaustive analysis of each of the 52 recognized phyla is not possible within the confines of this review, several phyla are discussed in detail. Finally, does the appearance of gene clusters in environmental gene clone libraries reveal aspects of the evolution of species in nature?

A CURRENT PICTURE OF THE BACTERIAL PHYLA

Bacterial Diversity

In the late 1970s and early 1980s, Woese and colleagues published a series of landmark papers in which they employed the comparative analysis of 16S rRNAs from cultivated microorganisms, obtained mostly via oligonucleotide cataloging, to sketch a framework for the phylogenetic diversity of Bacteria [e.g., (31, 37, 94, 114)], culminating in 1987 with a review aptly titled *Bacterial Evolution* (112). From the 16S rRNA sequences and catalogs available through 1987, Woese and colleagues were able to delineate 11 major groups or lineages, which since have variously been referred to at the taxonomic rank of kingdom (4), phylum (49), class (92, 93), order (87), and division (50, 51). Hereafter we follow the example set by one of the leading taxonomic resources for microbiologists, *Bergey's Manual of Systematic Bacteriology*, and refer to these groups as phyla (10, 49).

A 16S rRNA phylogenetic tree that broadly depicts relationships within the domain Bacteria is shown in Figure 2. The 11 original phyla recognized in 1987 are shown as 12 black wedges—the gram-positive bacteria are now recognized as two separate phyla, the *Firmicutes* (low G+C) and *Actinobacteria* (high G+C). Among the original phyla are the well-known *Proteobacteria* (92). These are the classical gram-negative bacteria and, on the basis of cultivation-dependent and -independent approaches, are generally recognized as one of the most successful microbial groups on the planet [e.g., (13, 38, 119)]. This phylum includes two of the most studied genera of microorganisms, *Escherichia* and *Pseudomonas* (35). Also included in the original 12 phyla are the *Cyanobacteria*, or oxygenic photosynthetic bacteria (and chloroplasts). Other identified phyla include the *Thermotogae*, which are sheathed, obligately anaerobic, fermentive heterotrophs (47). A few surprises were also included in the original 12 phyla. For example, it was discovered that the *Chloroflexi* (green non-sulfur) phylum, as originally defined, included the thermophilic phototroph *Chloroflexus*, the mesophilic, gliding chemoheterotroph *Herpetosiphon*, and the hyperthermophilic chemoheterotroph *Thermomicrobium* (77). This was unexpected, as these microbes exhibit divergent metabolic strategies and limited evidence existed previously which pointed to a shared evolutionary


0.05
Verrucomicrobia
vadinBE97
Chlamydiae
OP3
Planctomycetes
WS3
BRC1
NKB19
Firmicutes
OP9
WS2
Cyanobacteria
Fusobacteria
OP10
SC4
Actinobacteria
NC10
Bacteroidetes
Chlorobi
Marine Group A
Caldithrix
Gemmatimonadetes
Fibrobacteres
Proteobacteria
Deferribacteres
Chrysiogenes arsenatis
SBR1093
Acidobacteria
OP8
OS-K
Nitrospira
Termite Group 1
TM6
Synergistes
OP5
Spirochaetes
ABY1
BD1-5 group
OP11
WS6
TM7
Guaymas1
WS5
SC3
Chloroflexi
Deinococcus-Thermus
Thermodesulfobacteria
OP1
Thermotogae
Coprothermobacter
Dictyoglomus
Aquificae
Desulfurobacterium

history. Also a surprise was the discovery that the genera *Cytophaga*, *Bacteroides*, and *Flavobacterium* form a major lineage (82), known now as the *Bacteroidetes* phylum (36). Again, no phenotypic evidence was recognized previously to indicate that these microorganisms formed a single, coherent bacterial phylum.

An additional 14 (or so) phyla with cultivated representatives, shown as white wedges in Figure 2, have been identified since 1987 (49, 50, 78). Included in this group are several phyla of predominantly thermophilic microorganisms, including the *Aquificae*, *Thermodesulfobacteria*, *Dictyoglomi*, *Coprothermobacteria*, *Caldithrix*, and *Desulfurobacteria*. The cultivated members of many of these phyla grow chemolithoautotrophically: The phylum *Aquificae* is best known for the ability of some of its members to oxidize hydrogen gas as an energy source for chemolithotrophic growth (48), the *Desulfurobacteria* grow by anaerobic sulfur reduction (57), and *Thermodesulfobacterium hydrogenophilum* grows by sulfate reduction (54). Interestingly, many of the recently recognized thermophilic phyla also appear to branch early in the bacterial phylogenetic tree (Figure 2).

In addition to the thermophilic phyla, the recently recognized phylum *Acidobacteria* (63) is noteworthy because it is apparently ubiquitous and abundant in nature, as based on results from cultivation-independent molecular ecology studies, especially in soils [e.g., (4, 13)]. This phylum includes three species of divergent physiology (*Acidobacterium capsulatum*, *Holophaga foetida*, and '*Geothrix fermentans*'), making it difficult to predict many characteristics of the *Acidobacteria*-related microorganisms detected in environmental samples (4, 50). The phylum *Nitrospira* is also noteworthy in that it includes some interesting members such as the obligately chemolithotrophic, nitrite-oxidizing genus *Nitrospira*,

←

Figure 2 Phylogenetic tree illustrating the major lineages (phyla) of the domain Bacteria. Wedges shown in black are the 12 original phyla, as described by Woese (112), in white are the 14 phyla with cultivated representatives recognized since 1987, and in gray are the 26 candidate phyla that contain no known cultivated representatives. This evolutionary distance dendogram was constructed by the comparative analysis of over 600 nearly full-length 16S ribosomal RNA gene sequences using the ARB sequence analysis software package (65), selected from a larger database of over 12,000 sequences. A modified version of the "Lane mask" was employed in this analysis (55), along with the Olsen evolutionary distance correction and neighbor-joining tree-building algorithm. Horizontal wedge distances indicate the degree of divergence within a given phylum. The scale bar corresponds to 0.05 changes per nucleotide position. Phylum names are designated by selecting the first applicable option out of the following: (*a*) their convention in *Bergey's Manual of Systematic Bacteriology*, if it exists (36); (*b*) the first described representative genus within the phylum if it has cultivated representatives; (*c*) the first label given to a candidate phylum if previously published; or (*d*) the first clones or environment where the first clones were retrieved, for previously unnamed candidate phyla.

for which the phylum is named (29), and the obligately chemolithotrophic, ferrous iron-oxidizing genus *Leptospirillum* (45, 69). Members of this phylum are also apparently ubiquitous in the natural environment, judging by the large number of *Nitrospira* rRNA gene clones that have been reported in the past few years [e.g., (44, 46, 102)].

Phyla that are Underrepresented in Culture

With the advent of cloning and sequencing ribosomal RNA genes directly from the environment to access and explore the phylogenetic diversity of natural microbial communities, it has become clear that we have barely scratched the surface at obtaining representative cultures that span the phylogenetic breadth of most of the major phyla of Bacteria. While it would be easy to argue that a majority of the approximately 26 bacterial phyla currently with cultivated representatives are not well represented by these isolates, we highlight only 4 of special interest owing to their ubiquity in cultivation-independent surveys of microbial diversity. This topic has been reviewed recently (50); here we update two of the phyla included in that review (the *Verrucomicrobia* and *Chloroflexi*) and draw attention to two additional phyla that warrant further exploration.

The phylum *Verrucomicrobia* has been recognized as a separate phylum of Bacteria since 1995 (43, 109), but currently counts only a small number of cultivated microorganisms as members. This phylum was originally known from two genera and five species of prosthecate, aerobic heterotrophs isolated from freshwater environments: *Verrucomicrobium vinosum*, *Prosthecobacter fusiformis*, *P. debontii*, *P. vanneervenii*, and *P. dejongeii* (42, 43), although an obligately anaerobic, heterotrophic genus (*Opitutus terrae*) from rice paddy field soil has recently been isolated and characterized (16). In addition to the named microorganisms, three taxonomically uncharacterized isolates of "ultramicrobacteria" from rice paddy field soil with 16S rRNA gene sequences closely related to *O. terrae* have also been described (53). A dendogram of over 80 nearly complete 16S rRNA gene sequences depicting phylogenetic relationships within the *Verrucomicrobia* shows this phylum to consist of five clearly distinguishable subgroups (Figure 3, follow the Supplemental Material link from the Annual Reviews home page at http://www.annualreviews.org). This is congruent with the analysis of Hugenholtz and colleagues and indicates that no new major divisions within this phylum have been recovered since 1998 (50). Three of the five subgroups within the *Verrucomicrobia* contain no cultivated representatives, and environmental gene clones dominate the remaining two as well (Figure 3). Subdivision 1 contains the genera *Prosthecobacter* and *Verrucomicrobium* and appears to consist of bacteria of predominantly freshwater origin. Subgroups 2–4 are dominated by gene sequences of soil and freshwater origin, including *O. terrae* and related isolates in subgroup 4, while subgroup 5 is dominated by gene sequences recovered from marine sediments (Figure 3). From a review of Figure 3, it is readily apparent that a rich diversity of microorganisms awaits exploration within this phylum, which includes

major clusters of rRNA gene clones that are ubiquitous in natural freshwater and soil microbial communities (13, 119).

The phylum *Chloroflexi*, known colloquially as the green non-sulfur bacteria, was one of the initial group of phyla recognized early on in the application of small subunit ribosomal RNA sequencing to taxonomic questions within the Bacteria, and is generally considered to be a deep-branching lineage of the domain Bacteria (37, 77, 112). The limited number of cultivated members of this phylum is an interesting group of diverse phenotypes, which include gliding, filamentous isolates that contain some form of bacteriochlorophyll frequently arranged in chlorosomes (*Chloroflexus*, *Oscillochloris*, *Chloronema*, and *Heliothrix*); gliding, filamentous, mesophilic, strict aerobic chemoheterotrophs (*Herpetosiphon*); a hyperthermophilic, irregular rod-shaped, nonmotile aerobic chemoheterotroph (*Thermomicrobium roseum*); and an irregular cocci-shaped isolate able to reductively dechlorinate tetrachloroethene ('*Dehalococcoides ethenogenes*'). Unexpectedly, cultivation-independent investigations of the diversity present in a wide variety of microbial communities have found members of this phylum to be ubiquitous in the natural environment (50) (Figure 4, follow the Supplemental Material link from the Annual Reviews home page at http://www.annualreviews.org). For example, *Chloroflexi* rRNA genes appear frequently in clone libraries constructed from subsurface oceanic bacterioplankton [e.g., (3, 38, 40)], freshwater bacterioplankton [e.g., (104, 119)], soils [e.g., (15, 27)], sediments [e.g., (12, 20, 100)], and geothermal hot springs [e.g., (51, 110)], among others.

The phylogenetic analysis presented in Figure 4 differs from one published previously for the phylum *Chloroflexi* in that we identify eight monophyletic clades within the *Chloroflexi*, as opposed to four (50). This is not surprising; even if employing identical sequence datasets for the *Chloroflexi*, different phylogenetic trees can arise from the choice of outgroup sequences, variation between sequence masks, and choice of method of phylogenetic inference. We believe that the main reason our phylogenetic tree differs from previous analyses is that the *Chloroflexi* sequence database has exploded in numbers and now includes a large number of full-length 16S rRNA sequences. As shown in Figure 4, the majority of cultivated *Chloroflexi* cluster together in subgroup 7, the only one that appears well circumscribed by cultivated representatives. '*Dehalococcoides ethenogenes*' is the only cultivated representative of subgroup 6 (70), while an unnamed isolate (sludge isolate strain UNI-1) is the only cultivated representative of subgroup 1 (89). The remaining five subgroups contain no cultivated members (Figure 4). Sekiguchi and coworkers employed a directed cultivation effort to isolate *Chloroflexi* strain UNI-1 after first identifying and localizing members of this phylum in methanogenic granular sludge via 16S rRNA gene cloning-based analyses coupled with fluorescence in situ hybridization (89). Ribosomal RNA-directed cultivation efforts such as that employed by Sekiguchi and coworkers hold promise for isolating other members of this phylum from the many natural habitats in which it thrives.

The *Planctomycetes* and the *Chlamydia* are the only phyla of bacteria known to have cell walls that are not composed of peptidoglycan. They have a number

of other odd features as well, particularly the presence of an intracellular compartment in some species that contains the cell's DNA (60). The first members of the phylum *Planctomycetes* were isolated in the 1970s as heterotrophs growing in dilute media (5, 88, 95), but the known physiological diversity within the phylum expanded with the discovery that the microorganisms responsible for anaerobic ammonia oxidation (the anammox reaction) were *Planctomycetes* (96). Four genera have been described: *Pirellula*, *Planctomyces*, *Isosphaera*, and *Gemmata*. These genera form coherent phylogenetic clades, as shown in the phylogenetic tree in Figure 5 (follow the Supplemental Material link from the Annual Reviews home page at http://www.annualreviews.org). At a relatively early stage in the application of environmental gene cloning methods, it became apparent that the *Planctomycetes* were far more prevalent in the environment than would have been suspected from their scant presence in culture collections [e.g., (23, 59)]. It is now apparent that they are common in soils and sediments, as well as the fresh, marine and hot spring environments with which they were originally associated. Figure 5 does not show diversity within the four described genera but instead shows the considerable diversity of *Planctomycetes* that lies outside these clades.

There has been some progress in the cultivation of deep-branching *Planctomycetes*. The anammox organism has been studied in enrichment cultures and given the provisional name *Candidatus* 'Brocadia anammoxidans' (96). Recently, Zengler and colleagues (117) developed new methods for isolating microorganisms in gel microdroplets. Among the many novel strains that were transiently cultivated by this method were multiple members of a diverse clade of uncultivated *Planctomycetes* (117) (Figure 5). This promising development suggests that the introduction of novel cultivation methods will soon lead to better representation of *Planctomycetes* in culture collections. Given the remarkable physiology of the planctomycetes responsible for the anammox reaction, there are good reasons to suspect this deeply diverging phylum may hold future surprises.

The phylum *Gemmatimonadetes* has only recently been recognized as a main line of descent within the Bacteria: Identified first as a candidate phylum in 2001 in two independent reports (52, 66), it has since been proposed as a phylum with the cultivation of *Gemmatimonas aurantiaca* (118). *G. aurantiaca* is the only cultivated representative of what has become a diverse assemblage of environmental rRNA gene clone sequences (Figure 6, follow the Supplemental Material link from the Annual Reviews home page at http://www.annualreviews.org). At least four subgroups can be clearly delineated within this phylum (118) (Figure 6). On the basis of the environments where the rRNA gene clones were retrieved, subgroup 1 appears to be restricted to soil and other terrestrial ecosystems, while subgroups 2 and 4 appear to be predominantly marine in origin, having been retrieved from marine sediments and in symbioses with marine sponges (Figure 6). *G. aurantiaca*, a gram-negative aerobic heterotroph that appears to accumulate polyphosphate, was isolated from an anaerobic-aerobic sequential batch reactor by employing plating methods aimed at targeting slowly growing cells (118). Zhang and coworkers (118) hypothesize that one identifying feature of this phylum may be that its

members possess a gram-negative cell envelope lacking diaminopimelic acid in their peptidoglycan similar to one possessed by *G. aurantiaca*.

Emergence of the Candidate Phyla of Uncultured Microorganisms

The widespread application of 16S rRNA gene cloning and sequencing methods to identify microorganisms in natural samples has revealed an extensive and, in many cases, unsuspected diversity within bacterial phyla thought to be well known to microbiologists, such as the *Planctomycetes* and *Chloroflexi* discussed in the previous section. Early on, it was also apparent that some of the recovered clone sequences simply did not appear to belong to any of the known bacterial phyla [e.g., (34, 59, 91, 107, 110)]. It was later discovered that a portion of these "unaffiliated" environmental gene clone sequences were providing scientists with the first evidence of such ubiquitous, but not yet recognized, phyla as the *Verrucomicrobia* and *Acidobacteria* (63, 109). While a wide variety of artifacts (e.g., chimeric gene clones, PCR errors, sequencing errors) and methodological errors (e.g., improper reference or outgroup taxon selection, inadequate quantity of sequence information, improper alignment, use of an inappropriate alignment mask) can (and still does) cause the misplacement of gene clone sequences in phylogenetic trees, the observation that unaffiliated clones from different studies frequently clustered together in later analyses to form monophyletic groups supported the conclusion that many of these sequences were in fact real and formed major lines of descent within the bacterial domain which did not contain cultivated relatives. These lineages have since become known as candidate divisions or phyla, with the term candidate implying that no cultures yet exist to represent the group (50, 51). Herein we use the phrase "candidate phylum" for these deeply diverging clusters of sequences that are phylogenetically equivalent to phyla of cultured microorganisms as delineated in *Bergey's Manual of Systematic Bacteriology* (10).

While the concept of candidate phyla existed prior to 1998, it was rather amorphous and not well defined in phylogenetic (or any other) terms. Before 1998 three groups of sequences composed solely of environmental gene clones were generally thought to form main lines of descent within the domain Bacteria: the OS-K group, named after a 16S rRNA gene clone recovered from a microbial mat of thermal Octopus Spring (107, 110); Marine Group A, named after gene clones recovered from Pacific Ocean bacterioplankton samples (34, 41); and Termite Group 1, named after gene clones recovered from the intestine of the termite *Reticulitermes speratus* (75). The year 1998 saw the publication of three papers from the Pace laboratory that helped form a coherent concept and framework for identifying and characterizing candidate bacterial phyla (26, 50, 51). In a nutshell, a candidate phylum is defined by Hugenholtz and coworkers as "an unaffiliated lineage in multiple analyses of datasets with varying types and number of taxa and having $<85\%$ identity to reported sequences, indicating its potential to represent a new bacterial division [phylum]" (51) and "as a lineage consisting of two or more 16S rRNA sequences

that are reproducibly monophyletic and unaffiliated with all other division [phylum]-level relatedness groups that constitute the bacterial domain" (21, 50).

Implicit in the above definition are two points that need to be seriously considered by scientists wishing to describe a new candidate phylum. First, the 16S rRNA gene sequences in question should be nearly complete (> ~1300 nucleotides). Although we feel this point cannot be overemphasized, it is frequently disregarded. While short (~300–600 nucleotides) sequences are usually ample to identify the nearest neighbors of an unknown gene sequence if it has close phylogenetic relatives within a well-established phylum, sequences of this length are usually not sufficient for phylogenetic placement at or near the base of the same established phylum or for lineages that potentially do not belong to any of the currently recognized phyla of Bacteria. Second, it is highly recommended that the definition of new candidate phyla be based on the analysis of two or more gene clone sequences from the lineage in question, preferably derived from clone libraries constructed from different PCR products. This can assist in multiple ways: It has the potential to permit a more reliable phylogenetic placement of the unidentified lineage by reducing treeing artifacts such as long-branch attraction (83), and it provides confidence that the lineage in question is not simply an undetected artifact that does not code for a native 16S rRNA gene from a microorganism in the sample of interest. Rather than propose candidate phyla on the basis of phylogenetically novel single sequences or small clusters of novel sequence fragments, we support the route taken by some of our colleagues who have chosen to label these as novel or unaffiliated bacterial lineages until data become available that support or refute their status as candidate phyla [e.g., (24, 46, 86, 100)].

In the process of providing the scientific community with a working definition, Hugenholtz and coworkers (51) described 12 lineages that potentially represented candidate phyla from a Yellowstone hot spring. Seven of these (prefix OP) appear in Figure 2 because they meet all the criteria described above. In that same year, Dojka and coworkers (26) described 6 lineages that potentially formed candidate phyla from a hydrocarbon- and chlorinated-solvent-contaminated aquifer, five of which appear in Figure 2 because they also fulfill the abovementioned criteria (prefix WS). The number of identified candidate phyla continues to grow: In the past two years alone, at least eight candidate phyla have emerged from the expanding public databases of 16S rRNA gene sequences (8, 65, 68). The following recently identified lineages are shown in Figure 2: SC3 and SC4, identified from arid soil (27); NC10, identified from aquatic microbial formations in flooded caves (46); BRC1, identified from bulk soil and rice roots of flooded rice microcosms (24); ABY1, identified from lake bacterioplankton and deep-sea sediments (104); Guaymas1, identified from hydrothermally active marine sediments (100); NKB19, identified from deep-sea sediments and activated sludge (52, 58); and SBR1093, identified from activated sludge (52, 56). Some uncertainty exists in determining the total number of candidate phyla; however, our estimate puts this number currently at about 26 (Figure 2).

Little information is available regarding most candidate phyla save for what can be gleaned from the environment(s) where the environmental DNA was retrieved.

Two exceptions to this include TM7 and WS6, which have been the focus of intense individual study (25, 52). Here we review the status of candidate phylum Marine Group A, with the aim of introducing our colleagues to a major group of environmentally significant bacteria for which little is currently known.

Candidate phylum Marine Group A has been known since 1993 from short sequence fragments recovered from Pacific Ocean bacterioplankton community DNA (34); the first full-length representatives were sequenced from Pacific Ocean and Sargasso Sea bacterioplankton samples a few years later (41). The analysis presented by Gordon & Giovannoni (41) is a good example of the type of information required to define a candidate phylum: Two gene clones from separate samples were sequenced completely, the clone sequences were scrutinized to confirm that they did not contain artifacts introduced in vitro, extensive database searching was performed to find closely related relatives such as the sequences generated by Fuhrman and coworkers (34), and the sequences were rigorously analyzed phylogenetically to determine their most closely related reference sequences (41). Initial phylogenetic analyses with the complete rRNA gene sequences indicated that this lineage represented a distinct line of descent within the bacterial domain bearing distant phylogenetic affinity to the phyla *Fibrobacteres* and *Chlorobi* (41). In more recent analyses, we frequently recover Marine Group A in a higher-order grouping with the phyla *Fibrobacteres*, *Gemmatimonadetes*, *Caldithrix*, *Chlorobi*, and *Bacteroidetes* (Figure 2). The phylum *Caldithrix*, named after a recently described genus of strictly anaerobic, mixotrophic thermophiles isolated from a hydrothermal vent (71), frequently branches as a sister phylum of Marine Group A, although this relationship is frequently not well-supported statistically.

The small number of sequences that originally comprised Marine Group A form a tightly knit cluster, sharing $>94\%$ similarity in pairwise sequence comparisons. This cluster is labeled as subgroup 1 in Figure 7 and can now be seen to make up only a small portion of the phylogenetic breadth of this lineage. In addition to other, more diverse environmental gene clones from oxic seawater habitats [e.g., (3, 32, 62, 85, 97)], members of this candidate phylum have also been recovered in abundance from anoxic seawater habitats (28, 67) and Antarctic marine sediments (11). While the sequences that comprise candidate phylum Marine Group A are exclusively marine in origin, the seven subgroups that can currently be phylogenetically delineated within this group can also be segregated on the basis of the environment from which they were recovered (Figure 7). Subgroups 1-3 contain sequences recovered almost exclusively from oxic seawater habitats, subgroups 4 and 5 are of marine sediment origin, subgroup 6 is composed solely of clone sequences recovered from anoxic seawater, and subgroup 7 originates from oxic seawater environmental DNA (Figure 7). Although little beyond a possible requirement for salt water and the ability of some members to grow anaerobically can be inferred about the physiology of Marine Group A, this group likely plays a significant role in the marine environment.

There are several reasons why certain ambiguity exists in determining the current number of recognizable bacterial phyla. For instance, several of the phylum-level lineages shown in Figure 2 are represented by a small number of 16S rRNA

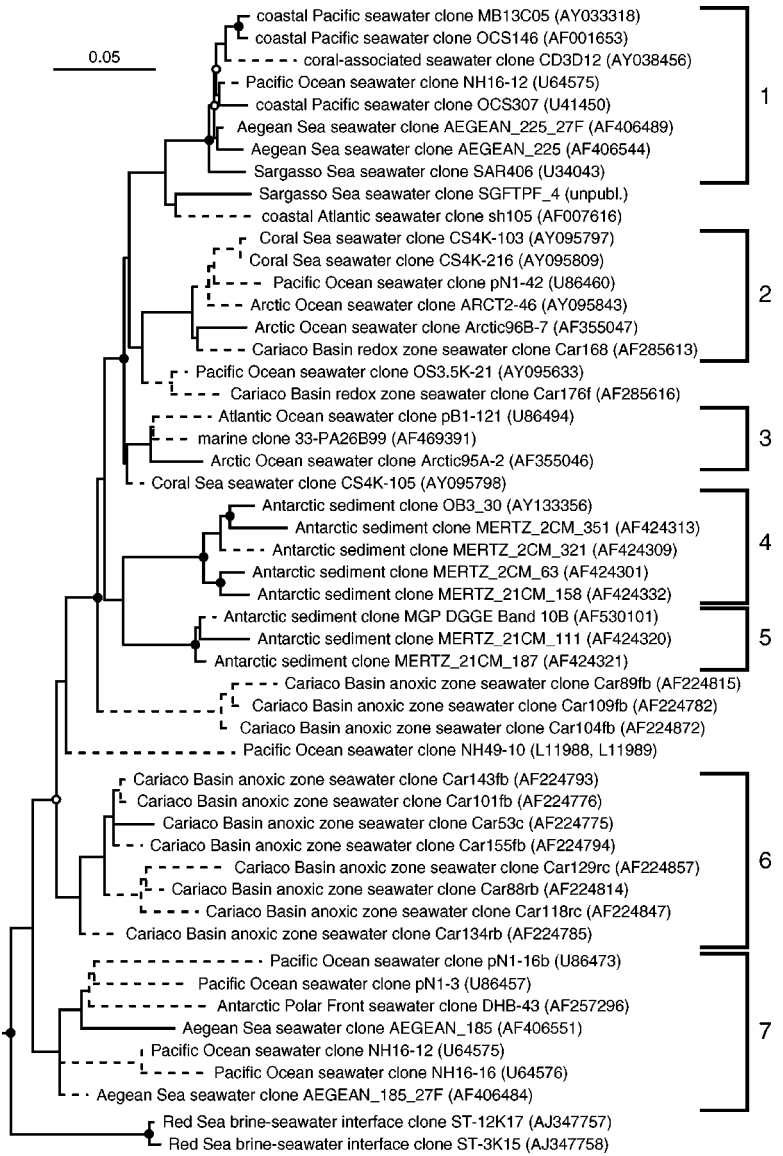


Figure 7 Phylogenetic tree showing relationships among representative members of the candidate phylum Marine Group A. In ARB, sequences >1000 nucleotides in length were used to construct the backbone of the dendogram using the Olsen evolutionary distance correction and neighbor-joining tree-building algorithm (solid branches), to which shorter-length sequences (dashed branches) were subsequently added with the ARB parsimony insertion tool (65). Bootstrap proportions were determined as in Figure 3 from the >1000 nucleotide dataset. Bootstrap, scale bar, and terminal node label conventions are as in Figure 3.

sequences. Thirteen of the phyla shown in Figure 2 include 5 or fewer sequences greater than 1300 nucleotides in length. One lineage shown in Figure 2 that potentially represents a bacterial phylum, *Chrysiogenes arsenatis* [the single-species phylum *Chrysiogenetes* in *Bergey's Manual for Systematic Bacteriology* (36)], is the only member of this lineage. Some of these are long phylogenetic branches that are separated from other phylum-level lineages by short internal nodes. As these lineages fill out, potential relationships with other phyla may become resolved, allowing the reorganization of these taxonomic groups inside previously recognized phyla. In addition, some of the previously recognized bacterial phyla, such as the phylum *Firmicutes* and class delta *Proteobacteria*, are not consistently resolved as monophyletic groups and may in fact be composed of several distinct, deep-branching lineages (i.e., phyla) of the domain Bacteria (64). We are optimistic that, as sequence databases and phylogenetic methods continue to evolve and/or comparative genomic approaches are employed [e.g., (90, 115)], the resolution of these relationships will become clear.

One result of the dramatic increase in the number of bacterial phyla is that the order of branching of these lineages no longer resembles an ordered progression of resolvable bifurcating nodes. Instead, the shape of the tree shown in Figure 2 represents a star-like radiation. While a few phyla such as *Aquificae, Thermotogae*, *Desulfurobacteria*, and *Thermodesulfobacteria* potentially branch early, most appear in rapid succession. We mention this to emphasize two points: First, the diversification of the Bacteria apparently occurred in a relatively short period of time. Second, the 16S rRNA gene does not appear able to resolve the order of branching between the major bacterial phyla with the analysis methods currently available.

EVOLUTION AND THE SIGNIFICANCE OF GENE CLUSTERS

Ribosomal RNA genes retrieved from nature almost invariably appear in phylogenetic trees as clusters of related genes; infrequently are two identical gene sequences retrieved, even from the same clone library. The name cluster was applied to these groupings to differentiate them from clades, a term that implies a set of related organisms (39). Genes cloned by PCR, as most environmental rRNAs are, can vary slightly in sequence owing to a number of causes. Early on it became clear that the patterns of nucleotide substitutions within gene clusters, particularly compensatory substitutions across helices, could only be explained by evolution. Although it can be proven that the gene clusters have an evolutionary basis, it remains uncertain to this day how much of the diversity within clusters is due to natural variation and how much is due to artifacts, such as chimeric genes formed from closely related sequences (105) and variation within paralogous rRNA operons (116). Diversity generated in vitro is nearly impossible to identify when the parent genes are highly related. Accurate measurements of the diversity within gene clusters are slowly coming forth from the sequencing of genes from cultured

isolates and of environmental genomic DNA cloned into bacterial artificial chromosome vectors (6).

Evolutionary processes in large populations of bacteria are not well understood at present. Perhaps the three most abundant groups of microorganisms known are the oceanic planktonic organisms 'Pelagibacter' [SAR11; (84)], the Group I Marine Crenarchaea (22, 33), and *Prochlorococcus* (17, 18). Population genetic theory predicts that advantageous mutations spread easily in such large populations because bottlenecks rarely occur. Theory also predicts that these organisms are well optimized for their environment because there is a large pool of diversity for selection to act upon.

Cohan (19) developed ideas that are useful for thinking about bacterial speciation in the context of phylogenetic trees. He suggests that "ecotypes" are revealed in the fine structure of phylogenetic trees and that it might be a good idea to define bacterial species as ecotypes. The concept of periodic selection is central to this idea. Periodic selection, also known by the term selective sweep, occurs when a new variant of a species, which has a significantly higher fitness than its conspecifics, arises. In a large asexual population such a variant eventually replaces its conspecifics. If environmental gene clusters are the consequence of neutral substitutions accumulating in large asexual populations as a function of time, then selective sweeps eliminate diversity within gene clusters, only to have them reform as descendents of the successful new cell variant accumulate neutral substitutions. An interesting feature of this model is that it predicts that phylogenetic trees made from environmental gene clones will contain a record of speciation and periodic selection. Below, we use two examples to illustrate this idea: the marine oxyphototroph clade (*Prochlorococcus* and marine *Synechococcus*) and the SAR11 gene cluster. Although this discussion is speculative, it serves to develop an important idea: Our modern view of microbial diversity is emerging from environmental studies, in which we increasingly interpret the evolution of microorganisms by considering their adaptations to a specific environmental context. Previously, it was more common to speculate about their evolution of organisms from their physiological attributes alone.

For example, consider the case of the marine oxyphototroph clade, which consists of the cyanobacteria *Prochlorococcus* and marine *Synechococcus* [Figure 8, *top*; also known as the marine picophytoplankton clade (103)]. Many examples of these microorganisms have been cultured [e.g., (17, 80)]. They are ubiquitous in the oceans, environmentally significant, and physiologically diverse, but they form a single clade of relatively shallow divergence in ribosomal RNA gene trees (103). The most distant sequences in this clade have sequence identities of 0.96, well above the boundary typically used to place microbial strains in separate genera. Assuming that 16S rRNA genes diverge in these organisms at a uniform rate that is about the same as that observed in other microorganisms [0.01–0.02 substitutions per site per 50 Ma (73)], the calculated age of this clade would be between 100 and 200 million years—far less than the probable history of unicellular cyanobacteria in seawater.

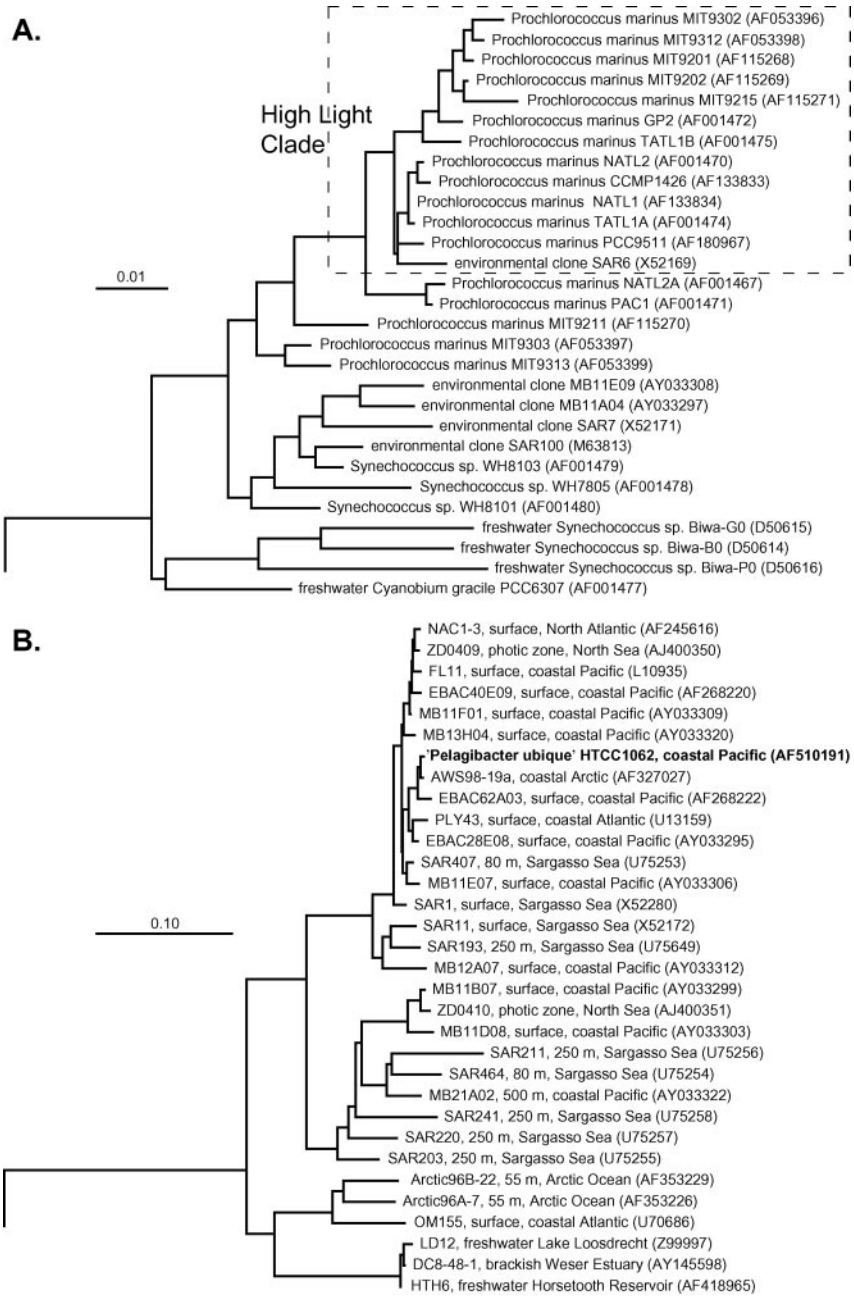


Figure 8 Phylogenetic trees showing relationships among representative members of the marine oxyphototroph clade (*panel A*) and the 'Pelagibacter' (SAR11) clade (*panel B*). The tree topologies were inferred with the ARB software package, using the Olsen evolutionary distance correction and neighbor-joining tree-building algorithm.

The marine oxyphototroph phylogenetic tree in Figure 8 has several interesting topological features that correlate with the physiology of the organisms. The deepest branch in the oxyphototroph clade separates *Prochlorococcus* from the marine *Synechococcus*, which differ markedly in their light-harvesting structures and ecology [(101); see also (81)]. The prochlorophytes lack the phycobiliprotein light-harvesting systems that are typical of other cyanobacteria and instead use a chlorophyll *a/b* system. The deepest branches within the *Prochlorococcus* subclade are ecotypes that are found near the bottom of the photic zone, have high Chl b/a_2 ratios, and can use NO_2^- or NH_4^+ as sources of inorganic nitrogen (72). Within the *Prochlorococcus* radiation, a subclade emerges which circumscribes strains adapted to higher light conditions; these strains are found at shallower depths in the water column, have low Chl b/a_2 ratios, and appear to rely exclusively on NH_4^+ as a source of inorganic nitrogen (72) (Figure 8).

The surprisingly high sequence identities within the marine oxyphototroph clade could be interpreted as evidence of a periodic selection event in which a single unicellular planktonic cyanobacterium acquired an advantageous genetic change that caused it to sweep competitors from the oceanic photic zone. The genetic data roughly place this event in the Mesozoic era, long after the origins of cyanobacteria and the oceans (2). The topology of the phylogenetic tree shown in Figure 8 also suggests that *Prochlorococcus* and marine *Synechococcus* diverged subsequently. There is no indication of what genetic change might have led to a selective sweep truncating diversity in the oxyphototroph clade, but insight into the evolution of this group may be forthcoming from the several marine oxyphototroph genomes that are now being sequenced.

Alpha proteobacteria of the SAR11 clade, another exclusively planktonic microbial group, form a far deeper clade with sequence identities as low as 0.87. In contrast to the marine oxyphototroph clade, the SAR11 clade currently contains only a small number of cultivated representatives, all of which contain nearly identical rRNA gene sequences (84) (Figure 8). The topology of this cluster also suggests evolutionary divergence into ecotypes (30) (Figure 8, *bottom*). The deepest branch of the SAR11 clade includes freshwater species. The strictly marine part of the SAR11 clade is divided into two subclades, which appear to roughly correspond to surface and deep ecotypes of SAR11 (30). The long internal segments that connect the subclades of SAR11 suggest a lengthy evolutionary history punctuated by the divergence of ecotypes and episodes of periodic selection. Likewise, the long segment connecting the SAR11 clade to the most closely related outgroups suggests that diversity within this clade was truncated by an episode of periodic selection that predated the divergence of the major SAR11 subclades.

The two clades of marine organisms that we have used as examples provide evidence that the fine structure of phylogenetic trees can provide useful information about evolution and functional specialization in natural microbial populations. Other reports not discussed here support similar conclusions [e.g., (106, 108)]. Most phylogenetic trees based partially on environmental gene clones show topological features that suggest evolutionary diversification, but rarely is there enough

information about the ecological distributions and phenotypic attributes of the organisms to enable an analysis such as that given above for the SAR11 and marine oxyphototroph clades. It appears likely that the fine structure of phylogenetic trees will increasingly be interpreted as more information becomes available. The extent to which Cohan's ideas will influence taxonomy remains to be seen, but the merit of his thinking is evident in the examples given above.

CLOSING REMARKS

Knowledge of microbial diversity, as it is revealed in the topology of 16S rRNA phylogenetic trees, will probably continue to expand the outlines of the domain Bacteria, but at an ever-decreasing rate. It seems likely that the major advances ahead will come about by genome sequencing and functional prediction from genome sequences [e.g., (7)]. In their landmark 1997 paper on the evolution of protein families, Tatusov and colleagues (99) remarked that clusters of orthologous genes (a set of homologs from different species) appeared to be "unique as a natural system" for protein classification. They were referring to their system of mapping genes into families, not cells. But, since it is genes that give cells the unique properties that allow them to expand their domains in nature, it seems likely that correlating the evolution of gene families with cellular evolution (still a solid concept) is the most illuminating source of new information about microbial diversity.

ACKNOWLEDGMENTS

The authors wish to thank past and present members of the Giovannoni laboratory for stimulating interactions and discussion. We also thank Hui Zhang for sharing results prior to publication, Susan Barns for sharing unpublished analyses, Jang-Cheon Cho for sharing results regarding candidate phylum vadinBE97, and Phillip Hugenholtz for making his ARB database freely available for public download at the Ribosomal Database Project II website. Support from the National Science Foundation, the Oregon State University Molecular and Cellular Biology Program, and Diversa Corporation is gratefully acknowledged.

The *Annual Review of Microbiology* is online at http://micro.annualreviews.org

LITERATURE CITED

1. Amann RI, Ludwig W, Schleifer K-H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59:143–69
2. Anbar AD, Knoll AH. 2002. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* 297:1137–42
3. Bano N, Hollibaugh JT. 2002. Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl. Environ. Microbiol.* 68:505–18

4. Barns SM, Takala SL, Kuske CR. 1999. Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment. *Appl. Environ. Microbiol.* 65:1731–37
5. Bauld J, Staley JT. 1976. *Planctomyces maris* sp. nov.: a marine isolate of the *Planctomyces-Blastocaulis* group of budding bacteria. *J. Gen. Microbiol.* 97:45–55
6. Béjà O, Koonin EV, Aravind L, Taylor LT, Seitz H, et al. 2002. Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* 68:335–45
7. Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, et al. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2:516–29
8. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. 2002. GenBank. *Nucleic Acids Res.* 30: 17–20
9. Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, et al. 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407:623–26
10. Boone DR, Castenholz RW, Garrity GM, eds. 2001. *Bergey's Manual of Systematic Bacteriology*, Vol. 1. New York: Springer
11. Bowman JP, McCammon SA, McCuaig RD, Gibson JAE, Nichols PD. 2002. Characterization of continental shelf sediments collected off Antarctica: microbial metabolic activity, community structure and biogeography. http://www.ncbi.nlm.nih.gov/Entrez/
12. Bowman JP, Rea SM, McCammon SA, McMeekin TA. 2000. Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. *Environ. Microbiol.* 2:227–37
13. Buckley DH, Schmidt TM. 2002. Exploring the diversity of soil—a microbial rain forest. See Ref. 95a, pp. 183–208
14. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1058–73
15. Chandler DP, Brockman FJ, Bailey TJ, Fredrickson JK. 1998. Phylogenetic diversity of *Archaea* and *Bacteria* in a deep subsurface paleosol. *Microb. Ecol.* 36:37–50
16. Chin K-J, Liesack W, Janssen PH. 2001. *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division '*Verrucomicrobia*' isolated from rice paddy soil. *Int. J. Syst. Evol. Microbiol.* 51:1965–68
17. Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, et al. 1992. *Prochlorococcus marinus* nov. gen. Nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch. Microbiol.* 157:297–300
18. Chisholm SW, Olsen RJ, Zehler ER, Goericke R, Waterbury JB, Welschmeyer NA. 1988. A novel free-living prochlorophyte abundant in oceanic euphotic zone. *Nature* 334:340–43
19. Cohan FM. 2001. Bacterial species and speciation. *Syst. Biol.* 50:513–24
20. Coolen MJL, Cypionka H, Sass AM, Sass H, Overmann J. 2002. Ongoing modification of Mediterranean Pleistocene sapropels mediated by prokaryotes. *Science* 296:2407–10
21. Dalevi D, Hugenholtz P, Blackall LL. 2001. A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. *Int. J. Syst. Evol. Microbiol.* 51:385–91
22. DeLong EF. 1992. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* 89:5685–89
23. DeLong EF, Franks DG, Alldredge AL. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine

bacterial assemblages. *Limnol. Oceanogr.* 38:924–34
24. Derakshani M, Lukow T, Liesack W. 2001. Novel bacterial lineages at the subdivision level as detected by signature nucleotide targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Environ. Microbiol.* 67:623–31
25. Dojka MA, Harris JK, Pace NR. 2000. Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of bacteria. *Appl. Environ. Microbiol.* 66:1617–21
26. Dojka MA, Hugenholtz P, Haack SK, Pace NR. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* 64:3869–77
27. Dunbar J, Barns SM, Ticknor LO, Kuske CR. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. *Appl. Environ. Microbiol.* 68:3035–45
28. Eder W, Schmidt M, Koch M, Garbe-Schönberg D, Huber R. 2002. Prokaryotic phylogenetic diversity and corresponding geochemical data of the brine-seawater interface of the Shaban Deep, Red Sea. *Environ. Microbiol.* 4:758–63
29. Ehrich S, Behrens D, Lebedeva E, Ludwig W, Bock E. 1995. A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch. Microbiol.* 164:16–23
30. Field KG, Gordon D, Wright T, Rappé M, Urbach E, et al. 1997. Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* 63:63–70
31. Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, et al. 1980. The phylogeny of prokaryotes. *Science* 209:457–63
32. Fuhrman JA, Davis AA. 1997. Widespread *Archaea* and novel *Bacteria* from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* 150: 275–85
33. Fuhrman JA, McCallum K, Davis AA. 1992. Novel major archaebacterial group from marine plankton. *Nature* 356:148–49
34. Fuhrman JA, McCallum K, Davis AA. 1993. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Appl. Environ. Microbiol.* 59:1294–302
35. Galvez A, Maqueda M, Martinez-Bueno M, Valdivia E. 1998. Publication rates reveal trends in microbiological research. *ASM News* 64:269–75
36. Garrity GM, Holt JG. 2001. The road map to the *Manual*. See Ref. 10, 1:119–66
37. Gibson J, Ludwig W, Stackebrandt E, Woese CR. 1985. The phylogeny of the green photosynthetic bacteria: absence of a close relationship between Chlorobium and Chloroflexus. *Syst. Appl. Microbiol.* 6:152–56
38. Giovannoni S, Rappé M. 2000. Evolution, diversity and molecular ecology of marine prokaryotes. In *Microbial Ecology of the Oceans*, ed. DL Kirchman, pp. 47–84. New York: Wiley
39. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345: 60–63
40. Giovannoni SJ, Rappé MS, Vergin KL, Adair NL. 1996. 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria. *Proc. Natl. Acad. Sci. USA* 93:7979–84
41. Gordon DA, Giovannoni SJ. 1996. Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific Oceans. *Appl. Environ. Microbiol.* 62:1171–77
42. Hedlund BP, Gosink JJ, Staley JT. 1996. Phylogeny of *Prosthecobacter*, the fusiform caulobacters: members of a

recently discovered division of the *Bacteria*. *Int. J. Syst. Bacteriol.* 46:960–66
43. 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
44. Hentschel U, Hopke J, Horn M, Fredrich AB, Wagner M, et al. 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. *Environ. Microbiol.* 68:4431–40
45. Hippe H. 2000. *Leptospirillum* gen. nov. (ex Markosyan 1972), nom. rev., including *Leptospirillum ferrooxidans* sp. nov. (ex Markosyan 1972), nom. rev. and *Leptospirillum thermoferrooxidans* sp. nov. (Golovacheva et al. 1992). *Int. J. Syst. Evol. Microbiol.* 50:501–3
46. Holmes AJ, Tujula NA, Holley M, Contos A, James JM, et al. 2001. Phylogenetic structure of unusual aquatic microbial formations in Nullarbor caves, Australia. *Environ. Microbiol.* 3:256–64
47. Huber R, Langworthy TA, König H, Thomm M, Woese CR, et al. 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* 144:324–33
48. Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, et al. 1992. *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst. Appl. Microbiol.* 15:340–51
49. Hugenholtz P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3:reviews0003.1–03.8
50. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180:4765–74
51. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. 1998. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180:366–76
52. Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL. 2001. Investigation of candidate division TM7, a recently recognized major lineage of the domain *Bacteria* with no known pure-culture representatives. *Appl. Environ. Microbiol.* 67: 411–19
53. Janssen PH, Schuhmann A, Mörschel E, Rainey FA. 1997. Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl. Environ. Microbiol.* 63:1382–88
54. Jeanthon C, L'Haridon S, Cueff V, Banta A, Reysenbach AL, Prieur D. 2002. *Thermodesulfobacterium hydrogeniphilum* sp. nov. a thermophilic chemolithoautotrophic sulfate reducing bacterium isolated from a deep sea hydrothermal vent at Guaymas basin and emendation of the genus *Thermodesulfobacterium*. *Int. J. Syst. Evol. Microbiol.* 52:765–72
55. Lane DJ. 1991. 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*, ed. E Stackebrandt, M Goodfellow, pp. 115–75. New York: Wiley
56. Layton AC, Karanth PN, Lajoie CA, Meyers AJ, Gregory IR, et al. 2000. Quantification of *Hyphomicrobium* populations in activated sludge from an industrial wastewater treatment system as determined by 16S rRNA analysis. *Appl. Environ. Microbiol.* 66:1167–74
57. L'Haridon S, Cilia V, Messner P, Raguenes G, Gambacorta A, et al. 1998. *Desulfurobacterium thermolithotophum* gen. nov. sp. nov. a novel autotrophic, sulphur-reducing bacterium isolated from a deep sea hydrothermal vent. *Int. J. Syst. Bacteriol.* 48:701–11
58. Li L, Kato C, Horikoshi K. 1999. Microbial diversity in sediments collected from the deepest cold-sea area, the Japan Trench. *Mar. Biotechnol.* 1:391–400
59. Liesack W, Stackebrandt E. 1992. Occurrence of novel groups of the domain

Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* 174:5072–78
60. Lindsay MR, Webb RI, Strous M, Jetten MSM, Butler MK, et al. 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch. Microbiol.* 175: 413–29
61. Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol.* 34:153–62
62. López-García P, López-López A, Moreira D, Rodríguez-Valera F. 2001. Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol. Ecol.* 36:193–202
63. Ludwig W, Bauer SH, Bauer M, Held I, Kirchhof G, et al. 1997. Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol. Lett.* 153:181–90
64. Ludwig W, Klenk H-P. 2001. Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. See Ref. 10, 1:49–65
65. Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, et al. 1998. Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* 19:554–68
66. Madrid VM, Aller JY, Aller RC, Christoserdov AY. 2001. High prokaryote diversity and analysis of community structure in mobile mud deposits off French Guiana identification of two new bacterial candidate divisions. *Environ. Microbiol.* 37:197–209
67. Madrid VM, Taylor GT, Scranton MI, Christoserdov AY. 2001. Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. *Appl. Environ. Microbiol.* 67:1663–74
68. Maidak BL, Cole JR, Lilburn TG, Parker CT Jr, Saxman PR, et al. 2001. The RDP-II (Ribosomal Database Project). *Nucleic Acids Res.* 29:173–74
69. Markosyan GE. 1972. A new iron-oxidizing bacterium, *Leptospirillum ferrooxidans* gen. et sp. nov. *Biol. Zh. Armen.* 25:26
70. Maymó-Gatell X, Chien Y-t, Gossett JM, Zinder SH. 1997. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* 276:1568–71
71. Miroshnichenko ML, Kostrikina NA, Chernyh NA, Pemenov NV, Tourova TP, et al. 2003. *Caldithrix abyssi* gen. nov., sp. nov., a nitrate-reducing, thermophilic, anaerobic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent, represents a novel bacterial lineage. *Int. J. Syst. Evol. Microbiol.* 53:323–29
72. Moore LR, Post AF, Rocap G, Chisholm SW. 2002. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* 47:989–96
73. Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. London Ser. B* 253:167–71
74. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–29
75. Ohkuma M, Kudo T. 1996. Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl. Environ. Microbiol.* 62:461–68
76. Olsen GJ, Lane DL, Giovannoni SJ, Pace NR. 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* 40:337–65
77. Oyaizu H, Debrunner-Vossbrinck B, Mandelco L, Studier JA, Woese CR. 1987. The Green Non-Sulfur Bacteria: a deep branching in the eubacterial line of

descent. *System. Appl. Microbiol.* 9:47–53
78. Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–40
79. Pace NR, Stahl DA, Lane DJ, Olsen GJ. 1986. The analysis of natural microbial populations by ribosomal RNA sequences. *Adv. Microb. Ecol.* 9:1–55
80. Parpais J, Marie D, Partensky F, Morin P, Vaulot D. 1996. Effect of phosphorus starvation on the cell cycle of the photosynthetic prokaryote *Prochlorococcus* spp. *Mar. Ecol. Prog. Ser.* 132:265–74
81. Partensky F, Hess WR, Vaulot D. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63:106–27
82. Paster BJ, Ludwig W, Weisburg WG, Stackebrandt E, Hespell RB, et al. 1985. A phylogenetic grouping of the Bacteroides, Cytophagas, and certain Flavobacteria. *Syst. Appl. Microbiol.* 6:34–42
83. Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* 8:616–23
84. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418:630–33
85. Rappé MS, Vergin K, Giovannoni SJ. 2000. Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol. Ecol.* 33:219–32
86. Reysenbach A-L, Longnecker K, Kirshtein J. 2000. Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl. Environ. Microbiol.* 66:3798–806
87. Schlesner H, Stackebrandt E. 1986. Assignment of the genera *Planctomyces* and *Pirella* to a new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. *Syst. Appl. Microbiol.* 8:174–76
88. Schmidt JM. 1978. Isolation and ultrastructure of freshwater strains of *Planctomyces*. *Curr. Microbiol.* 1:65–71
89. Sekiguchi Y, Takahashi H, Kamagata Y, Ohashi A, Harada H. 2001. In situ detection, isolation, and physiological properties of a thin filamentous microorganism abundant in methanogenic granular sludges: a novel isolate affiliated with a clone cluster, the green non-sulfur bacteria subdivision I. *Appl. Environ. Microbiol.* 67:5740–49
90. Snel B, Bork P, Huynen MA. 1999. Genome phylogeny based on gene content. *Nat. Genet.* 21:108–10
91. Stackebrandt E, Liesack W, Göebel BM. 1993. Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rRNA analysis. *FASEB J.* 7:232–36
92. Stackebrandt E, Murray RGE, Truper HG. 1988. Proteobacteria classis nov., a name for the phylogenetic taxon that includes the purple bacteria and their relatives. *Int. J. Syst. Bacteriol.* 38:321–25
93. Stackebrandt E, Rainey FA, Ward-Rainey NL. 1997. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int. J. Syst. Bacteriol.* 47:479–91
94. Stackebrandt E, Woese CR. 1981. The evolution of prokaryotes. In *Molecular and Cellular Aspects of Microbial Evolution*, ed. MJ Carlile, JR Collins, BEB Mosely, 32:1–31. Cambridge: Cambridge Univ. Press
95. Staley JT. 1973. Budding bacteria of the *Pasteuria-Blastobacter* group. *Can. J. Microbiol.* 19:609–14
95a. Staley JT, Reysenbach A-L, eds. 2002. *Biodiversity of Microbial Life: Foundations of Earth's Biosphere*. New York: Wiley-Liss
96. Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, et al. 1999. Missing lithotroph identified as new planctomycete. *Nature* 400:446–49
97. Suzuki MT, Béjà O, Taylor LT, DeLong

EF. 2001. Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ. Microbiol.* 3:323–31

98. Swofford D. 2000. *PAUP*. Phylogenetic Analysis Using Parsimony* (**and Other Methods*). Sunderland, MA: Sinauer
99. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–37
100. Teske A, Hinrichs KU, Edgcomb V, Gomez AdV, Sylva SP, et al. 2002. Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotropic communities. *Appl. Environ. Microbiol.* 68:1994–2007
101. Ting CS, Rocap G, King J, Chisholm SW. 2002. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol.* 10:134–41
102. Todorov JR, Chistoserdov AY, Aller JY. 2000. Molecular analysis of microbial communities in mobile deltaic muds of Southeastern Papua New Guinea. *FEMS Microbiol. Ecol.* 33:147–55
103. Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW. 1998. Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J. Mol. Evol.* 46:188–201
104. Urbach E, Vergin KL, Young L, Morse A, Giovannoni SJ. 2001. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. *Limnol. Oceanogr.* 46:557–72
105. Wang GC-Y, Wang Y. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* 63:4645–50
106. Ward DM. 1998. A natural species concept for prokaryotes. *Curr. Opin. Microbiol.* 1:271–77
107. Ward DM, Bateson MM, Weller R, Ruff-Roberts AL. 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. *Adv. Microb. Ecol.* 12:219–86
108. Ward DM, Papke T, Nübel U, McKitrick MC. 2002. Natural history of microorganisms inhabiting hot spring microbial mat communities: clues to the origin of microbial diversity and implications for microbiology and macrobiology. See Ref. 95a, pp. 25–48
109. Ward-Rainey N, Rainey FA, Schlesner H, Stackebrandt E. 1995. Assignment of hitherto unidentified 16S rDNA species to a main line of descent within the domain *Bacteria*. *Microbiology* 141:3247–50
110. Weller R, Bateson MM, Heimbuch BK, Kopczynski ED, Ward DM. 1992. Uncultivated cyanobacteria, *Chloroflexus*-like inhabitants, and spirochete-like inhabitants of a hot spring microbial mat. *Appl. Environ. Microbiol.* 58:3964–69
111. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* 95:6578–83
112. Woese CR. 1987. Bacterial evolution. *Microbiol. Rev.* 51:221–71
113. Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87:4576–79
114. Woese CR, Stackebrandt E, Macke TJ, Fox GE. 1985. A phylogenetic definition of the major eubacterial taxa. *Syst. Appl. Microbiol.* 6:143–51
115. Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV. 2001. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.* 1:8
116. Yap WH, Zhang Z, Wang Y. 1999. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chomogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* 181:5201–9
117. Zengler K, Toledo G, Rappé M, Elkins J, Mathur EJ, et al. 2002. Cultivating the

uncultured. *Proc. Natl. Acad. Sci. USA* 99:15681–86

118. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, et al. 2003. *Gemmatimonas aurantiaca* gen. nov., sp. nov., a Gram-negative aerobic polyphosphate-accumulating microorganism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phy. nov. *Int. J. Syst. Evol. Microbiol.* In press
119. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han S-K. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* 28:141–55