

A Molecular View of Microbial Diversity and the Biosphere

Norman R. Pace

Over three decades of molecular-phylogenetic studies, researchers have compiled an increasingly robust map of evolutionary diversification showing that the main diversity of life is microbial, distributed among three primary relatedness groups or domains: Archaea, Bacteria, and Eucarya. The general properties of representatives of the three domains indicate that the earliest life was based on inorganic nutrition and that photosynthesis and use of organic compounds for carbon and energy metabolism came comparatively later. The application of molecular-phylogenetic methods to study natural microbial ecosystems without the traditional requirement for cultivation has resulted in the discovery of many unexpected evolutionary lineages; members of some of these lineages are only distantly related to known organisms but are sufficiently abundant that they are likely to have impact on the chemistry of the biosphere.

Microbial organisms occupy a peculiar place in the human view of life. Microbes receive little attention in our general texts of biology. They are largely ignored by most professional biologists and are virtually unknown to the public except in the contexts of disease and rot. Yet, the workings of the biosphere depend absolutely on the activities of the microbial world (1). Our texts articulate biodiversity in terms of large organisms: insects usually top the count of species. Yet, if we squeeze out any one of these insects and examine its contents under the microscope, we find hundreds or thousands of distinct microbial species. A handful of soil contains billions of microbial organisms, so many different types that accurate numbers remain unknown. At most only a few of these microbes would be known to us; only about 5000 noneukaryotic organisms have been formally described (2) (in contrast to the half-million described insect species). We know so little about microbial biology, despite it being a part of biology that looms so large in the sustenance of life on this planet.

The reason for our poor understanding of the microbial world lies, of course, in the fact that microbes are tiny, individually invisible to the eye. The mere existence of microbial life was recognized only relatively recently in history, about 300 years ago, with Leeuwenhoek's invention of the microscope. Even under the microscope, however, the simple morphologies of most microbes, usually nondescript rods and spheres, prevented their classification by morphology, the way that large organisms had always been related to one another. It

was not until the late 19th century and the development of pure-culture techniques that microbial organisms could be studied as individual types and characterized to some extent, mainly by nutritional criteria. However, the pure-culture approach to the study of the microbial world seriously constrained the view of microbial diversity because most microbes defy cultivation by standard methods. Moreover, the morphological and nutritional criteria used to describe microbes failed to provide a natural taxonomy, ordered according to evolutionary relationships. Molecular tools and perspective based on gene sequences are now alleviating these constraints to some extent. Even the early results are changing our perception of microbial diversity.

A Sequence-Based Map of Biodiversity

Before the development of sequence-based methods, it was impossible to know the evolutionary relationships connecting all of life and thereby to draw a universal evolutionary tree. Whittaker, in 1969, just as the molecular methods began to develop, summarized evolutionary thought in the context of the "Five Kingdoms" of life: animals, plants, fungi, protists ("protozoa"), and monera (bacteria) (3). There also was recognized a higher, seemingly more fundamental taxonomic distinction between eukaryotes, organisms that contain nuclear membranes, and prokaryotes, predecessors of eukaryotes that lack nuclear membranes (4). These two categories of organisms were considered independent and coherent relatedness groups. The main evolutionary diversity of life on Earth, four of the five traditional taxonomic kingdoms, was thought to lie among the eukaryotes, par-

ticularly the multicellular forms.

The breakthrough that called to question many previous beliefs and brought order to microbial, indeed biological, diversity emerged with the determination of molecular sequences and the concept that sequences could be used to relate organisms (5). The incisive formulation was reached by Carl Woese who, by comparison of ribosomal RNA (rRNA) sequences, established a molecular sequence-based phylogenetic tree that could be used to relate all organisms and reconstruct the history of life (6, 7). Woese articulated the now-recognized three primary lines of evolutionary descent, termed "urkingdoms" or "domains": Eucarya (eukaryotes), Bacteria (initially called eubacteria), and Archaea (initially called archaebacteria) (8).

Figure 1 is a current phylogenetic tree based on small-subunit (SSU) rRNA sequences of the organisms represented. The construction of such a tree is conceptually simple (9). Pairs of rRNA sequences from different organisms are aligned, and the differences are counted and considered to be some measure of "evolutionary distance" between the organisms. There is no consideration of the passage of time, only of change in nucleotide sequence. Pair-wise differences between many organisms can then be used to infer phylogenetic trees, maps that represent the evolutionary paths leading to the modern-day sequences. The tree in Fig. 1 is largely congruent with trees made using any molecule in the nucleic acid-based, information-processing system of cells. On the other hand, phylogenetic trees based on metabolic genes, those involved in the manipulation of small molecules and in interaction with the environment, commonly do not concur with the rRNA-based version [see (10, 11) for reviews and discussions of phylogenetic results with different molecules]. Incongruities in phylogenetic trees made with different molecules may reflect lateral transfers or even the intermixings of genomes in the course of evolution. Some metabolic archaeal genes, for instance, appear much more highly related to specific bacterial versions than to their eucaryal homologs; other archaeal genes seem decidedly eukaryotic in nature; still other archaeal genes are unique. Nonetheless, the recently determined sequence of the archaeon *Methanococcus jannaschii* shows that the evolutionary lineage Archaea is independent of both Eucarya and Bacteria (12).

Interpreting the Molecular Tree of Life

"Evolutionary distance" in this type of phylogenetic tree (Fig. 1), the extent of sequence change, is read along line segments.

The author is in the Departments of Plant and Microbial Biology and Molecular and Cell Biology, University of California, Berkeley, CA 94720-3102, USA. E-mail: nrpace@nature.berkeley.edu



The tree can be considered a rough map of the evolution of the genetic core of the cellular lineages that led to the modern organisms (sequences) included in the tree. The time of occurrence of evolutionary events cannot be extracted reliably from phylogenetic trees, despite common attempts to do so. Time cannot be accurately correlated with sequence change because the evolutionary clock is not constant in different lineages (7). This disparity is evidenced in Fig. 1 by the fact that lines leading to the different reference organisms are not all the same length; these different lineages have experienced different extents of sequence change. Nonetheless, the order of occurrence of branchings in the trees can be interpreted as a genealogy, and intriguing insights into the evolution of cells are emerging.

A sobering aspect of large-scale phylogenetic trees such as that shown in Fig. 1 is the graphical realization that most of our legacy in biological science, historically based on large organisms, has focused on a narrow slice of biological diversity. Thus, we see that animals (represented in Fig. 1 by *Homo*), plants (*Zea*), and fungi (*Coprinus*) constitute small and peripheral branches of even eukaryotic cellular diversity. If the animals, plants, and fungi are taken to comprise taxonomic "kingdoms," then we must recognize as kingdoms at least a dozen other eucaryotic groups, all microbial, with as much or more independent evolutionary history than that which separates the three traditional eukaryotic kingdoms (13).

The rRNA and other molecular data solidly confirm the notion stemming from the last century that the major organelles of eukaryotes—mitochondria and chloroplasts—are derived from bacterial symbionts that have undergone specialization through coevolution with the host cell. Sequence comparisons establish mitochondria as representatives of Proteobacteria (the group in Fig. 1 including *Escherichia* and *Agrobacterium*) and chloroplasts as derived from cyanobacteria (*Synechococcus* and *Gloeobacter* in Fig. 1) (14). Thus, all respiratory and photosynthetic capacity of eukaryotic cells was obtained from bacterial symbionts; the "endosymbiont hypothesis" for the origin of organelles is no longer hypothesis but well-grounded fact. The nuclear component of the modern eukaryotic cell did not derive from one of the prokaryotic lineages, however. The rRNA and other molecular trees show that the eukaryotic nuclear line of descent extends as deeply into the history of life as do the bacterial and archaeal lineages. The mitochondrion and chloroplast came in relatively late. This late evolution is evidenced by the fact that mitochondria and chloroplasts diverged

from free-living organisms that branch peripherally in molecular trees. Moreover, the most deeply divergent eukaryotes even lack mitochondria (15). These latter organisms, little studied but sometimes troublesome creatures such as *Giardia*, *Trichomonas*, and *Vairimorpha*, nonetheless contain at least a few bacterial-type genes (16). These genes may be evidence of an earlier mitochondrial symbiosis with Eucarya that was lost (11) or perhaps other symbiotic or gene-transfer events between the evolutionary domains.

The root of the universal tree in Fig. 1, the point of origin of the modern lineages, cannot be established using sequences of only one type of molecule. However, recent phylogenetic studies of gene families that originated before the last common ancestor of the three domains have positioned the root of the universal tree deep on the bacterial line (10). Therefore, Eucarya and Archaea had a common history that excluded the descendants of the bacterial line. This period of evolutionary history shared by Eucarya and Archaea was an important time in the evolution of cells, during which the refinement of the primordial information-processing mechanisms occurred. Thus, modern representatives of Eucarya and Archaea share many properties that differ from bacterial cells in fundamental ways. One ex-

ample of similarities and differences is in the nature of the transcription machinery. The RNA polymerases of Eucarya and Archaea resemble each other in subunit composition and sequence far more than either resembles the bacterial type of polymerase. Moreover, whereas all bacterial cells use sigma factors to regulate the initiation of transcription, eucaryal and archaeal cells use TATA-binding proteins (17, 18).

Because of the shared history of Eucarya and Archaea, we should, perhaps, look to the Archaea to identify fundamental properties of far more complex cells such as our own. The eukaryotic nuclear membrane, for instance, is considered by cell biologists to be an intrinsic component of the nucleus, somehow responsible for its integrity. The fact that Archaea remained "prokaryotic," that is, did not develop a nuclear membrane, indicates that a membrane is not required for nuclear function, which Archaea certainly achieve (as do Bacteria, for that matter). Indeed, the archaeal nuclear zone even seems to exclude ribosomes (19), and the genome of *M. jannaschii* is sprinkled with homologs of eucaryal nuclear and nucleolar structural genes (12). What constitutes a "nucleus?" Certainly the acquisition of the nuclear membrane was a relatively late event in the establishment of the eu-

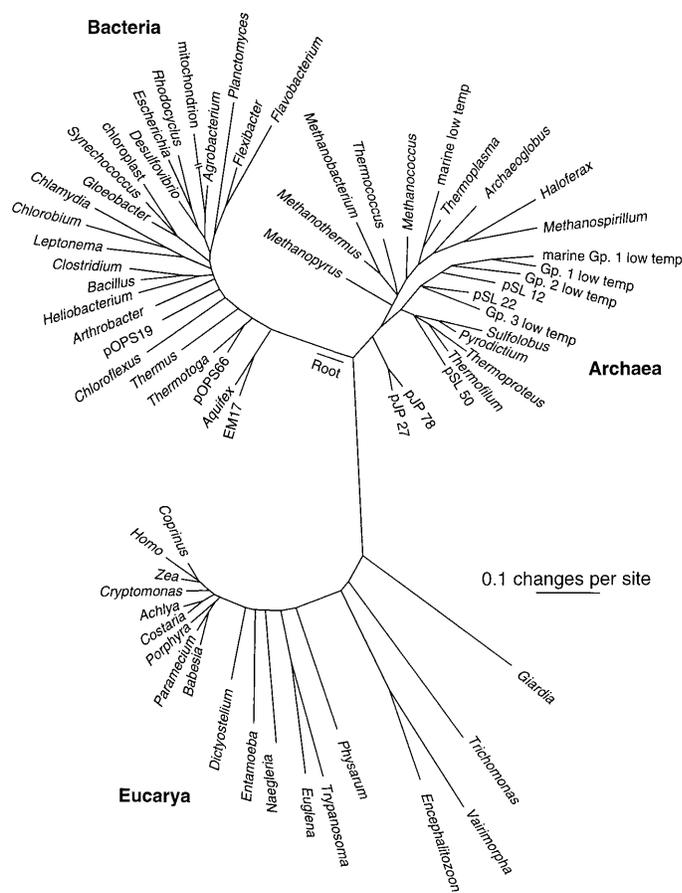


Fig. 1. Universal phylogenetic tree based on SSU rRNA sequences. Sixty-four rRNA sequences representative of all known phylogenetic domains were aligned, and a tree was produced using FASTD-NAML (43, 52). That tree was modified, resulting in the composite one shown, by trimming lineages and adjusting branch points to incorporate results of other analyses. The scale bar corresponds to 0.1 changes per nucleotide.

caryal line of descent, occurring only after the separation from Archaea. Perhaps the nuclear membrane is after all not fundamental to the function of the nucleus but rather is a relatively late-arriving embellishment. One hypothesis would be that the nuclear membrane was an invention derived from the Golgi apparatus to serve as a gathering basket for nuclear products, for distribution by the Golgi throughout the cell. The properties of nuclear pores would be consistent with this hypothesis; they are large orifices, typically >10 nm in diameter, unlikely to gate anything except large molecules (20). The evolutionary record suggests, then, that we look to something more fundamental than the nuclear membrane for the integrity of the nucleus and by which to define the essential quality of the eukaryotic cell. The shared evolutionary history of Eucarya and Archaea suggests that we may be able to recognize the most fundamental elements of our own nucleus through study of the archaeal version.

The Metabolic Diversity of Life

The molecular-phylogenetic perspective (Fig. 1) is a reference framework within which to describe microbial diversity; the sequences of genes can be used to identify organisms. This capability is an important concept for microbial biology. It is not possible to describe microorganisms as traditionally done with large organisms, through their morphological properties. To be sure, some microbes are intricate and beautiful in the microscope, but they are mainly relatively unfeared at the resolution of routine microscopy. Therefore, in order to distinguish different types of microbes, microbiologists early turned to metabolic properties such as utilizable sources of nutrition, for instance, sources of carbon, nitrogen, and energy. Microbial taxonomy accumulated as anecdotal descriptions of metabolically and morphologically distinct types of organisms that were essentially unrelatable. Molecular phylogeny now provides a framework within which we can relate organisms objectively, and also through which we can interpret the evolutionary flow of the metabolic machineries that constitute microbial diversity.

Laboratory studies of microbial metabolism have focused mainly on organisms such as *Escherichia coli* and *Bacillus subtilis*. In the broad sense, such organisms metabolize much as animals do; we are all “organotrophs,” using reduced organic compounds for energy and carbon. Organotrophy is not the prevalent form of metabolism in the environment, however. Autotrophic metabolism, fixation of CO₂ to reduced organic compounds, must necessarily contribute to a greater biomass than organotrophic metabo-

lism, which it supports (a principle long appreciated by ecologists). Energy for fixing CO₂ is gathered in two ways: by phototrophy (photosynthesis) or lithotrophy (coupling the oxidation of reduced inorganic compounds such as hydrogen, hydrogen sulfide, or ferrous iron to the reduction of a chemical oxidant, a terminal electron acceptor such as oxygen, nitrate, sulfate, sulfur, or carbon dioxide). Thus, metabolic diversity can be generalized in terms of organotroph or autotroph, phototroph or lithotroph, and the nature of the electron donor and acceptor.

The phylogenetic distributions of different types of carbon and energy metabolism among different organisms do not necessarily follow the evolutionary pattern of rRNA (Fig. 1). Presumably, this lack of correspondence is because of past lateral transfers of those metabolic genes and larger scale symbiotic fusions. Nonetheless, there are domain-level tendencies that may speak to the ancestral nature of the three domains of life (21). The perspective here is currently limited mainly to Archaea and Bacteria. Such broad generalities cannot yet be assessed for the Eucarya because so little is known about the metabolic breadth of the domain, the properties of the most deeply divergent lineages. There is considerable information about one pole of eukaryotic diversity, that represented by animals, plants, and fungi. We know little about the other pole, the amitochondriate organisms that spun off of the main eucaryal line early in evolution (22). The known instances of such lineages, represented by *Trichomonas*, *Giardia*, and *Vairimorpha* in Fig. 1, are primarily pathogens. Pathogenicity to humans is a rare trait among the rest of eucaryotes and bacteria, and no archaeal pathogen is known. This correlation may indicate that nonpathogenic, deeply divergent eucaryotes are abundant in the environment but not yet detected. They should be sought in anaerobic ecosystems, possibly coupled metabolically to other organisms. A driving theme of the eucaryal line seems to be the establishment of physical symbiosis with other organisms. Beyond that, the general metabolism of the rudimentary eukaryotic cell seems simple, based on fermentative organotrophy. By virtue of symbiotic partners, however, eucaryotes are able to take on phototrophic or lithotrophic life-styles and to use the electron-acceptor oxygen (23).

Symbiotic microbes commonly confer the lithotrophic way of life even on animals, although this was only recently recognized. The 2-m-long tubeworm *Riftia pachyptila*, for instance, lives in the vicinity of sea-floor hydrothermal vents and metabolizes hydrogen sulfide and carbon dioxide by means of sulfide-oxidizing, carbon diox-

ide-fixing bacterial symbionts (24). This invertebrate and metabolically similar ones may contribute significantly to primary productivity in the ocean (25). It is not necessary to go to unusual (from our perspective) places such as ocean-floor vents to encounter other equally fascinating hydrogen sulfide-dependent eucaryotes (26). Underfoot at the ocean beach, for example, microbial respiration of seawater sulfate creates a hydrogen sulfide-rich ecosystem populated by little-known creatures such as *Kentrophoros*, a flat, gulletless ciliate that under the microscope appears fuzzy because it cultivates on its outer surface a crop of sulfide-oxidizing bacteria (27). These bacteria are ingested by endocytosis and thereby provide nutrition for *Kentrophoros*. In other anaerobic environments, methanogens, members of Archaea, live intracellularly with eucaryotes and serve as metabolic hydrogen sinks (28). Still other symbioses based on inorganic energy sources are all around us and are little explored for their diversity of microbial life (26).

Many lithotrophic, but comparatively few organotrophic, representatives of Archaea have been obtained in pure culture (29). There are primarily two metabolic themes, both relying on hydrogen as a main source of energy. Among the known Euryarchaeota—one of the two archaeal kingdoms known through cultivated organisms—the main electron acceptor is carbon dioxide, and the product, methane—“natural gas.” Most of the methane encountered in the outer few kilometers of Earth’s crust or on the surface is determined by isotopic analysis to be the product of methanogenic Archaea, past and present. Such organisms probably constitute a large component of global biomass. They certainly offer an inexhaustible source of renewable energy to humankind.

The general metabolic theme of the other established kingdom of Archaea, Crenarchaeota, is also the oxidation of molecular hydrogen, but with a sulfur compound as the terminal electron acceptor. All of the cultivated representatives of Crenarchaeota also are thermophiles. Consequently, such organisms have been referred to as “thermoacidophilic” or “hyperthermophilic” Archaea; some grow at the highest known temperatures for life, up to 113°C in the case of *Pyrolobus fumaris* (30). These crenarchaeotes might seem bizarre, capable of thriving at temperatures above the usual boiling point of water on a diet of H₂, CO₂, and elemental sulfur and exhaling hydrogen sulfide. Yet, in terms of the molecular structures of the basic cellular machineries, these creatures resemble eucaryotes far more closely than either resembles the bacterium *E. coli* (17).

The metabolic diversity of microbes is



usually couched in terms of the utilization of complex organic compounds. From that standpoint, metabolic diversity seems, on the basis of cultivated instances of organisms, to have flowered mainly among the Bacteria. Even here, however, reliance on organic nutrients probably was not ancestral. The most deeply branching of the cultured bacterial lineages, represented by *Aquifex* and *Thermotoga* in Fig. 1, are basically lithotrophs that use hydrogen as an energy source and electron acceptors such as sulfur compounds (*Thermotoga*) or low levels of oxygen (*Aquifex*) (31). Cultivated instances of these deeply branching bacterial lineages also are all thermophilic and thus share two important physiological attributes with the deeply branching and slowly evolving Archaea: a hydrogen-based energy source and growth at high temperatures. This coincidence suggests that the last common ancestor of all life also metabolized hydrogen for energy at high temperatures. This inference is consistent with current notions regarding the origin of life, that it came to be in a geothermal setting at high temperature (32).

Chlorophyll-based photosynthesis was a bacterial invention. It seems to have appeared well after the establishment of the bacterial line of descent, at or before the divergence of the line in Fig. 1 leading to *Chloroflexus*, a photosynthetic genus (33), and after the deeper divergences such as those leading to *Aquifex* and *Thermotoga*, genera that are not known to have photosynthetic representatives. Most bacterial photosynthesis is anaerobic, however. Oxygenic photosynthesis, the water-based photosynthetic mechanism that produces the powerful electron acceptor oxygen, arose only in the kingdom-level lineage of cyanobacteria. This invention changed the surface of Earth profoundly and conventionally is thought to be the basis, directly or indirectly, of most present-day biomass.

Anaerobic photosynthesis is widely distributed in the late-branching bacterial kingdoms. The more ancient theme of lithotrophy, metabolism of inorganic compounds, is also widely distributed phylogenetically, intermixed with organotrophic organisms. The pattern suggests that organotrophy arose many times from otherwise photosynthetic or lithotrophic organisms. Indeed, many instances of Bacteria can switch between these modes of nutrition, carrying out photosynthesis in the light and lithotrophy or organotrophy in the dark. Particularly among Bacteria, this type of energy metabolism seems highly volatile in evolution: Bacteria that are closely related by molecular criteria can display strikingly different phenotypes when assessed in the laboratory through the nature of their car-

bon and energy metabolism. In the relatively closely related "gamma subgroup" of the kingdom of Proteobacteria (delineated by the genus *Escherichia* in Fig. 1), for instance, we find the phenotypically disparate organisms *E. coli* (organotroph), *Chromatium vinosum* (hydrogen sulfide-based phototroph), and the symbiont of the tubeworm *R. pachyptila* (hydrogen sulfide-based symbiont). The superficial metabolic diversity of these types of Bacteria belies their underlying close evolutionary relatedness, giving no hint of the close similarities of their basic machineries. The versatility of Bacteria makes the metabolic machineries of Archaea and Eucarya seem comparatively more monotonous. As the sequences of diverse genomes are compared, it will be possible to map the flow of metabolic genes onto the rRNA-based tree and thereby see how metabolic diversity has been molded through evolution.

The molecular perspective gives us more than just a glimpse of the evolutionary past; it also brings a new future to the discipline of microbial biology. Because the molecular-phylogenetic identifications are based on sequence, as opposed to metabolic properties, microbes can be identified without being cultivated. Consequently, all the sequence-based techniques of molecular biology can be applied to the study of natural microbial ecosystems, heretofore little known with regard to organismal makeup.

A Sequence-Based Glimpse of Biodiversity in the Environment

Knowledge of microorganisms in the environment has depended in the past mainly on studies of pure cultures in the laboratory. Rarely are microbes so captured, however. Studies of several types of environments estimate that more than 99% of organisms seen microscopically are not cultivated by routine techniques (34). With the sequence-based taxonomic framework of molecular trees, only a gene sequence, not a functioning cell, is required to identify the organism in terms of its phylogenetic type. The occurrence of phylogenetic types of organisms, "phylotypes," and their distributions in natural communities can be surveyed by sequencing rRNA genes obtained from DNA isolated directly from the environment. Analysis of microbial ecosystems in this way is more than a taxonomic exercise because the sequences provide experimental tools—for instance, molecular hybridization probes—that can be used to identify, monitor, and study the microbial inhabitants of natural ecosystems (35).

Ribosomal RNA genes are obtained by cloning DNA isolated directly from the environment. "Shotgun libraries" of random

DNA fragments are a source of rRNA, as well as other genes, but require sorting of rRNA genes from the others. The quickest way to survey the constituents of microbial ecosystems is through the use of the polymerase chain reaction (PCR) (36). The highly conserved nature of rRNA allows for the synthesis of "universal" PCR primers that can anneal to sequences conserved in the rRNA genes from all three phylogenetic domains. In principle, PCR carried out with these primers amplifies the rRNA genes of all types of organisms present in an environmental sample. Individual types of genes in the mixture are separated by a cloning step and then sequenced.

A molecular-phylogenetic assessment of an uncultivated organism can provide insight into many of the properties of the organism through comparison with its studied relatives. One example of the perspective that phylogeny can offer on an otherwise unknown organism is seen with the sulfur-oxidizing microorganisms that provide nutrition to symbiotic invertebrates such as the vent tubeworm *R. pachyptila* (24). Although many attempts to cultivate the symbionts for phenotypic characterization failed, rRNA analyses revealed that many of the basic cellular properties of the symbionts were already familiar to us. The *Riftia* symbiont and a number of other sulfur-oxidizing symbionts associated with invertebrate animals all proved to be fairly closely related to one another, close relatives also to the intensively studied organisms *E. coli* and *Pseudomonas aeruginosa* (37). Because of their phylogenetic proximity, many of the properties of the symbionts can be inferred from those of the well-studied organisms. For instance, we can predict with good confidence the nature of the ribosome and antibiotic-susceptibility patterns, the nature of the DNA-replicative machinery, the character of the RNA polymerase complex, the character of biosynthetic pathways and their regulatory mechanisms, the nature of the cell envelope and energy transduction schemes, and many other cellular properties of the symbionts. On the other hand, because *E. coli* and *P. aeruginosa* do not oxidize sulfur, these relations cannot provide insight into the sulfur-oxidative pathways of the symbionts. The rRNA sequence does, however, identify free-living and cultivated (but less-studied) close relatives of the symbionts—for example *Thiomicrospira* sp. L-12 (a hydrothermal-vent isolate) and *Thiothrix* sp.—that also rely on sulfur oxidation and so are likely to provide good models for this process in the symbionts.

Every nucleic acids-based study of natural microbial ecosystems so far performed has uncovered novel types of rRNA se-

quences, often representing major new lineages only distantly related to known ones. The discovery of rRNA sequences in the environment that diverge more deeply in phylogenetic trees than those of cultivated organisms is particularly noteworthy. It means that the divergent organisms recognized by rRNA sequence are potentially more different from known organisms in the lineage than the known organisms are from one another. The deepest divergences in both the Bacteria and Archaea were first discovered in rRNA-based surveys of hot spring-associated communities in Yellowstone National Park.

The geothermal features of Yellowstone National Park have been favorite haunts of high-temperature biologists for decades (38). Currently, rRNA-based methods are being used to survey phylogenotypes present in a number of Yellowstone hot springs with disparate chemical settings. One of these, Octopus Spring (Fig. 2A)—a near-boiling, slightly alkaline, extremely low-nutrient flow near Old Faithful geyser containing an abundant community of pink filaments—yielded the first evidence for the lineage currently thought to be the most deeply divergent in the Bacteria. When this lineage, represented by *Aquifex* and EM17 (pink filament clone) (39) in Fig. 1, was first encountered by 5S rRNA sequence (40), little could be inferred about the physiology of the associated organism because no cultivated specific relative had yet been described. Subsequent clues to the nature of the pink filaments came with the discovery of *A. pyrophilus*, cultured from an Icelandic hot spring with a chemical character similar to that of Octopus Spring (41), and determination of the 16S rRNA sequence for the pink filaments (39). The *A. pyrophilus* and pink filament sequences are sufficiently closely related (Fig. 1) that many of their properties are likely to be shared. The mode of nutrition of the pink filaments, for instance, is predicted to be that of *Aquifex*, consumption of hydrogen with low levels of oxygen and fixation of carbon dioxide. Many other representatives of the *Aquifex*-EM17 relatedness group (*Aquificales*) have now been cultured, mainly from high-temperature settings, and all are thermophilic hydrogen oxidizers (31).

Hot springs on the northern flank of the Yellowstone caldera usually have high concentrations of iron (II), hydrogen sulfide, hydrogen, and carbon dioxide—a wealth of foodstuffs for compatible physiologies. Ongoing sequence-based studies of the microbial inhabitants of one of these springs, Obsidian Pool (Fig. 2B), have radically revised our view of the phy-

logenetic diversity of Archaea. All cultivated Crenarchaeota branch in the cluster bracketed by *Pyrodictium* and *Thermofilum* in Fig. 1. Discovery of a rich abundance of diverse crenarchaeal rRNA genes in Obsidian Pool sediment (for example, pSL sequences in Fig. 1), scores of new genera, expanded the known phylogenetic diversity (estimated by specific line-segment lengths) of Crenarchaeota severalfold (42). More surprising, other sequences from Obsidian Pool (pJP27 and pJP28 in Fig. 1) seem to branch so deeply in the overall archaeal tree that they constitute a new kingdom-level branch of Archaea, recognized provisionally as “Korarchaeota” (43). It now will be interesting to study other genes from these novel organisms. These genes, as well as information on the physiology and other properties of the organisms, will be obtained most readily if they can be cultured. Even without cultivation, however, cloning large fragments of environmental DNA and then “chromosome walking” to assemble contiguous clones offers access to the genomes of these or other uncultivated organisms (44).

Continuing study of Obsidian Pool is expanding the known extent of bacterial, as well as archaeal, diversity. Obsidian Pool, judged extremely inhospitable from the human standpoint, contains a rich diversity of sequence types representing most of the known bacterial kingdoms, as well as kingdom-level divergences never described by cultivation (45). Phylogenetic studies of cultured and environmental sequences have expanded substantially our appreciation of the scope of bacterial diversity: In 1987, only about 12 phylogenetic kingdoms (main phyla) of Bacteria were recognized (Fig. 3, inset) (7), but now, at least 25 to 30 distinct, kingdom-level phylogenetic divergences are resolved (Fig. 3). The topology of the bacterial tree is remarkable. Bacterial diversity seems to have arisen mainly from an

explosive radiation of lineages, rather than from the sequential divergence of main lines seen, for instance, in the eucaryal domain (Fig. 1). Preliminary results from Obsidian Pool also call into question another supposition based on culture studies, that Archaea dominate high-temperature environments. Quantitative hybridization of domain-specific oligonucleotide probes to rRNA genes obtained by PCR indicates that bacterial genes outnumber archaeal genes by 50:1 in this environment. Such conclusions, of course, are compromised to an unknown extent by considerations such as nonuniform amplification of different rRNA genes, but the trend seems to indicate that bacteria dominate this environment.

It is not necessary to go to extreme environments to encounter exotic diversity; it is all around us. Phylogenotypes that, because of their abundance, must be significant contributors to the biosphere have escaped detection until the sequence-based methods developed. One example of an arena for research opened by the molecular methods involves the recently discovered mesophilic (low-temperature) Crenarchaeota (represented in Fig. 1 by pGrf and marSBAR). On the basis of culture-studies, crenarchaeotes had been thought to be restricted to high-temperature environments. Cloned rRNA gene analysis shows, however, that low-temperature versions of Crenarchaeota are abundant globally in marine (19, 46) and terrestrial (47) environments, in typically 30 to 50% of planktonic rRNA genes in limited samplings of Atlantic, Pacific, and Antarctic waters (48). The physiologies of the low-temperature crenarchaeotes are unknown; none has yet been cultivated. The properties of their remote relatives—the cultivated, high-temperature Crenarchaeota—hint that the mesophilic types might also engage in hydrogen metabolism, perhaps using some oxidation state of sulfur as an electron acceptor.

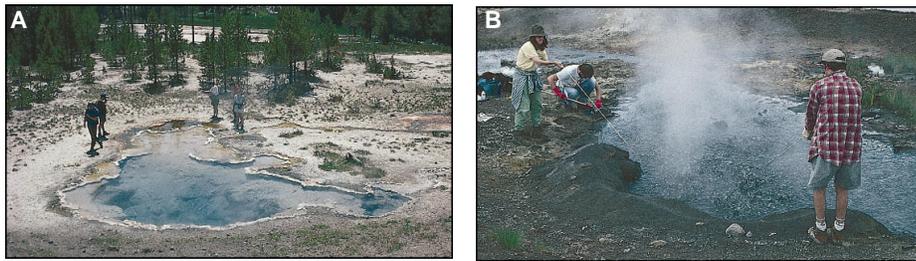


Fig. 2. Yellowstone National Park hot springs rich in microbial diversity. **(A)** Octopus Spring. The source pool of this hot spring is 90° to 93°C and extremely low in nutrients but contains abundant biomass and the deepest known evolutionary divergences in the domain Bacteria. **(B)** Obsidian Pool. Molecular studies find that the inhospitable environment of this hot spring, 75° to 95°C in temperature and containing high concentrations of iron (II) and hydrogen sulfide, supports an extensive diversity of previously unknown microbial life, both archaeal and bacterial.

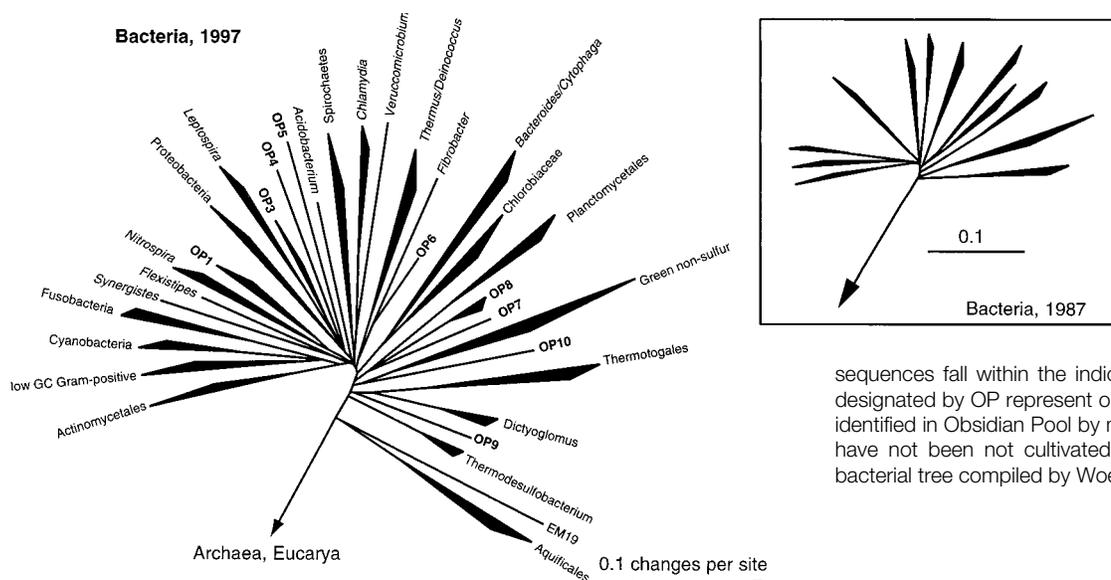


Fig. 3. Diagrammatic representation of the known phylogenetic span of Bacteria in 1987 (inset) and today. Phylogenetic trees containing sequences from the indicated organisms or groups of organisms, chosen to represent the broadest diversity of Bacteria, were used as the basis for this diagram (compiled with P. Hugenholtz). Filled sectors indicate that several representative

sequences fall within the indicated depth of branching. Lines designated by OP represent one or more phylotypes that were identified in Obsidian Pool by means of molecular methods but have not been cultivated. The inset is an outline of the bacterial tree compiled by Woese in 1987 (7).

Microbial Diversity and the Limits of the Biosphere

Textbooks generally portray only a part of the global distribution of life, the part that is immediately dependent on either the harvesting of sunlight or the metabolism of the decay products of photosynthesis. The molecular phylogenetic record shows, however, that lithotrophic metabolism preceded and is more widespread phylogenetically and geographically than is either phototrophy or organotrophy. The lithotrophic biosphere potentially extends kilometers into the crust of Earth, an essentially unknown realm (49). These considerations may indicate that lithotrophy contributes far more to the biomass of Earth than currently thought.

Part of that lithotrophic biomass is in microhabitats all around us, usually away from light and oxygen. It is not necessary to look far to find such environments: the rumens of cattle and the guts of termites and humans, for example, are significant sources of methane, a signature of hydrogen metabolism. Most life that depends on inorganic energy metabolism, however, probably is in little-known environments, based on poorly understood geochemistries. The oceans, for instance, cover 70% of Earth's surface to an average depth of 4 km. Most life in the ocean is microbial, and the metabolic patterns of such organisms are not understood: Large standing crops of low-temperature crenarchaeotes, potential hydrogen oxidizers, may indicate an unsuspected, lithotrophy-based food chain in the oceans. Another little-studied environment with global significance is the deep subsurface (50). There is increasing evidence that the crust of Earth is

shot through with biomass, wherever the physical conditions permit. Metabolism of hydrogen is a dominant theme among organisms isolated from geothermal settings or deep aquifers (51). Hydrogen is generated readily by abiotic mechanisms such as interaction of water with iron-bearing basalt, the main stuff of Earth's crust; consequently, a food source is unlikely to be limiting in most subterranean environments. Rather, it is likely to be the oxidant, the terminal electron acceptor, that limits growth. Nonetheless, it seems possible that much, perhaps most, of the biomass on Earth is subterranean, a biological world based on lithotrophy. Although the metabolic rate of this subterranean biosphere is likely to be far slower than in the more dynamic, photic environment, life is likely to be as pervasive in occurrence, and perhaps in cellular diversity, as we experience on the surface.

Opportunity for an Environmental Genome Survey

It is clear from even the small number of environments so far studied with the molecular methods that our understanding of the makeup of the natural microbial world is rudimentary. The sequence-based methods, however, now provide a way to survey biodiversity rapidly and comprehensively. Ribosomal RNA genes gathered from the environment are snapshots of organisms, representatives of different types of genomes, targets for further characterization if they seem interesting or useful. If we want to understand the biosphere, I think it important, even essential, that we undertake a representative survey of microbial diversity in the environment. A complete cataloging

of Earth's microbial biota is needless and, of course, impossible. A representative survey, however, is worthwhile. A representative survey could be achieved with modest effort, with the use of automated sequencing technology. Analysis of 1000 clones (to detect the most abundant genome types) from each of 100 chemically different environments would be comparable to the effort to sequence a single microbial genome. The questions are large and many: What kinds of organisms do we share this planet with and depend on? What model systems should we choose for laboratory studies of environmental processes? How extensive is the fund of biodiversity from which we can draw useful lessons and resources? Can we use the distribution of microbes as a biosensor array to map and monitor the chemistries of Planet Earth? Are there deeper branchings in the tree of life than the lineages we know?

The opportunities for the discovery of new organisms and the development of resources based on microbial diversity are greater than ever before. Molecular sequences have finally given microbial biologists a way to define their subjects, through molecular phylogeny. The sequences also are the basis of the tools that will allow microbial biologists to explore the distribution and roles of the organisms in the environment. Microbial biology can now be a whole science; the organism can be studied in the ecosystem.

REFERENCES

1. M. T. Madigan, J. M. Martinko, J. Parker, *Brock Biology of Microorganisms* (Prentice-Hall, Upper Saddle River, NJ, ed. 8, 1996).
2. A. T. Bull, M. Goodfellow, J. H. Slater, *Annu. Rev. Microbiol.* **46**, 219 (1992).

3. R. H. Whittaker, *Science* **163**, 150 (1969).
4. E. Chatton, *Titres et Travaux Scientifiques* (Sete, Sottano, Italy, 1937).
5. E. Zuckerkandl and L. Pauling, *J. Theor. Biol.* **8**, 357 (1965); R. M. Schwartz and M. O. Dayhoff, *Science* **199**, 395 (1978).
6. C. R. Woese and G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
7. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
8. _____, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
9. D. L. Swofford, G. J. Olsen, P. J. Waddell, D. M. Hillis, in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, MA, 1996), pp. 407-514.
10. W. F. Doolittle and J. R. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6721 (1994).
11. J. D. Palmer, *Science* **275**, 790 (1997).
12. C. J. Bult *et al.*, *ibid.* **273**, 1058 (1996).
13. The taxonomic term "kingdom" has no molecular definition. I use the term to indicate main lines of radiation in the particular domain; 14 such "kingdom-level" lines are associated with the eucaryal line of descent in Fig. 1 [see also (22)].
14. J. Sapp, *Evolution by Association: A History of Symbiosis* (Oxford Univ. Press, New York, 1994).
15. T. Cavalier-Smith, *Microbiol. Rev.* **57**, 953 (1993).
16. E. T. N. Bui, P. J. Bradley, P. J. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9651 (1996); A. Germot, H. Philippe, H. le Guyader, *ibid.*, p. 14614; A. J. Roger, C. G. Clark, W. F. Doolittle, *ibid.*, p. 14618.
17. T. L. Marsh, C. I. Reich, R. B. Whitelock, G. J. Olsen, *ibid.* **91**, 4180 (1994).
18. T. Rowlands, P. Baumann, S. P. Jackson, *Science* **264**, 1326 (1994).
19. E. F. DeLong, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5685 (1992).
20. L. I. Davis, *Annu. Rev. Biochem.* **64**, 865 (1995).
21. O. Kandler, in *Early Life on Earth*, S. Bengtson, Ed. (Nobel Symp. 84, Columbia Univ. Press, New York, 1993), pp. 152-160.
22. M. L. Sogin, *ibid.*, pp. 181-192.
23. D. C. Smith and A. E. Douglas, *The Biology of Symbiosis* (Edward Arnold, London, 1987).
24. V. Tunnicliffe, *Am. Sci.* **80**, 336 (1992).
25. R. A. Lutz *et al.*, *Nature* **371**, 663 (1994).
26. T. Fenchel and B. J. Finlay, *Ecology and Evolution in Anoxic Worlds*, R. M. May and P. H. Harvey, Eds. (Oxford Ser. Ecol. Evol., Oxford Univ. Press, Oxford, 1995).
27. _____, *Ophelia* **30**, 75 (1989).
28. T. M. Embley and B. J. Finlay, *Microbiology* **140**, 225 (1994).
29. M. Kates, D. J. Kushner, A. T. Matheson, Eds., *The Biochemistry of Archaea (Archaeobacteria)* (Elsevier, Amsterdam, 1993).
30. K. O. Stetter, *Am. Soc. Microbiol. News* **61**, 285 (1995).
31. C. Pitulle *et al.*, *Int. J. Syst. Bacteriol.* **44**, 620 (1994).
32. N. R. Pace, *Cell* **65**, 531 (1991).
33. B. K. Pierson, in *Early Life on Earth* (Nobel Symp. 84, Columbia Univ. Press, New York, 1993), pp. 161-180.
34. R. I. Amann, W. Ludwig, K. H. Schleifer, *Microbiol. Rev.* **59**, 143 (1995).
35. N. R. Pace, D. A. Stahl, D. J. Lane, G. J. Olsen, *Am. Soc. Microbiol. News* **51**, 4 (1985); P. Hugenholz and N. R. Pace, *Trends Biotechnol.* **14**, 190 (1996).
36. R. K. Saiki *et al.*, *Science* **239**, 487 (1988).
37. D. L. Distel *et al.*, *J. Bacteriol.* **170**, 2506 (1988).
38. T. D. Brock, *Thermophilic Microorganisms and Life at High Temperatures* (Springer-Verlag, New York, 1978).
39. A.-L. Reysenbach, G. S. Wickham, N. R. Pace, *Appl. Environ. Microbiol.* **60**, 2113 (1994).
40. D. A. Stahl, D. J. Lane, G. J. Olsen, N. R. Pace, *ibid.* **49**, 1379 (1985).
41. R. Huber *et al.*, *Syst. Appl. Microbiol.* **15**, 340 (1992).
42. S. M. Barns, R. E. Fundyga, M. W. Jeffries, N. R. Pace, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1609 (1994).
43. S. M. Barns, C. F. Delwiche, J. D. Palmer, N. R. Pace, *ibid.* **93**, 9188 (1996).
44. J. L. Stein, T. L. Marsh, K. Y. Wu, H. Shizuya, E. F. DeLong, *J. Bacteriol.* **178**, 591 (1996).
45. P. Hugenholz, C. Pitulle, K. L. Hershberger, N. R. Pace, in preparation.
46. J. A. Fuhrman, K. McCallum, A. A. Davis, *Nature* **356**, 148 (1992).
47. T. Ueda, Y. Suga, T. Matsuguchi, *Eur. J. Soil Sci.* **46**, 415 (1995); K. L. Hershberger, S. M. Barns, A.-L. Reysenbach, S. C. Dawson, N. R. Pace, *Nature* **384**, 420 (1996); S. B. Bintrim, T. J. Donohue, J. Handelsman, G. P. Roberts, R. M. Goodman, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 277 (1997); G. Jurgens, K. Lindstrom, A. Saano, *Appl. Environ. Microbiol.* **63**, 803 (1997).
48. E. F. DeLong, K. Y. Wu, B. B. Prezelin, R. V. M. Jovine, *Nature* **371**, 695 (1994).
49. W. C. Ghiorse, *Science* **275**, 789 (1997).
50. T. Gold, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6045 (1992); J. K. Fredrickson and T. C. Onstott, *Sci. Am.* **275**, 68 (October 1996); D. Lovely and F. Chapelle, *Rev. Geophys.* **33**, 365 (1995).
51. T. O. Stevens and J. P. McKinley, *Science* **270**, 450 (1995); K. Pedersen, *Earth Sci. Rev.* **34**, 243 (1993).
52. B. L. Maidak *et al.*, *Nucleic Acids Res.* **25**, 109 (1997).
53. I thank S. Kustu, G. Olsen, and C. Woese for helpful comments on the manuscript, and S. Barns, S. Dawson, C. Delwiche, and P. Hugenholz for assistance with the figures. Research in my laboratory is supported by grants from the U.S. Department of Energy and NIH.

Quantum leap.

The SCIENCE World Wide Web site just got even better. Now you can access the fully searchable database of SCIENCE research paper abstracts and news story summaries for current and back issues. Jump onto the Internet and discover a whole new world of SCIENCE at the Web address below.

www.sciencemag.org

SCIENCE