Aquatic Geomicrobiology

By

DONALD E. CANFIELD

Danish Center for Earth System Science and Institute of Biology, University of Southern Denmark, Odense, Denmark

ERIK KRISTENSEN

Institute of Biology, University of Southern Denmark, Odense, Denmark

BO THAMDRUP

Danish Center for Earth System Science and Institute of Biology, University of Southern Denmark, Odense, Denmark

ELSEVIER
ACADEMIC
PRESS

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Aquatic Geomicrobiology

By

DONALD E. CANFIELD

Danish Center for Earth System Science and Institute of Biology,
University of Southern Denmark, Odense, Denmark

ERIK KRISTENSEN

Institute of Biology, University of Southern Denmark,
Odense, Denmark

BO THAMDRUP

Danish Center for Earth System Science and Institute of Biology,
University of Southern Denmark, Odense, Denmark

ELSEVIER
ACADEMIC
PRESS

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier Academic Press
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ♾

For all information on all Academic Press publications
visit our Web site at www.academicpress.com

ISBN: 0-12-026147-2 Case Bound
ISBN: 0-12-158340-6 Paperback

PRINTED IN THE UNITED STATES OF AMERICA
05 06 07 08 9 8 7 6 5 4 3 2 1

This book is dedicated to our families:

Andreas, Ellen, Marianne, Daniela,
Mie, Karen, Lea, and Kirsten

CONTENTS

4. Carbon Fixation and Phototrophy

5. Heterotrophic Carbon Metabolism

6. The Oxygen Cycle

7. The Nitrogen Cycle

8. The Iron and Manganese Cycles

9. The Sulfur Cycle

10. The Methane Cycle

11. The Phosphorus Cycle

12. The Silicon Cycle

13. Microbial Ecosystems

PREFACE

This volume offers a comprehensive overview of the biogeochemistry of element cycling from the perspective of the organisms involved. We interface the disciplines of microbial ecology, in which biogeochemistry is often given only cursory attention, with biogeochemistry, in which the organisms are frequently viewed as mere catalysts. We report that microbes (meaning prokaryotes here) are exquisitely adapted to the environment in which they live, with often unexpected and complex behavior and physiologies allowing for extremely efficient energy utilization. Thus, individually and collectively, microbes conduct a bewildering array of different metabolisms aimed at efficient resource management, including the utilization of each other's waste products. Microbes have individual physiological adaptations allowing life in almost incomprehensible extremes of hot, cold, acid, base and salt. We learn that microbes have evolved intricate biochemical machineries capable of conducting efficient and complex chemistries. We also learn, through various pages of this volume, that evolution has been a major player in shaping the structure and function of the microbial world, and new tools in molecular biology are allowing us to explore the details of microbial evolution for the first time. Most importantly, we learn that microbes are the drivers of elemental cycling.

This volume is cross-disciplinary, so our first five Chapters include largely introductory material designed to set a common ground for the following Chapters on elemental cycling. Thus, we start in Chapter 1 by looking at microbial systematics and phylogeny from a molecular perspective. This is a rapidly evolving field, and the reader is forewarned that progress in this field is likely to continue at a rapid pace. This is followed by Chapter 2 outlining the general principles of microbial ecology, including aspects of microbial growth and population structure. Microbes operate in a chemical environment, and we devote Chapter 3 to outlining the basic chemical principles needed to understand microbial energetics and aspects of microbial competition in nature. This Chapter is followed by two Chapters (4 and 5) overviewing first phototrophy and carbon fixation and next heterotrophic carbon metabolism. The next Chapters (6 through 12) outline the interface between the biogeochemistry of elemental cycling and the organisms involved, focusing in turn on the cycling of oxygen, nitrogen, Fe and Mn, sulfur, methane, phosphorus and silicon. There is a great deal of interrelatedness between these Chapters, emphasizing the intricate and often extraordinary coupling of element cycles. The cycling of phosphorus (Chapter 11) and silicon (Chapter 12) involves microbial interaction, but not as much so as the

other elements discussed. In these Chapters, the role of phosphorus and silicon as nutrients, as well as aspects of their cycling by eukaryotic organisms, is also emphasized. We finish with Chapter 13 exploring several specific microbial ecosystems.

This volume has been written both for students and for our colleagues. We hope to offer a source of reference and inspiration for those wishing to mix disciplines traditionally rooted in the biological or the geological sciences. There has been a heightened awareness recently of the role of microbes in controlling Earth-surface processes. This awareness is not new, but it is now taken seriously. Out of this, the field of geobiology has emerged. Again, this field is not new, but its prominence is clear. Although ours is not a volume specifically focusing on geobiology, we believe that the principles and core material outlined here have broad relevance to geobiological endeavors.

We have benefited from the expert advice of a number of good colleagues who have taken the time to read and comment on one or more of the Chapters. Thus, we are indebted to Daniel Alongi, Rudy Amann, Gary Banta, David Burdige, Dan Conley, Ralf Conrad, Raymond Cox, Kai Finster, Ronnie Glud, Ellery Ingall, Andy Knoll, Joel Kostka, Bente Lomstein, Claus Olesen, Jesper Pedersen, Eric Roden, Alan Southward and an anonymous reviewer for comments and advice. We are especially thankful to Raymond Cox and Ralf Conrad, who each endured multiple Chapters of the volume. We also thank Raymond Cox for suggesting our remarkably straightforward title. Our submission would have been considerably delayed beyond its already embarrassingly tardy date without the expert technical skills of Mette Andersen, who helped in so many ways. We also thank Yvonne Mukherjee for earlier help with the volume. Jacob Zopfi kindly allowed us to reproduce one of his photographs, and Peter Søholt patiently sorted through countless computer and networking problems. The extreme patience of our editor, suffering seemingly endless delays, cannot be overlooked. Finally, we thank our funding agencies, the Danish National Research Foundation (Dansk Grundforskningsfond) and the Danish National Research Council (SNF), for generous support of our work.

Donald E. Canfield
Erik Kristensen
Bo Thamdrup
Odense, Denmark

Chapter 1

Systematics and Phylogeny

1. INTRODUCTION

The classification of microorganisms[1] is a problem that has plagued microbiologists for more than a century. The problem is introduced by considering the macroscopic world of plants, animals, and fungi. An astounding array of morphological differences can be naturally grouped into two simple physiological types. Thus, all plants share photosynthesis as their basic means of energy metabolism but display an enormous range of obvious

[1]Microorganisms include a broad range of different types of organisms, including prokaryotes, and many single-celled eukaryotes such as protists and some algae. However, unless otherwise stated, we will use the term microorganisms synonymously with prokaryotes. We will try to avoid using bacteria as general designator for prokaryotes, as reclassification of the Tree of Life (Woese, 1990) has provided a very precise definition for *Bacteria* (see below) that does not encompass the whole of the prokaryote world. We frequently use the term bacteria to designate organisms within the bacterial domain.

differences in morphology, size, reproductive strategy, and habitat. These differences can be used as indices of classification. Likewise, essentially all animals are aerobic heterotrophs, with classification based on similar considerations. Furthermore, through comparative biology, classification schemes for plants and animals can be constructed to reflect the evolutionary relationships (phylogeny) among organisms. For the microscopic world of prokaryotic organisms, that is, organisms lacking a membrane-bound nucleus and containing no organelles (with a few exceptions), the situation is quite the opposite. Rather limited differences in size, shape, and mode of locomotion are combined with an enormous range of physiological diversity. Furthermore, while some groups of microorganisms, such as cyanobacteria, for example, may be identified in ancient rocks, microbial fossils are of little use in constructing phylogenetic relationships. Fossil microbes are generally difficult to distinguish based on preserved morphology alone, and furthermore, the fossil record of microbes is spotty and poorly preserved in early Earth history, when the pace of new metabolic developments might have been most rapid.

Early microbial classification schemes sought a natural order from combined considerations of morphological and physiological characteristics (for example, earlier editions of *Bergey's Manual of Determinative Bacteriology*; Breed *et al.*, 1948). These schemes provided exact prokaryote identification by using a checklist of properties including cell shape, cell organization, basic physiology and details of cell motility, somewhat mimicking the strategy used to classify macroscopic organisms. In general, the properties of cell morphology and motility provided a higher order classification than did considerations of cell physiology. While these schemes produced an organized catalog of the prokaryote world, they provided no indication of evolutionary relationships, unlike the schemes used to classify plants and animals. Also, incorrect associations between organisms were sometimes generated. In one example, older classification schemes (Reichenbach and Dworkin, 1981) grouped filamentous cyanobacteria and filamentous colorless-sulfur bacteria together, with the filamentous nature of the organisms a principal consideration in forming the group. This grouping, like many others, has not survived in modern phylogenetic classification schemes. It is apparent that the metrics of comparison used to classify and organize members of the macroscopic world are insufficient to provide a phylogeny of organisms in the microscopic world.

A phylogenetic grouping of prokaryotic organisms became possible after the structure of DNA was discovered and its role in inheritance was made clear. It was a logical step to speculate, as Zuckerkandl and Pauling did in 1965, that "the comparison of the structure of homologous informational macromolecules allows the establishment of phylogenetic relationships" (Zuckerkandl and Pauling, 1965). Informational macromolecules include proteins, DNA and RNA, while homologous macromolecules are genes

or proteins of common function inherited from the same ancestor. The cataloging of prokaryotic organisms (and indeed all organisms) based on the molecular sequences of genetic material (DNA and RNA) and proteins has advanced with the advent of sequencing techniques and the ability to sequence more complicated, and larger, molecules. The earliest sequences were dominantly of proteins, whereas now sequences for whole prokaryote genomes are becoming routine. In order to appreciate the construction and interpretation of phylogenies based on molecular sequence comparisons, we first briefly review some key aspects of basic genetics.

2. REVIEW OF BASIC GENETICS

All information related to the biochemical function of an organism is carried by the DNA molecule. DNA is double-stranded, with each strand consisting of four different nucleotides linked in long succession. For prokaryotes, DNA consists of, typically, 1 to 6 million nucleotides per strand. A nucleotide contained in the DNA molecule consists of one of four nitrogen bases: adenine (A), guanine (G), thymine (T), and cytosine (C), bonded to the sugar deoxyribose and a phosphate molecule (Figure 1.1). The two strands of DNA are connected by hydrogen bonds where A is linked to T and G is linked to C, forming base pairs (Figure 1.2).

The DNA molecule is organized into a series of genes encoding for the production of various proteins and RNA molecules as well as some non-coding regions sometimes referred to as "junk DNA." Proteins control

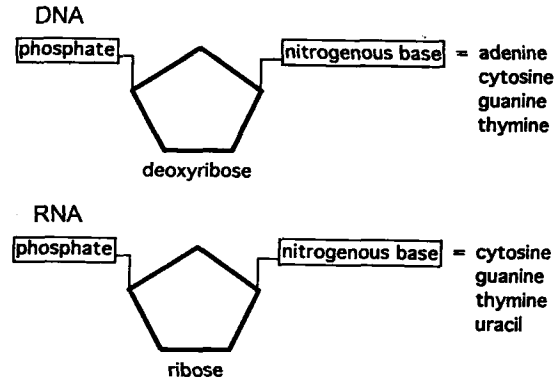

Figure 1.1 The basic structure of the nucleotides making up RNA and DNA. In RNA, the deoxyribose of DNA is replaced by ribose, and adenine is replaced by uracil.

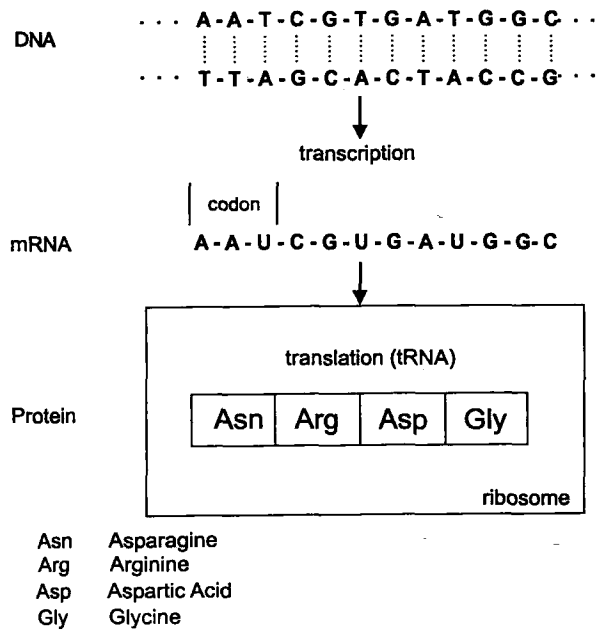

Figure 1.2 Cartoon displaying key aspects of the protein synthesis process.

the biochemical machinery of a cell, catalyzing countless chemical reactions. They may also serve a structural function in cell membranes and in various cytoplasmic components. RNA acts as an information shuttle, transferring the information from the DNA molecule to the site of protein synthesis, and it is also involved directly in protein synthesis itself (Figure 1.2). The molecular structure of RNA is quite similar to DNA, except that thymine is replaced by uracil (U), and the sugar ribose replaces deoxyribose (Figure 1.1). Furthermore, RNA is single stranded. Protein synthesis takes place in two principal steps (Figure 1.2). In the first step, genetic information from the DNA molecule is "transcribed" onto a messenger RNA molecule (mRNA). After formation, mRNA moves to the ribosome, consisting of various ribosomal RNAs (rRNA) intimately associated with a suite of ribosomal proteins. The ribosome is the place of protein synthesis within the cell. In the ribosome, information from the mRNA is "translated" into the amino acid sequence of a protein (Figure 1.2). Each three consecutive nucleotides from mRNA represent a codon that specifies for one of 20 possible amino acids that are moved into place by transfer RNA (tRNA). This process is continued until a stop codon is reached, signifying the end of the

translation process. The basic process of transcription, translation and protein synthesis is common to all living organisms on Earth.

3. MOLECULAR BASIS FOR EVOLUTION

Darwin proposed that during evolution new traits are randomly acquired by organisms during the passing of generations (descent with modification), and if advantageous to survival, traits will be selected for and passed to subsequent generations (natural selection). In fact, there is a molecular basis for this proposition. Random changes (mutations) occasionally occur in the nucleotide sequence of the DNA molecule during the replication process. This changes the nucleotide sequence of the genes affected by mutation, the sequence of the RNA molecules transcribed from the genes and, in principle, the amino acid sequence of proteins encoded for by a gene.

In some cases, mutations of the DNA nucleotide sequence are "silent," as more than one three-nucleotide sequence (codon) may code for the same amino acid. Also, mutations can occur in the non-coding portion of the DNA, the "junk DNA," with no apparent consequence to the organism. However, mutations can lead to changes in the amino acid sequence of proteins, and this can affect the organism in various ways. In some cases, a change is deleterious to the function of the protein, and hence to the organism, and will not be accepted. In other cases, a change may not alter the function of the protein (or the RNA molecule) but will be transferred to subsequent generations and will provide a reminder of the mutation event. These changes in the sequences of homologous genes (and in the sequence of the protein or RNA encoded for by the gene) are the basis of the phylogenetic reconstructions imagined by Zuckerkandl and Pauling (see below). Gene mutations can also prove advantageous to the organism and can even promote new gene function or spawn the emergence of new species. In addition, the chance duplication of genes, followed by mutation and modification of the new gene, could also produce new gene function or could provide superior mediation of the original gene function (Maynard Smith, 1998). Gene deletions may also occur, altering the biochemical machinery of the organism.

Other forces acting at the molecular level may also promote evolution. Thus, the lateral transfer of genes between organisms may provide new opportunities for the host that could not be derived from a linear sequence of gene mutation and change (Maynard Smith, 1998). Lateral gene transfer can occur through a variety of paths; viral infection, in which the virus may swap genes with the host as a result of infection (see Chapter 3), is one

example. Some organisms may also directly incorporate DNA from the environment into their genome.

At different levels, gene sequences, proteins and RNA molecules provide a comparative evolutionary history of organisms. The most information, which is also the most difficult to interpret, comes from whole-genome sequences. These provide an accumulated history of molecular-level changes that have shaped the evolution of an organism (Fitz-Gibbon and House, 1999; Lin and Gerstein, 2000; Clarke *et al.*, 2002; House *et al.*, 2003). The interpretation of whole-genome sequences is a rapidly evolving field. At another level, the comparison of sequences from homologous genes, RNA molecules and proteins provides the basis for constructing evolutionary relationships that have been extensively used to formulate organismal phylogenies among prokaryotes (e.g. Woese, 1987; Doolittle *et al.*, 1996; Brown and Doolittle, 1997).

4. CONSTRUCTING PHYLOGENIES

A great deal of thought and consideration has gone into the construction of organismal phylogenies from molecular sequence data. This matter is not trivial. Certain criteria must be met before a genetic element (gene, protein or RNA molecule) may be considered as appropriate for constructing phylogenies. These criteria have been carefully elucidated by Carl Woese and his associates (Woese, 1987; Olsen and Woese, 1993), and some of the important considerations are outlined below.

1. Phylogenies are generally constructed from genetic elements with the same function (homologous function). Furthermore, phylogenies constructed from genetic elements are only as inclusive as the distribution of the given genetic element among organisms. The more widely distributed the element, the more inclusive the phylogeny, and the less widely distributed, the more specialized the phylogeny. If a phylogeny is constructed, for example, from the nitrogenase enzyme, then the phylogeny will include only those organisms capable of fixing nitrogen.

2. Larger genetic elements provide more sequence differences among organisms. More differences provide better resolution and more statistically valid phylogenies.

3. The genetic element should accumulate mutations at a rate commensurate with the resolution required. Organisms separated by large evolutionary distances are best compared using sequences in which mutations accumulate only slowly. By contrast, a more rapid accumulation of mutations would benefit the comparison of organisms recently diverged from one another. In this case, distant evolutionary relationships might become obscured by too many mutational changes.

4. Do the phylogenies constructed from the given genetic element reflect the phylogeny of the organism? They do if, as elegantly stated by Olsen and Woese (1993), "the clear majority of the essential genes in a genome share a common heritage." This should be true if, in turn, the lateral transfer of genetic material from one organism to another has not been the predominant means of genome construction. This issue will be considered in some detail below.

5. Related to the previous consideration, the construction of phylogenies implicitly assumes that modern organisms have been ultimately derived from one common ancestor (or a suite of ancestors freely exchanging a common genetic pool) (Woese, 2000).

Independent of the type of genetic element used to construct the phylogeny, the basis for phylogenetic reconstructions is comparative sequence analysis. As an example, the phylogeny of seven hypothetical organisms based on a 10-base sequence of either "X's" or "O's" is illustrated in Figure 1.3. It is assumed that organism "a" is the last common ancestor to all of the others. Also demonstrated in Figure 1.3 is how quickly phylogenetic resolution can be lost if the information content of the genetic element is too small.

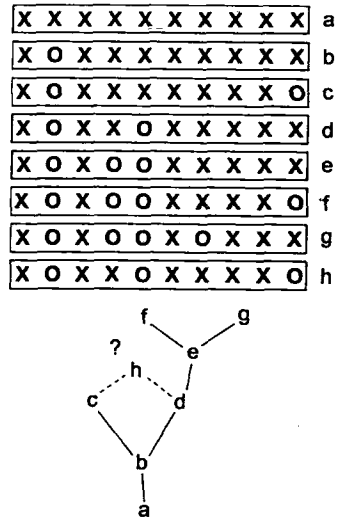

Figure 1.3 Phylogeny drawn from a hypothetical group of organisms with characteristic protein sequences designated either by an O or an X. The placement of organism h is uncertain with the limited sequence information available. It could have evolved from either c or d. Back mutations are not considered.

In practice, the analysis of molecular sequences to obtain phylogenetic information is quite complex and requires sophisticated computer algorithms. Corrections must be made for the possibility of back mutations and for homoplasy, the independent mutation of the same site in two lineages. Furthermore, analogous regions of the genetic element must be correctly identified and aligned for all of the organisms considered, and sequences must be statistically treated to provide the most probable order of emergence and its statistical validity (Hillis *et al.*, 1993; Huelsenbeck and Rannala, 1997).

5. PHYLOGENIES FROM RIBOSOMAL RNA MOLECULES

Phylogenies have been constructed from a range of different DNA, RNA, and protein sequences. By far, the most widely used and comprehensive phylogenies have been constructed from the small subunits of the rRNA molecule. As mentioned previously, the ribosome is the location of protein synthesis in all organisms. The ribosome of prokaryotic organisms is constructed of 30S and 50S subunits, yielding together 70S ribosomes. For eukaryotic organisms, 40S and 60S subunits make up 80S ribosomes. The "S" refers to Svedberg units and is the sedimentation coefficient of ribosomes or ribosomal subunits subjected to ultracentrifugation.

Each of the ribosome subunits is a complex of proteins and various rRNA molecules. For prokaryotes, the 30S subunit contains 16S rRNA and the 50S subunit contains both 5S and 23S rRNA, while in eukaryotes, the 40S subunit contains 18S rRNA and the 60S subunit contains 5.8S and 28S rRNA (Table 1.1). Each of the comparable rRNA subunits in prokaryotes and eukaryotes (e.g., 16S rRNA in prokaryotes and 18S rRNA in eukaryotes) is functionally equivalent and structurally similar (Figure 1.4). The

Table 1.1 Comparison of ribosomal RNA (rRNA) composition of prokaryotes and eukaryotes

	Prokaryotes	Eukaryotes
Overall size	70S	80S
Small subunit	30S subunit 16S rRNA (1500)	40S subunit 18S rRNA (2300)
Large subunit	50S subunit 5S rRNA (120) 23S rRNA (2900)	60XS subunit 5S rRNA (120) 5.8S rRNA (160) 28S rRNA (4200)

Number of bases in parenthesis. Modified after Madigan *et al.* (2003).

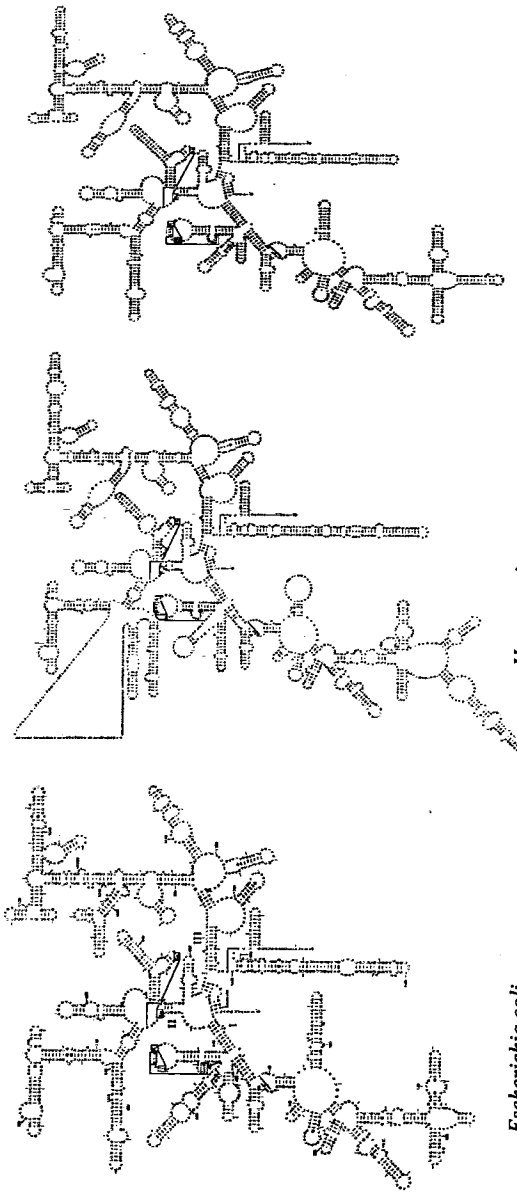

Figure 1.4 Sequences of 16S rRNA from members of the *Bacteria* and *Archaea* and 18S rRNA for a representative *Eukarya*. Emphasized is the secondary structure of the molecule, showing the broad similarity in structure for organisms throughout the Tree of Life. Structures from Maidak *et al.* (1999).

reader is referred to standard textbooks (e.g. Madigan *et al.*, 2003) for further discussion of ribosome function and structure.

Early phylogenies from rRNA molecules were constructed from sequence comparisons of the 5S rRNA subunit (Woese and Fox, 1977). These sequence comparisons demonstrated the potential for rRNAs to produce a template for organizing prokaryotic taxonomy and for generating a phylogeny for prokaryotic organisms as was never before possible. Unfortunately, the number of nucleotides in the 5S rRNA molecule is small (approximately 120), severely limiting the phylogenetic resolution of the molecule. As sequencing techniques developed, it became possible to sequence the larger 16S and 18S rRNAs, collectively known as the small subunit (SSU) rRNAs (Table 1.1). This development led to a widely accepted molecular-based "Tree of Life" and has forever changed our views of the taxonomic organization of the prokaryotic world, as well as our understanding of organismal evolution on Earth.

Molecular sequences from the SSU rRNA molecule have proven ideal for taxonomic organization and phylogenetic reconstructions for several reasons (Pace *et al.*, 1986; Woese, 1987; Olsen and Woese, 1993). First, the SSU rRNA molecule is distributed among all organisms on Earth. Second, the function of the ribosome, protein synthesis, is tightly constrained so the molecular sequences of the SSU rRNA are highly conserved, accumulating mutations only slowly; this allows comparison between organisms of distant evolutionary association. However, different regions of the SSU rRNA molecule accumulate mutations at different rates, providing resolution also for fairly closely associated organisms. Next, the SSU rRNA is relatively large (approximately 1500 nucleotides for prokaryotes and 2300 nucleotides for eukaryotes) (Table 1.1), allowing good phylogenetic resolution. Finally, the transcription–translation process of protein synthesis is so fundamental to the livelihood of an organism that the lateral transfer of rRNA between cells is unlikely to be a common mode of rRNA acquisition by the cell.

6. TREE OF LIFE FROM SSU rRNA COMPARISONS

At present (June, 2004) 97,000 prokaryote 16S rRNA sequences have been determined, aligned and archived for use by the scientific community (see http://rdp.cme.msu.edu). In fact, the number of different prokaryotic "species" known from molecular sequences is far greater than the number of "species" (more on the species concept in Chapter 2) actually cultured in the lab. This is because while it is relatively easy to extract DNA for molecular sequencing, it is difficult to find the appropriate culture conditions by which to enrich and isolate microorganisms. A general consensus guess is that only about 1% of the prokaryotic organisms present in nature have

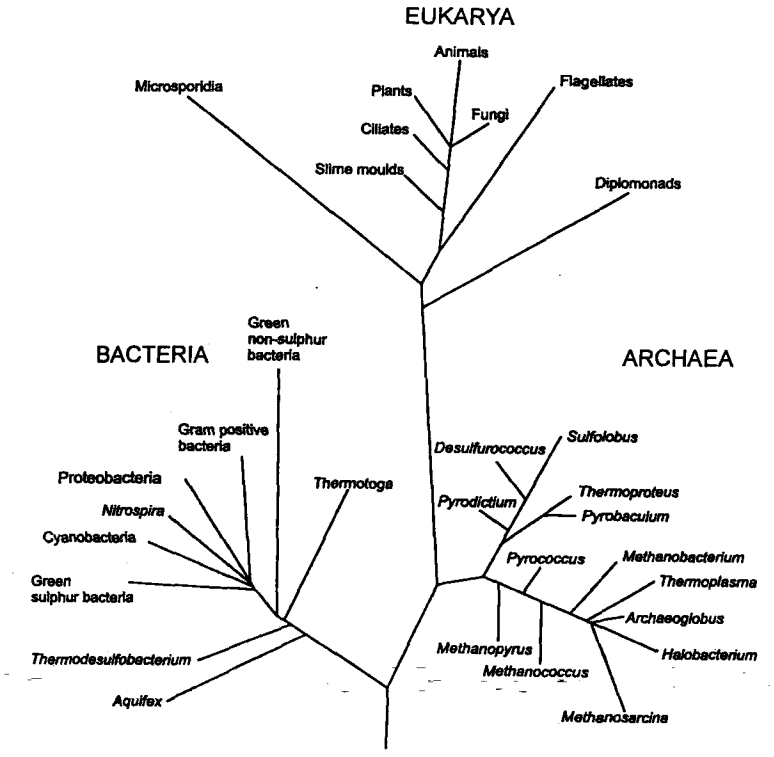

Figure 1.5 The Tree of Life from SSU rRNAs. See text for details. The tree is adapted from Olsen and Woese (1993), with additional lineages from Pace (1997).

been brought into culture so far (Pace, 1997). Nevertheless, from the available prokaryote and eukaryote sequences a Tree of Life emerges (Figure 1.5).

Molecular sequences from the SSU rRNAs naturally divide all life into three principal domains: *Bacteria*, *Archaea*, and *Eukarya*. Two of the domains, *Archaea* and *Bacteria*, are composed of prokaryotic organisms, while the *Eukarya* domain consists of the eukaryotic organisms. This tree has been rooted based on the presence or absence of a duplicated gene coding for the enzyme ATPase. This comparison suggests that the *Bacteria* are the out group (they do not contain the duplicated gene) (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989), placing the root between the *Bacteria* and the domains *Eukarya* and *Archaea*. However, there is still much discussion about the placement of the root, and this issue is not yet fully resolved.

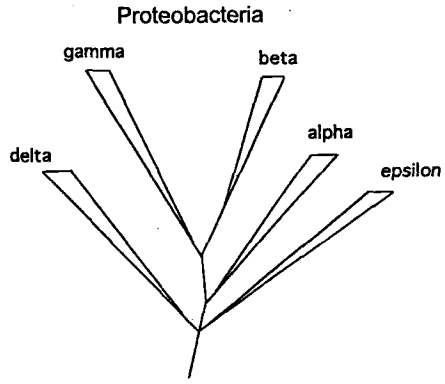

Figure 1.6 The branching order of the principal divisions within the proteobacteria (purple bacteria). See Figure 1.5 for the placement of the proteobacteria in the Tree of Life and the text for additional details.

Taken at face value, deeper branching lineages within the Tree of Life represent organisms evolved from more ancient ancestors than lineages branching further up on the tree. Therefore, the *Aquifex* lineage, for example, was presumably established before the earliest evolving cyanobacteria, and plants were established well after the emergence of diplomonads and slime molds. Likewise, the evolutionary history of a given lineage is represented by the branching order of major divisions, or groups of organisms, in the lineage. Thus, there are five major divisions within the proteobacteria (Figure 1.6): α, β, δ, ε, and γ, and of these, the γ and β subdivisions separated most recently, while the δ and the ε subdivisions are the most ancient.

The similarity or difference in the SSU rRNA sequences of two organisms is a measure of the evolutionary distance separating the two organisms. Trees are constructed so that the branch lengths represent the differences in SSU rRNA sequence relative to nearest neighbors, as well as to all other organisms on the Tree of Life. Evolutionary distance between two organisms is measured as the fractional difference between two sequences and is determined as the sum of the branch lengths connecting organisms through the shortest path. Thus, for sulfate-reducing organisms of the δ subdivision of the proteobacteria (Devereux and Stahl, 1993) (Figure 1.7), representing most known sulfate reducers, the evolutionary distance between species of the genus *Desulfobacter* is much less than the evolutionary distance between species of the genus *Desulfovibrio*. Furthermore, the evolutionary distance between *Desulfobacter* sp. and *Desulfobacterium* sp. is less than the distance between *Desulfobacter* sp. and *Desulfovibrio* sp. These evolutionary distances are all much shorter than the distance between members of the proteobacteria and members of the domains *Archaea* or *Eukarya* (see Figure 1.5).

Extraction, amplification and sequencing of DNA for phylogenetic analysis

The construction of phylogenies from molecular data begins, most frequently, with the extraction of DNA from organisms. In both lab cultures and environmental samples DNA is extracted after cell disruption, with various strategies, including beating the sample with small beads, sonication, freeze/thaw cycles and chemical lysis, used either singly or in combination. The DNA is subsequently purified by a variety of means, including phenol-chloroform extraction, dialysis and ethanol precipitation. After this, segments of DNA are amplified by polymerase chain reaction (PCR). The segment of DNA chosen (typically 100 to over 1000 base pairs in length) will be part or all of a gene sequence coding for specific enzymes or the RNA molecules used in constructing phylogenics. To begin a PCR cycle, DNA is melted, separating the strands, to which primers are attached. Primers consist of short stretches of nucleotides complementary to the regions of the DNA sequence flanking the target gene. Thus, the sequence of the desired gene(s) must be known so that primers can be constructed with good gene specificity yet wide applicability to a variety of organisms with the same gene.

The primers are added in great excess, and as the primer–DNA mixture cools, the primers fix to the DNA. After this, a thermostable DNA polymerase (the DNA polymerase from Thermus aquaticus is often used) extends the primer along the target gene of interest. This mixture is reheated, separating the DNA again into single strands and doubling the target sites for a new cycle of PCR. Typically, 30–40 cycles are run, yielding an increase in the number of sequences of 10^9 times or more. In environmental samples, a large yield of gene sequence is possible from organisms present in only small numbers. Obviously, contamination is also of great concern.

Since mutations in the SSU rRNA accumulate over time, in principle, the Tree of Life might be used to tell when in Earth history major episodes of biological evolution and innovation occurred. Important biological milestones include the evolution of individual metabolic capabilities such as methanogenesis, sulfate reduction, nitrogen fixation, denitrification and, of particular interest, oxygenic photosynthesis. Furthermore, the Tree of Life might help resolve when plants and animals first evolved.

Unfortunately, the use of SSU rRNA sequences to constrain the timing of major biological innovations has, thus far, proven to be more of a hope than a reality. It is clear that mutation rates may vary among different major lineages

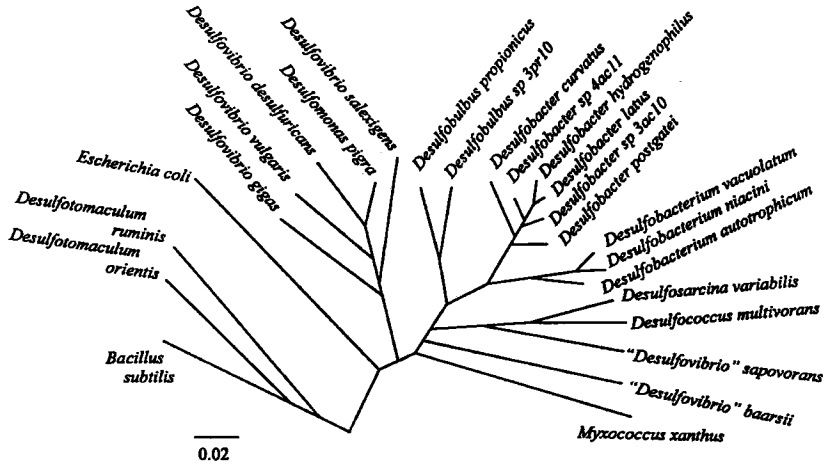

Figure 1.7 Phylogeny of members of selected genera of sulfate reducers from the proteobacteria and the gram-positive bacteria from 16S rRNA sequences. Redrawn from Devereux and Stahl (1993).

within the Tree of Life, and furthermore, mutation rates may also change over time. Early attempts to "calibrate" the Tree of Life utilized a general SSU rRNA mutation rate determined from SSU rRNA sequence differences among organisms calibrated with evolutionary events constrained by the fossil record (Ochman and Wilson, 1987). The calibration points were mostly from fossils of the Phanerozoic (geologic eon comprising the last 544 million years), and hence were young compared to the full time over which biological evolution has occurred on Earth. This approach, when extrapolated back in time, revealed unrealistically young ages for important evolutionary events such as the emergence of oxygenic photosynthesis (cyanobacteria). The value of SSU rRNA molecular clocks could improve with a better understanding of mutation rate variability within the tree, constrained, possibly, by calibration points further back in time. Also, alternative molecular clocks, based on comparisons of gene sequences coding for proteins, show promise in unraveling the timing of major biological innovations in Earth history (e.g. Feng *et al.*, 1997).

7. COMPARISONS WITH PREVIOUS TREES OF LIFE

Previous divisions of the natural world have been heavily influenced by what we can see. Whittaker and Margulis (1978) proposed that the diversity of life on Earth could be represented by five kingdoms: *Animalia, Plantae, Fungi,*

Protista, and *Monera* (prokaryotes), of which the first four are eukaryotic organisms, and the first three comprise most of the macroscopic living world. In an updated version, Margulis and Schwartz (1998) incorporated aspects of the SSU rRNA phylogeny but retained the five-kingdom division of life.

The SSU rRNA-based Tree of Life, in contrast with previous trees, emphasizes the diversity of prokaryotic organisms. Indeed, three of the prominent kingdoms in the tree of Whittaker and Margulis (1978), *Animalia, Plantae,* and *Fungi,* represent late-evolving lineages in the SSU rRNA Tree of Life, with little sequence variation separating them. As prokaryotes encompass such an enormous range of metabolic diversity and have been so prominent in the history of life on Earth, we feel that the three-domain Tree of Life, as derived from molecular sequences, provides a better phylogeny of the living world. Therefore, in the remainder of this work, we adopt the organization of the living world as provided by the SSU rRNA Tree of Life.

8. A BRIEF TOUR THROUGH THE TREE

8.1. Domain *Bacteria*

The *Bacteria,* one of the two prokaryotic domains, is divided into 30 currently recognized phyla. For example, the cyanobacteria, the proteobacteria (also known as the purple bacteria) and the green sulfur bacteria are all recognized as phyla. Some of these phyla, however, contain relatively few species, and taxonomic designations within the bacterial domain are at present rather imprecise. In many cases international standards have not been established for classification much beyond the genus level. This situation is changing, with new editions of *Bergey's Manual of Determinative Microbiology* and *The Prokaryotes* now being assembled and disseminated.

The bacterial domain is genotypically (based on genetic information) quite distinct from the *Archaea,* the other prokaryotic domain (see below), but there are many phenotypic characteristics (observable properties) that also distinguish the *Bacteria* from the *Archaea.* First, great metabolic diversity is housed in the *Bacteria,* including many physiologically unique groups of organisms. For example, among the prokaryotes, oxygenic photosynthesis (the cyanobacteria) is unique to the *Bacteria.* The *Bacteria* is also the principal domain for anoxygenic photosynthetic organisms. Other metabolisms such as methanotrophy[2] (methane oxidation), methylotrophy (oxidation of

[2]Methanotrophy may also occur among some methanogens during the anaerobic oxidation of methane, although the details of this process are still unclear (see Chapter 10).

one-carbon compounds, not including methane) and nitrification are also, as far as we currently know, restricted to the *Bacteria*. All known endospore-forming prokaryotes are found within the bacterial domain (gram-positive bacteria). Organisms from the bacterial domain also contain unique cell wall features. For example, with only a couple of exceptions, members of the *Bacteria* have the cell rigidifier peptidoglycan.

Other distinguishing features of the bacterial domain include the deep branching of high temperature-adapted organisms. These include the chemolithoautotrophic *Aquifex-Hydrogenobacter* group, sulfate reducers of the genera *Thermodesulfobacterium* and *Thermodesulfatator* (Moussard *et al.*, 2003) and the dominantly fermentative organisms of the Thermotogales group. All of these organisms have temperature optima in the 80–90 °C range (Stetter, 1996). Also deep branching are anoxygenic photosynthetic organisms of the family *Chloroflexiaceae*. It would appear that great metabolic diversity in hot environments accompanied early bacterial evolution (Stetter, 1996). An inspection of tree topology shows that the majority of the domain, including most of the bacteria with which we are most familiar, including members of the cyanobacteria, the proteobacteria, and the nitrospira group, emerged from approximately the same point in the tree. This could represent a massive diversification of bacterial life, possibly in association with cyanobacterial evolution (Knoll and Bauld, 1989; Canfield and Raiswell, 1999).

8.2. Domain *Archaea*

The *Archaea* is the second of the two prokaryotic domains and is organized into three principal kingdoms: *Crenarchaeota, Euryarchaeota*, and *Korarchaeota*. Of the three archaeal kingdoms, only the *Crenarchaeota* and *Euryarchaeota* have members with well-studied physiologies. The *Korarchaeota* are known only from molecular sequences (e.g., Barns *et al.*, 1996), with no cultured representatives so far. Molecular sequences have revealed, therefore, a fundamental new branch of life with possibly significant implications for the path of early evolution and physiological diversity within the *Archaea*. We wait eagerly for the isolation and physiological study of organisms from this kingdom.

Overall, physiological diversity within the *Archaea* is not as extensive as in the *Bacteria*. Nevertheless, there are some unique aspects. For example, methanogenesis is solely housed within the *Archaea*, and organisms with the highest temperature tolerance known, up to 121 °C, are also found in the *Archaea* (Kashefi and Lovley, 2003). Furthermore, within the *Archaea* are found the halophiles, organisms commonly growing at salt concentrations up to halite saturation (highly salt-tolerant organisms are, however, also found among the *Bacteria*). Some members of the halophiles provide the reddish pigmentation typical of salt-concentrating ponds and conduct a unique type of light-dependent energy conservation, a kind of photosynthesis, relying on the protein bacteriorhodopsin (see Chapter 4). This protein is functionally and structurally similar to the eye pigment rhodopsin. This type of photosynthesis is used to generate ATP and does not involve oxidation–reduction reactions. As with the *Bacteria*, certain cell wall features also characterize the *Archaea*. Thus, most *Archaea* have a cell wall surface layer consisting of protein or glycoprotein. Some methanogens contain a cell wall quite similar to the peptidoglycan cell walls of *Bacteria*. However, this pseudopeptidoglycan lacks muramic acid, a key component in bacterial peptidoglycan. Cell membranes in *Archaea* are also distinctive, as they contain ether-linked lipids as opposed to the ester-linked lipids found in organisms from the domains *Bacteria* and *Eukarya*.

Archaea have been isolated from environments considered to be the most extreme habitats for life on Earth. These environments include those with high temperature (hyperthermophiles), high salinity (halophiles), and extremes of pH. Recently, a much greater environmental range for *Archaea* has been established by the identification of *Archaea* as an important constituent of marine picoplankton (Delong, 1992; Karner *et al.*, 2001) and the discovery of symbiotic associations between *Archaea* and some marine animals (Preston *et al.*, 1996). Unfortunately, these *Archaea* have yet to be cultured.

Like the *Bacteria*, deep-branching organisms in the *Archaea* tend to be hyperthermophilic (thriving at high temperatures above 80 °C). The hyperthermophilic nature of deep-branching organisms has led to speculation that the last common ancestor to the three domains of life may also have been hyperthermophilic (Woese, 1987; Stetter, 1996) and, furthermore, that life itself may have originated in a high-temperature environment (Shock, 1996; Russell and Hall, 1997). This hypothesis has not been universally accepted (Galtier *et al.*, 1999) and, recently, low-temperature environments have also yielded molecular sequences from deep-branching *Archaea* (Delong, 1992; Delong *et al.*, 1994). Culturing of these organisms may generate further insights into the path of early biological evolution.

8.3. Domain *Eukarya*

Eukaryotic organisms are easily distinguished by their membrane-bound nucleus and by the presence of organelles.[3] Organelles are membrane-bound bodies, for example, mitochondria and chloroplasts, which specialize in carrying out specific functions within the cell. Compared to the prokaryotic domains, the metabolic diversity (as opposed to biochemical and

morphological diversity) within the *Eukarya* is quite limited. Thus, eukaryotic organisms are oxygen-producing phototrophs, are aerobic heterotrophs, or in some cases, survive by fermentation (yeast, for example). Although eukaryotes dominate the macroscopic world, most of the SSU rRNA diversity within the domain is accounted for by microscopic organisms including protozoans (single-celled eukaryotes without a cell wall, including many human pathogens), unicellular algae, and some fungal types including yeast. Indeed, as noted previously, higher plants and animals occupy only late-evolving lineages within the domain.

Whereas most metabolic diversity has evolved within the prokaryotic domains, cellular complexity is a hallmark of eukaryotes and is a product of eukaryote evolutionary history. The earliest stages of eukaryote evolution are unclear, but in one prominent model (Sogin, 1991) an early proto-eukaryote, lacking a nucleus and with a fragmented, possibly RNA-based genome, engulfed an early archaeon. A fusion between the DNA-based genome of the archaeon and the fragmented genome of the proto-eukaryote formed the basis of the early eukaryote genome. The engulfed archaeon developed into the membrane-bound nucleus that distinguishes eukaryotes. This scheme is consistent with, and derived partly from, the relatedness between *Archaea* and *Eukarya* in the Tree of Life (Figure 1.5). In a somewhat similar proposition, Zillig *et al.* (1992) proposed that eukaryotes originated from a fusion between an early bacterium and an archaeon, with, as before, the archaeon becoming the nucleus.

More certain is the origin of some of the organelles, in particular, the mitochondria and chloroplasts. Both of these were derived from endosymbiotic associations between eukaryote cells and members of the *Bacteria* (Margulis, 1970). Thus, mitochondria arose from an endosymbiotic proteobacterium, whereas chloroplasts developed from endosymbiotic cyanobacteria. Each of these organelles contains their own rRNA whose sequences clearly support their endosymbiotic origins in the *Bacteria*.

The deepest-branching members of the *Eukarya* (Diplomonads, Microsporidia, and Trichomonads) are anaerobic protozoans lacking mitochondria (Sogin, 1989). It has therefore been argued that these lineages originated from primitive eukaryotes that evolved before the endosymbiotic event forming mitochondria. However, the nuclear genomes of many of these deep-branching eukaryotes contain what are apparently mitochondrial-derived genes

[3]What might be considered organelles are also found among certain prokaryote groups, including the chlorosomes of the green sulfur bacteria (see Chapter 9) and the anammoxozomes of some members of the genus *Planctomycetes* involved in anaerobic ammonia oxidation (see Chapter 7) (Sinninghe Damsté *et al.*, 2002). Other members of the *Planctomycetes* also contain membrane-bound bodies, including a membrane-bound nuclear structure.

(Sogin, 1997). Therefore, it seems more likely that these deep-branching eukaryotes once had mitochondria but later lost them. The loss of mitochondria may have resulted from a dominantly parasitic lifestyle and a retreat to anoxic habitats (Sogin, 1997). Even though eukaryotes stand alone as a principal domain in the Tree of Life, their evolution and defining features (nucleus and organelles) are ultimately tied to organisms of the prokaryotic domains.

9. HOW TRUE ARE SSU rRNA-BASED PHYLOGENIES?

Phylogenies drawn from the SSU of the rRNA molecule provide an unparalleled view into the evolution of prokaryotic life. Furthermore, these phylogenies have given us the three-domain Tree of Life, have been instrumental in defining *Archaea* as a separate domain, and have provided solid support for the endosymbiotic origins of mitochondria and chloroplasts in eukaryotic cells. With all this success, there is still debate as to whether the Tree of Life, as constructed from SSU rRNA, is correct (Doolittle, 1999). There are two major concerns. The first is statistical: as phylogenies are drawn from statistical analysis of molecular sequences, there is some concern whether the branching patterns depicted on SSU rRNA-based trees are statistically valid (van de Peer *et al.*, 1994). Statistical treatments are basically exercises in hypothesis testing, and the validity of the given phylogeny depends on the validity of the hypothesis and the extent to which the statistical test shows valid differences among various molecular sequences (Hillis *et al.*, 1993; Huelsenbeck and Rannala, 1997). Statistical problems in determining phylogenetic relationships should become reduced as more sequences become available, particularly in the deep branches, and as statistical treatments improve.

A potentially far more serious concern is whether it is appropriate, or even valid, to think about organisms as having a unique phylogeny (Doolittle, 1999). If the lateral transfer of genetic material from one organism to another has been a major mode of genome construction, then different genes would be expected to yield different, equally valid, phylogenies. Yet, none of these phylogenies could be taken to represent the evolutionary history of the organism; in fact, the organism would have a complex history defined by several different episodes of gene acquisition (Hilario and Gogarten, 1993). The evidence for lateral gene transfer among prokaryotic organisms is now indisputable, including divergent phylogenies obtained using different genes and the apparent acquisition of archaeal genes by bacterial genomes (Brown and Doolittle, 1997; Jain *et al.*, 1999; Nelson *et al.*, 1999). An emerging concept is that informational genes, those genes involved in the basics of

cell function such as the transcription–translation process of protein synthesis (including ribosomal RNA genes), are relatively unsusceptible to gene transfer. By contrast, operational genes, those involved in cellular metabolism and "housekeeping," are more readily transferred (Jain *et al.*, 1999). The reality of lateral gene transfer has understandably given rise to a certain skepticism as to the validity of SSU rRNA-based phylogenies (Philippe and Laurent, 1998).

In principle, comparison of whole genome sequences should provide the best resolution of lateral gene transfer events between organisms and the most accurate organismal phylogenies. Thus far, whole genome sequences have provided equivocal answers to the question of organismal evolution and the phylogenetic validity of the SSU rRNA tree. Whereas lateral gene transfer is clearly indicated from whole genome sequences (e.g., Nelson *et al.*, 1999), phylogenetic analysis of whole genome gene families yields trees quite similar to the SSU rRNA tree (Fitz-Gibbon and House, 1999; Clarke *et al.*, 2002; House *et al.*, 2003). The principal domains are preserved, as is the differentiation of most phylogenetic groups as identified by SSU rRNA. Nevertheless, there are some important differences. For example, in the whole genome tree of House *et al.* (2003), an early evolution of the methanogens is not supported and some non-gram-positive bacteria cluster with the gram-positive group.

Whole genome phylogenetic analysis is in its infancy, and so far only a relatively small number of whole genomes have been sequenced and annotated (about 100, but the number is growing rapidly). Hopefully the pathways of genome acquisition by organisms will soon become clearer.

10. SSU rRNA SEQUENCES AS A TAXONOMIC TOOL

Despite current debate regarding the interpretation of organismal phylogenies, especially those obtained from deep branches in the SSU rRNA Tree of Life, organisms are not simply a homogeneous mix of shared genetic information. The Tree of Life provides countless examples of sensible groupings based on similar phenotypic expression or other distinct cell properties. As a conspicuous example, cyanobacteria represent a phylogenetically distinct monophyletic group with the common physiology of oxygenic photosynthesis.

If we look closely, related organisms based on SSU rRNA sequence comparisons (<5 to 10% difference) typically share common physiologies. For example, among the sulfate reducers of the Delta subdivision of the proteobacteria (Figure 1.7), species of the genus *Desulfobacter* group together and are related in their ability to oxidize acetate (see Chapter 9). Furthermore, all species of the genus *Desulfovibrio* are lactate utilizers, and

these are all vibrio shaped. Within the archaeal domain, sulfate reducers of the genus *Archaeoglobus* all group together. These similarities, and many others, have encouraged the use of the SSU rRNA tree as a fundamental tool in taxonomic classification. Thus, a complete description of a new organism now requires a SSU rRNA sequence, and genus-level classification is typically based on where in the tree the organism resides.

This SSU rRNA-based taxonomy has also required some changes in previous taxonomies. For example, the organism *Desulfobotulus sapovorans* (Figure 1.7) was previously classified as a *Desulfovibrio* species based on similarity in phenotypic characters, and shape, to the organisms of the genus *Desulfovibrio*. However, SSU rRNA sequences clearly indicated a different evolutionary history for this organism, and hence a new genus name was needed. Other instances are outlined in the following Chapters. The physiological diversity of organisms known only from their SSU rRNA sequences (molecular isolates) is sometimes inferred from placement within the tree, but one should be wary of assigning physiologies to organisms before they have been isolated and studied in the laboratory.