ELSEVIER

Molecular microbial ecology: land of the one-eyed king

Larry J Forney*, Xia Zhou and Celeste J Brown

Studies of microbial biodiversity have made astounding discoveries of late due to the use of methodologies based on phylogenetic analyses of small subunit ribosomal RNA sequences. Although there are limitations to these methods, they can nonetheless be very useful if these limitations are kept in mind. These limitations range from technical problems such as obtaining representative genomic DNA and suitable primers, to conceptual problems such as defining and using meaningful taxonomic units of diversity (species). Here we discuss several of the limitations inherent in studies of microbial diversity that must be considered when interpreting the results obtained using these approaches.

Addresses
Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA
*e-mail: lforney@uidaho.edu

Current Opinion in Microbiology 2004, **7**:210–220

This review comes from a themed issue on
Ecology and industrial microbiology
Edited by Elizabeth Wellington and Mike Larkin

Available online 18th May 2004

1369-5274/$ – see front matter

DOI 10.1016/j.mib.2004.04.015

Abbreviations

DGGE	denaturing gradient gel electrophoresis
RDP	ribosomal database project
SSU	small subunit
T-RFLP	terminal-restriction fragment length polymorphism

Introduction

In recent years the exploration of microbial biodiversity has taken a quantum leap forward. While microbiologists were previously limited by their inability to characterize uncultured organisms, the advent to so-called 'cultivation independent' methods has provided researchers with the ability to determine the composition of microbial communities and identify numerically important, but not yet cultured organisms. This has resulted in numerous remarkable discoveries. Included among these are the fact that Archaea constitute about one-third of pelagic marine ecosystems [1]; acidobacteria are common in many soils [2]; 'simple' hot spring microbial mat communities are not so simple [3]; most unusual environments (e.g. deep sea thermal vents) are populated by many taxa that have not previously been cultivated [3–7]; the microbial diversity of many familiar environments such as the gingival of humans are extraordinarily complex [8]; even the most exotic habitats have diverse kinds of prokaryotes; and to find novel taxa, one doesn't need to look further than their own backyard. Indeed, the scientific literature of the past two decades is replete with accounts of studies in which novel taxa have been discovered. These reports have been so abundant, that it is no longer considered novel to report the discovery of novel organisms.

These advances were possible because of the seminal work of Carl Woese [9–11] and Norm Pace [12,13], as well as their students and colleagues, who advocated the use of molecular phylogeny to characterize microbial diversity and to classify organisms based on their evolutionary relationships [14]. Their approach centered on determining and comparing the DNA sequences of small subunit (16S and 18S) rRNA (SSU rRNA) molecules to understand their phylogenetic relationships. The efforts of microbial explorers have led to the discovery of two large subdomains within the Archaea (Crenarchaeota and the Euryarchaeota), and a tripling of the number of identifiable bacterial divisions [14]. Many of the bacterial divisions including *Acidobacterium* [15], *Verrucomicrobia* [16], and many others are poorly represented by sequences from cultivated organisms [14,17] and virtually nothing is known about their physiology or ecology.

The advances made in research to define and understand microbial diversity have opened the eyes of biologists to the fact that the number of prokaryotic species may well exceed that of all other life forms on the planet [18••], and the vast majority of the biomass on Earth is comprised of prokaryotic cells [19]. The number of prokaryotes and the total amount of their cellular carbon on earth are estimated to be $4–6 \times 10^{30}$ cells and 350–550 pg of carbon ($1\ \text{pg} = 10^{15}$ g), respectively. Indeed, the total amounts of carbon, nitrogen and phosphorus are roughly equal to that in terrestrial plants. Species of prokaryotes possess immense genetic and metabolic diversity. They are key players in major geochemical cycles and climate change, and have practical importance in agriculture, disease prevention, animal nutrition, waste treatment, biotechnology and much more. Consequently, a better understanding of microbial community structure and function is critically important to sustaining life on Earth and gauging the impact of human activities on the functions within ecosystems. These discoveries about microbial diversity and abundance led Mark Wheelis to state: 'The Earth is a microbial planet, on which macroorganisms are recent additions — highly interesting and extremely complex in ways that most microbes aren't, but in the final analysis relatively unimportant in a global context.'

(as quoted in [11]). Given the prominent role of prokaryotes in the biosphere, defining the extent of microbial diversity found in microbial communities, as well as spatial and temporal changes in their composition and function have become increasingly more important to ecologists.

In this review, we discuss common pitfalls in research that has been done to characterize microbial diversity that are based on phylogenetic analyses of SSU rRNA, as well as some of the difficulties encountered in attempts to reconcile the phylogeny of organisms (based on SSU rRNA genes) and their taxonomic classification based on phenetic criteria.

The one-eyed king

Most studies of microbial community diversity are based on the extraction of total community DNA from samples followed by PCR amplification of SSU rRNA genes. While this approach sounds straightforward, it is, in fact, fraught with problems at almost every step along the way, from the extraction of DNA, to the selection of primers, right on through to the amplification of DNA [20–27]. The saying 'the one-eyed man is king in the land of the blind' seems to apply. While these methods are flawed and provide an incomplete and sometimes distorted view, it is far better than being completely blind.

Despite their technical limitations and biases, various approaches based on SSU rRNA genes from natural assemblages have proven to be quite useful to describe the structure of microbial communities. One such approach, namely the construction and analysis of clone libraries, provides detailed phylogenetic information about the members of communities. However, this approach is generally not well suited for the analysis of numerous samples because of the time and cost associated with the analysis of numerous clone libraries.

In studies whose aims are to understand spatial and temporal changes in community structure, the biogeography of prokaryotes, or how community structure changes in response to various perturbations, the ability to analyze numerous samples is critical. As part of such studies it is also important to devise statistically valid sampling schemes, and to have sufficient numbers of samples within 'treatment groups' so that specific hypotheses can be statistically tested. (It is embarrassing that the latter is almost entirely missing from studies of microbial diversity.) The failure to systematically analyze large numbers of samples to test specific hypotheses precludes microbial ecologists from testing ecological theories. Given current technology, these needs can only be addressed through the use of comparatively 'high-throughput' methods in which many samples can be processed simultaneously. Fingerprinting techniques, such as denaturing gradient gel electrophoresis (DGGE) [28] or terminal restriction fragment length polymorphism (T-RFLP) analysis [29–31] of SSU rRNA genes, offer the best compromise between the number of samples processed and the information obtained. Analyses of community structure by DDGE of SSU rRNA genes are hampered by the lack of satisfactory methods to quantify the results (through image processing). For these and other reasons it is difficult to compare the data obtained in different laboratories. However, a prime advantage of the method is its relative simplicity and low cost. T-RFLP analysis of SSU rRNA genes is currently one of the most powerful methods in microbial ecology for rapidly comparing the diversity of bacterial DNA sequences amplified by PCR from environmental samples. The method relies on variation in the position of restriction sites among sequences and determination of the length of fluorescently labeled terminal restriction fragments by high-resolution gel electrophoresis on automated DNA sequencers. The automated analysis of samples by capillary electrophoresis permits high sample throughput, and highly precise determination of fragment lengths. The data obtained can be compared with data from *in silico* analyses of sequence databases to infer the potential composition of samples. Importantly, the data are amenable to analyses using various statistical methods, such as similarity indices, hierarchical clustering algorithms, principal-component analyses, and self-organizing maps.

What you can't see with only one eye

In the paragraphs that follow we will briefly review some of the lesser-known (lesser discussed?) problems associated with methods used to characterize microbial diversity by cultivation-independent methods.

Universal primers are not universal

The choice of primers used in studies to assess the diversity of prokaryotes is nontrivial. Ideally, the primers used would be complimentary to the SSU-rRNA genes of all Eubacteria and Archaea. While these so-called 'universal' primers are widely used, they are far from 'universal'. For example, Baker *et al.* [32] determined the total number of sequences in the Ribosomal Database Project that are complimentary to various primers developed for the amplification of bacterial 16S rRNA genes (Table 1), and they also gauged their specificity. Two things become immediately clear upon perusal of these data: the primers differ dramatically in their 'universality' and most are not specific for Eubacteria. Similar difficulties are encountered in the selection of primers to amplify 16S rRNA genes from Archaea. Baker *et al.* [32] determined the complimentarity of 51 different archaeal primers to sequences of Kroarchaeota and Nanoarchaeota and found that 18 were complimentary to Kroarchaeota sequences, while only 11 were complimentary to Nanoarchaeota. Given this, whole groups might be excluded depending on the primers chosen.

Table 1

PCR primers commonly used for amplification of bacterial 16S rRNA genes [1].

Primer[a]	Sequence 5′ – 3′	Specificity			Matches[c]
		B	A	E	
E8F	AGAGTTTGATCCTGGCTCAG	+++	+	−	2424
E9F	GAGTTTGATCCTGGCTCAG	+++	+	++	2741
E334F	CCAGACTCCTACGGGAGGCAGGC	+++	−	−	13 172
E341F	CCTACGGGIGGCUGCA	+++	+	−	16 685
E786F	GATTAGATACCCTGGTAG	+++	+	−	12 616
E553R	TIACCGIIICTICTGGCAC	+++	+	++	18 724
E926R	CCGICIATTIITTTIAGTTT	+++	+++	++	19 950
E939R	CTTGTGCGGGCCCCCGTCAATTC	+++	−	−	8620
E1115R	AGGGTTGCGCTCGTTG	+++	−	−	9052
E1541R	AAGGAGGTGATCCANCCRCA	+++	+	−	1355

[a]Primer numbering relates to *E. coli* position complimentary to the 5′ end of the primer.
[b]Matches for Eubacteria (B), Archaea (A), and Eukarya (E) are represented as follows: −, no matches; +, <25 matches; ++, 25–100 matches; +++, >100 matches.
[c]Total number of matches in Ribosomal Database Project.

Degenerate primers (that have more than one nucleotide at a given location in the sequence), or primers with inosine at a given location (that can base-pair with all four nucleotides found in DNA) can be used to increase the universality of primers. In addition, the annealing temperature and the composition of reaction mixtures used for PCR can be manipulated to achieve the same end. While these parameters can be modulated in an attempt to increase 'universality' it is likely to be accompanied by a trade-off in specificity so that nontarget sequences are also amplified.

So why are so many universal primers not very universal? The obvious (and trivial) answer is that while there are conserved regions in the sequences of SSU rRNA genes, even the sequences of conserved regions are divergent. Consequently, the inability to identify truly universal primers is entirely due to the heterogeneity found in rRNA gene sequences. However, there are more insidious reasons. As reported by Baker *et al.* primers E8F and E9F were complimentary to comparatively few sequences, and this is disconcerting as they are commonly used. The mystery is partially solved when one aligns primer E8F to several sequences from the database. For example, we have aligned primer E8F with sequences of *Lactobacillus* strains, and found that roughly half would be scored as not being complimentary to the primer because the sequences of the region complimentary to B8F are unknown or ambiguous (Table 2)! The same explanation accounts for the low number of sequences that are complimentary to E1541R (data not shown). Obviously, assessments of primers should take into account the fact that there are many incomplete sequences in the database, and fair assessments of primers can only be made to those sequences that include the targeted region. Moreover, as the number of known sequences increases, researchers should continually reassess the specificity and utility of primers that were previously developed using a much smaller dataset.

It is also important to realize that just because a primer is complimentary to a large fraction of the sequences in the ribosomal database project (RDP) does not necessarily mean that it is a 'good' primer. While the number of SSU rRNA in databases has dramatically increased in recent years, they have not been collected in any systematic way. Moreover, if estimates of bacterial 'species' diversity are even close to being accurate, the sequences in the RDP database (~80,000) represent only a small fraction (<1%) of the total (~10 million; [18••]) that are estimated to exist. If these two facts are taken together, it becomes clear that efforts to develop 'universal' primers are crippled by the dearth of sequence data, and that various taxa could be under-sampled or entirely overlooked. In other words, you don't know what you don't know. However, in practice, we do the best we can with what we have available, and empirically test primers and PCR conditions to identify the combination that results in the highest apparent diversity with the kinds of samples being analyzed.

Language barriers

In discussions of biodiversity, it would be helpful if we all spoke the same language. Ecologists who study macrobiota define biodiversity in terms of species and differentiate diversity on three scales: α, β and γ. Alpha diversity is the diversity of species found within a site (local diversity), while a difference in species composition between sites is referred to as β diversity, and the diversity across a landscape (of all sites combined) is the gamma diversity. The diversity within a site is defined according to three parameters: species richness is the number of species in a given area, species evenness (equitability) is the relative abundance of various species,

Table 2

Alignment of Lactobacillus spp. 5′ end of 16S rRNA gene sequences.

8-Forward	----------AGAGTTTGATCCTGGCTCAG
Accession number	Aligned sequences
S000000152	----------GTGCCTAATACATGCAAGTCGAGCGCACTGGCCCAACTGATATGACGTGCTTGCACTGATTTGACGA
S000002363	----------GTGCCTAATACATGCAAGTCGAGCGCACTGGCCCAACAGAAATGACGTGCTTGCACTGATTTGACGT
S000005507	----------
S000015720	----------AGAGTTTGATNNTGGCTCAGGATGAACGCCGGCGGTGTGCCTAATACATGCAAGTCGAGCGCACTGGCCCAACTGATATGACGTGCTTGCACTGATTTGACGA
S000126953	----------AGAGTTTGATTATGGCTCAG-ATGAACGCCGGCGGTGTGCCTAATACATGCAAGTCGAGCGCACTGGCCCAACTGATATGACGTGCTTGCACTGATTTGACGA
S000007778	----------AGAGTTTGATCCTGGCTCAGGATGAACGCCGGCAGTGTGCCTAATACATGCAAGTCGTACGCACTGGCCCAACTAATTGATGGTGCTTGC--TGAATTGACGA
S000014648	----------AGAGTTTGATNNTGGCTCAGGATGAACGCCGGCGGTGTGCCTAATACATGCAAGTCGTACGCACTGGCCCAACTGATTGATGGTGCTTGCACCTGATTGACGA
S000001164	----------TTATATGAGAGTTTGATCCTGGCTCAGGATGAACGCCGGCGGTGTGCCTAATACNTGCNAGTCGAGCGCACTGGCCCANCTGATATGACGTGCTTGCACTNAATTGACGS
000019354	----------AGAGTTTGATCCTGGCTCAGGATGAACGCCGGCGGTGTGCCTAATACATGCAAGTCGAACGCGTTGGCCCAACTGATTGAACGTGCTTGCACGGACTTGACGT
S000011212	---------NNTTATATGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGTGTGCCTAATACATGCTAGTCGAACGCGTTGGCCCAATTGATTGATGGTGCTTGCACCTGATTGATTT
S000012733	----------
S000015716	----------
S000022065	----------NNAAAACGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAACCAACAGATTCA--------CTTCGGTGATGACGT
S000001162	--------NNNTANAATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCGGAACCAACAGATTTA--------CTTCGGTAATGACGT
S00000889	----------
S000009255	----------AGAGTTTGATNNTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCGGAACTAACAGATTTA--------CTTCGGTAATGACGT
S000011215	----------GAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGACGAGCGAGCAGAACCAGCACATCCA--------CTTCGGTAACGACGC
S000007776	----------AGAGTTTGATNNTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCAGAACTAACAGATCTA--------CTTCGGTAGTGACGT
S000109972	----------GGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCAGAACTAACAGATCTA--------CTTCGGTAGTGACGT
S000109599	----------GGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAACCAGCAGATTCA--------CTTCGGTGATGACGC
S000008510	---------NTCAAATTGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAATTCAAAGATCC---------CTTCGG-GGTGATTT
S000013906	---------NNCAAATTGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAATTCAAAGATTC---------CTTCGG-GAGGATTT
S000130019	----------CAAATTGAGAGTTNGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAATTCAAAGGACT---------CTTCGG-GGTGATTT
S000012578	----------G-ACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAATTCAAAGATYC---------CTTCGG-GRTGATTT
S000109603	----------TNCGNTCCTGGCTCAG-ACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTGAATTCAAAGATTC---------CTTCGG-GATGATTT
S000015357	--------NNTTAACATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCCGAACTAATTGATTA---------CTTCGGGTATGAAGT
S000003126	----------TGGCGGCGTGCTTAATACATGCAAGTCGAGCGAGCTTGCCTATTGAAATT---------CTTCGGAATGGACAT
S000088003	----------TGCCTAATACATGCAAGTCGAGCGAGCTTGCCTATAGAAGTT---------CTTCGGAATGGAAAT
S000130780	----------TGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTTGCCTATAGAAGTT---------CTTCGGAATGGAAAT
S000003516	----------NNAAAATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTTGCCTAGATGAATTTG-GTGCTTGCACCAGATGAAACT
S000006242	----------GGCGGCGTGCCTAATACATGCAAGTCGAGCGAGCTTGCCTAGATGATTTTA-GTGCTTGCACTAAATGAAACT
S000010822	----------
S000001993	----------NAAAATTGAGAGTTTGATCCTGGCTCAGGATNAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTGGTTAAGAGTRGCG-GTGCTTGCACCAAAGCRATTA
S000000429	---------NNATATATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCACTGACGTC---GACAGAAGGTGCTTGCAC-----TGGAAG
S000005109	---------NNTTATATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCACTGACGTC---GACCGAAGCTGCTTGCAG-----TGGACG
S000020742	---------NNTTAATCGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCACTCTCNTTTA-GATTGAAGGAGCTTGCTCCTC-ATTGATA
S000015718	---------NTTAATTTGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCACTCTCNTTTA-GATTGAAGGAGCTTGCTCCTC-ATTGATA

Table 2 Continued

```
S000004550 -------------------------------------GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTTTGGTC---GATGAACGGTGCTTGCACT----GWGATT
S000009879 -------------------------------------GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTTTGGTC---GATGAACGGTGCTTGCATC----GTGATT
S000010606 -------------------------------------GATSAACGCTSGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTCGTT---GATGATCGGTGCTTGCACC----GAGATT
S000015191 -------------------------------------GATSAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTCGTT---GATGATCGGTGCTTGCACC----GAGATT
S000010607 -------------------------------------GATSAACGSTSGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTCGTT---GATGATCGGTGCTTGCACC----GAGATT
S000109809 -------------------------------------GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTCGTT---GATGATCGGTGCTTGCACC----GAGATT
S000013699 ---------TTTTATATGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTTTGGTC---GATGAACGGTGCTTGCACT----GAGATT
S000016651 ---------TTTTATATGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTTTGGTC---GATGAACGGTGCTTGCACT----GAGATT
S000011211 ----------NTTATATGAGAGTTTGATCCTGGCTCAGGATNAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTGATT---ATTGAAAGGTGCTTGCATC----TTGATT
S000014980 -----------TNAAATGAGAGTNTGATCCTGGCTCNGGATNAACGCTGGCGGCGTGCCTNATNCNTGCNAGTCGAACGAGTNTTGGTC---GATGAACGGTGCTTGCTCT----NNNATT
S000000171 -----------------------------------------ACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCAATCT-TTGACC---AATGAGTGCTTGCACT-----CAGCN
S000129411 ------------------------------------------------TGGCGGCGTGCCTAATACATGCAAGTCGAACGCAATCT-TTGACT---AATGAGTGCTTGCACT-----CAGCG
S000005786 -CCGAATTCGTCGACAACAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAACTT-TCTTAC---ACCGAATGCTTGCRTT-----CA-TC
S000021671 ----------TTAAAATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAACTT-TCTTAC---ACCGAATGCTTGCATT-----CACTC
S000000428 ---------NNTAAATTGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGAATT-TCTTAC---ACCGAGTGCTTGCACT-----CA-CC
S000004337 ---------NCTAAAATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCTTTTT-TCAATC---ATCGTA-GCTTGC-TA-----CACCG
S000012369 ---------NATAAATTGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAGCTT-TCTTTC---ACCGAATGCTTGCATT-----CACCG
S000007322 ---------NTTNAATTGAGAGTTTGATCCTGGCTCAGGATNAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAACTTCTTTATC---ACCGAGTGCTTGCACT-----CACCG
S000021282 ---------NTTAAATTGAGAGTTTGATCCTGGCTCAGGATNAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAACTTCTTTATC---ACCGAGTGCTTGCACT-----CACCG
S000013439 ------------------------------------------------------------------------------------------------------------CT-----CACCG
S000000239 ---------NNTAATATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAACCATCCTGAA----GATTGAAGCTTGCTTCA----TGATT
S000013905 ----------NTAATATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAACTTTCCTATT----NATTGATGCTTGCATCA----TGATT
S000008817 -------------------------------------GACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAACCATCCTGAA----GATTGAAGCTTGCTTCA----TGATT
S000108883 ----------------------------------AGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAACTCTGGTAAT----GATTGGTGCTTGCATCA----TGAAT
S000109455 -------------------------------------GACGAACGCTGSCGGCGTGCCTAATACATGCAAGTCGAACGAACTCTGGTATT----GATTGGTGCTTGCATCA----TGATT
S000001896 ----------------------------------------GAACNCTAGCGGCGTGNCTAACACATGCAAGTCGAACGAACTCTGGTATT----GATTGGTGCTTGCATCA----TGACT
S000131235 ----------TTAATTTGAGAGTTNGATCCTGGCTCAGGACGAACGCTNGCGGCGTGCCTAATACATGCAAGTCGAACGANCTCTNGTATT----NATTGGTGCTTGCATCA----TGATT
S000005355 ------------------------TGATCCTGGCTCAGGATGAACCCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCATCCCGTTAA-----ATCAAGTGCTTGCA-CG----GATTT
S000016063 ------------------AGAGTTTGATCATGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGATTTTATTA-----ATTGATTACTTCGG-TA----TGATT
S000001003 ------------------AGAGTTTGATNNTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAGCTGCGCCTAATGATAGTTGATGCTTGCATTAGCTTGACTT
S000014646 ---------NCTAAAATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAGCTGCGCCTAATGATAGTTGATGCTTGCATTAGCTTGACTT
S000003125 ------------------AGAGTTTGATNNTGGCTCAGGACGAACGTTGGCGGCGTGCCTAATACATGCAAGTCGAACGAAGTCGCCCAATTGATTCTTAGTGCTTGCACTAAGATGATTT
S000005113 ------------------AGAGTTTGATNNTGGCTCAGGACGAACGTTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGGTCTCCTAACTGATAGCTGGTGCTTGCATCAGCTTGACGA
S000004837 -------------------------------------GACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGCTCTCCCAAATTG-ATTTTATGCTTGCATAAATGATTTTT
S000000624 ----------ATAAGATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACATGCAAGTCGAACGAGCTTCCGTTGA----ATGACGTGCTTGCACT----GATTTC
S000002726 -----------TAAGATGAGAGTTNGATCCTGGCTCAGGACGAACGCTGGCGGCATGCCTAATACNTGCAAGTCGAACGAGCTTCCGTTGA----ATGACGTGCTTGCACT----NATTTC
S000002357 ------------------------------------------------------------------------------------------------------------------------
S000011965 ---------NNTAAGATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGCGTCTCCGTTRATGAYTTTARGTGCTTGCAYTTGAAAGATTT
S000010035 ----------NTCAGATGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATNCATGCAAGTCGAACGCGTCTTGGTCAATGAAGTTGAGTGCTTGCATTTAACTNATTT
S000001598 -----------NNAATGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGGGA-ATCTTCGG----------------------------
S000008151 ----------NNAGATGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGGGNTGCCATTAG----------------------------
```

and species composition is the actual species that are present in a sample. Many investigators of microbial biodiversity employ the term 'diversity'; and make statements about a change or lack of change in diversity. These statements are, generally, imprecise, and not supported by the data. The structures of microbial communities are usually highly skewed such that a few species are abundant, and many are rare. By the very nature of the methods used (PCR amplification of SSU rRNA genes) the analyses done are limited to only the numerically abundant populations targeted by the primers used. Since PCR amplification of DNA is a competitive enzymatic reaction, the SSU rRNA templates in a sample are amplified roughly in accordance to their abundance. Consequently, the SSU rRNA genes of the numerically abundant populations are the most abundant amplicons following PCR and the ones represented in the community profiles obtained by DGGE or T-RFLP analyses of the amplicons mixture. Populations that constitute less than ~1% of the total community (yet may still be present in numbers $>10^5$ per g) are generally not represented in such profiles, and so this represents a threshold of detection. As a result, the actual species richness of a community (or phylogenetic group depending on the primers used) remains unknown and may be impossible to precisely determine. It can, however, be estimated using various mathematical approaches [33,34••]. If investigators see differences in profiles of SSU rRNA genes from communities, they often infer or conclude that there is a difference in species richness due to more or fewer DNA fragments; in other words, that one microbial community is less diverse than another microbial community. In reality, such differences could be entirely due to differences in the rank-abundance of populations with no change in species richness. While microbial ecologists are certainly aware of the high numbers and skewed distribution of species in a community, they are too often lax in their interpretations of differences in community structure as reflected in community profiles.

Representative: what matters most

The isolation of genomic DNAs from samples is one of the first steps in studies of microbial diversity using cultivation independent methods. In doing so it is obviously important to obtain genomic DNAs that are representative of the microbial communities present in samples. In other words, it is important that the efficiency of cells lysis be approximately the same for all taxa present in the sample so that all populations are fairly represented. The amount of DNA recovered is of almost no consequence, so long as the amount obtained is sufficient for subsequent analyses, and the quality permits subsequent manipulations (e.g. amplification by PCR). Ironically, most efforts made to improve or optimize the extraction of genomic DNA from environmental samples focus on the yield of DNA, and there is almost no heed given to whether the microbial populations are fairly represented in the sample [35–38]. While it is not apparent how one could determine if extracted DNA was representative in a simple, straightforward way, investigators should be aware of the possibility that a significant bias may have been introduced into their analyses during the isolation of genomic DNA and interpret the data accordingly.

Apples and oranges

The sequences of newly discovered SSU rRNA molecules are typically compared to those previously described by using various alignment tools and phylogenetic algorithms, then the results are graphically represented in dendrograms in which the branch lengths reflect the genetic distances between sequences. Quite naturally researchers are curious to know which known microbial taxa (species) are related to the newly discovered ones; and the urge to name the new populations using a Latin binomial is often irresistible. But this is a bit like comparing apples and oranges. Nowadays the taxonomic classification of bacteria is accomplished using a polyphasic approach [39] in which a spectrum of phenotypic criteria are determined and used to ascertain the similarities and genotypic differences among cultivated isolates. Ultimately, a definitive classification of prokaryotes requires that the genetic relatedness of new organisms be compared to type strains by DNA:DNA annealing of genomic DNAs. According to Wayne *et al.* [40] to be considered the same species the genomes of two strains must have more than 70% DNA:DNA relatedness and with 5°C or less ΔT_m. Thus, while the methodologies used to characterize prokaryotes has become more sophisticated and elaborate over the years, it remains an essentially Linnaean classification scheme based on a phenetic species concept. By contrast, the classification of prokaryotes based on the phylogeny of gene sequences (such as SSU rRNA genes), inherently employs a phylogenetic concept of species. These are two distinctly different concepts.

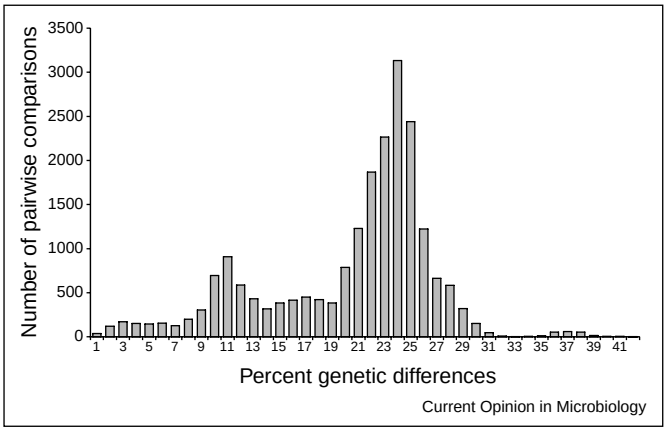
Since prokaryotes are asexually reproducing organisms, vertical inheritance predominates, and evolutionary mechanisms result in incremental changes in the genotypic and phenotypic characteristics of organisms. Given this, it would be expected that the phenetic criteria used for bacterial classification would be largely conserved among organisms that are related by evolutionary descent. Fortunately, this expectation is born out in so far as members of a phylogenetic clade share many common phenotypic traits. However, evolutionary changes resulting in the gain, loss, or modification of functions can and do occur along the course of evolutionary lineages. As a result, even strains that were recently derived from a common ancestor, and are highly related based on the phylogeny of specific genes, can differ in terms of the phenotypic criteria used for their classification. Indeed, disparities in the phenotypic traits of strains belonging to a single species are not uncommon, and in some cases can be numerous — even among strains of familiar and easily

recognizable species such as *Escherichia coli* [41]. The situation is further muddied because traits can also be acquired by the horizontal transfer of genes. The exchange of large chromosome segments can sometimes occur, but more often the exchange is more limited (<200 kb) and mediated by vectors such as bacteriophages and plasmids, as well as by the transformation of naked DNA. The donor and recipient organisms need not be phylogenetically close. Horizontal gene transfer in one form or another occurs at all taxonomic levels, and any given species may have acquired 20–25% of its genome via horizontal transfer. This reticulated evolutionary history is the basis for findings of phylogenetic incongruence among genes [42] and can readily lead to phylogenetically related strains that differ in terms of overall genetic relatedness and certain phenotypic properties that are important for taxonomic classification based on phenetic criteria. In other words, taxonomic groups based on phenetic criteria can be polyphyletic. By contrast, the purpose of schemes based on the phylogeny of genes is to classify organisms into monophyletic groups. Given the differences in these two approaches for microbial classification, and the mechanisms used for the adaptive evolution of prokaryotes, it is not surprising that there can be quite different outcomes.

In 1994, Stackebrandt and Goebel [43] reported studies they had done to compare the results of DNA–DNA reannealing to the relatedness of strains based on the similarity of 16S rRNA gene sequences. They concluded that if 16S rRNA sequences have less than 97% identity then the sequences are most likely to be derived from different species, and if the sequences have more than 97% identity, then they may either come from the same species or from different species. In the latter case, they suggested that DNA–DNA reannealing data would be needed to clarify the species relationships in such cases. Although this conclusion is clearly stated, many researchers have misinterpreted their findings and assert that if the 16S rRNA gene sequences of two strains are >97% identical then one can conclude that they are come from the same species. This is incorrect. Several recent examples of the misinterpreted '97% rule' can be found in the literature by searching in Science Citation Index for papers that reference the Stackebrandt and Goebel article. Looking for articles in 2003 yielded four papers that explicitly mis-state the rule in the way described above [44–47]. One of these papers even mentions that the rule seems to be incorrect for the genus of bacteria from which their sequences were derived because their studies included different species but there was >97% sequence similarity among their 16S rRNA genes [47].

To illustrate why it is important to accurately interpret Stackebrandt and Goebel's '97% rule', we determined the pair wise genetic distances among 206 type strains of bacteria and one archaeal species based on high quality, full-length 16S rRNA sequences from the RDP [48]. The sequences were aligned using ClustalW and the pair wise genetic distances were calculated using the Jukes Cantor correction for multiple changes as implemented in the 'distances' program of GCG (Accelrys). Three subspecies were included in the analyses so according to the misinterpreted '97% rule', there should be, at most, three genetic distances less than or equal to 3%. However, the results of our analysis indicate that there were 329 distances between 0 and 3%; not three as predicted (Figure 1). We also determined whether certain genera disproportionately contributed to this large number of closely related sequences. In Table 3, genera are sorted by the minimum pair wise genetic distances within a genus. Twenty-one of the thirty-one genera examined had at least one distance below 3%, and six of the sixteen genera with more than two species have greater than half of their distances less than 3% (see median column of Table 3). Thus, it is not at all uncommon for the 16S

Figure 1



Distribution of pairwise genetic differences among 16S rRNA sequences of 206 bacterial type strains.

Table 3

Summary statistics for pair wise genetic distances between species of genera.

Genus[a]	No. Sequences	% Difference Statistics			
		MIN	MAX	AVG	MEDIAN
Pseudomonas	31	0.07	7.66	3.65	3.36
Mycobacterium	12	0.14	6.25	3.57	4.08
Amycolatopsis	6	0.68	3.40	2.28	2.57
Micromonospora	12	0.75	2.88	1.94	2.03
Bacillus	6	0.76	6.42	3.49	4.38
Aeromonas	4	1.07	3.47	2.11	2.06
Streptomyces	9	1.08	6.91	4.17	4.14
Clostridium	3	1.48	1.76	1.60	1.48
Thermoactinomyces	5	1.71	10.54	6.87	7.03
Promicromonospora	3	1.71	2.25	1.94	1.71
Nocardia	9	1.83	4.55	3.16	3.19
Kibdelosporangium	3	1.83	4.75	3.67	1.83
Comamonas	3	2.32	2.68	2.49	2.32
Bifidobacterium	3	3.04	5.31	4.29	3.04
Corynebacterium	4	5.44	7.59	6.26	6.08
Actinomyces	3	8.65	9.77	9.30	8.65
Thermoleiphilum	2			0.07	
Tsukamurella	2			0.68	
Alicyclobacillus acidocaldarius	2			0.88	
Knoellia	2			1.36	
Leifsonia	2			2.39	
Caryophanon	2			2.79	
Kocuria	2			2.93	
Arthrobacter	2			2.99	
Thermomonas	2			3.47	
Vibrio	2			5.27	
Dermatophilus	2			5.49	
Cytophaga	2			5.71	
Nocardioides	2			6.44	
Facklamia	2			7.95	
Desulfovibrio	2			11.87	

Genera in red font do not have any pairwise genetic distances among species that are less than 3%.

rRNA sequences of distinct bacterial species to be ≥97% similar, and this appears to be true of species from all bacterial phyla. Given this, conclusions that two strains are the same species based on highly similar 16S rRNA gene sequences are subject to type 2 errors (concluding they are the same species when in fact they are different species). This contradicts Fox *et al.* [49] who explicitly stated that although 16S rRNA is not a suitable method for determining new species, it can be used for identifying strains of known species. The interests of investigators and the expectations of editors to name organisms has led to the creation of a '97% rule' that is inappropriate and inaccurate for many prokaryotic species. The problem is accentuated because microbiologists have not reached a consensus concerning the classification of prokaryotes at the species level based on phylogenetic data.

Data from studies on the phylogeny of prokaryotes has informed microbial taxonomists and stimulated them to reconsider and revise the classification of prokaryotic genera and species. For example, Wen *et al.* [50] determined the 16S rRNA sequence of *Comamonas acidovorans* (formerly *Pseudomonas acidovorans*), and found that it was phylogenetically distant from the type species of the genus, *Comamonas terrigena*. On the basis of this and other data, a new genus, *Delftia* was created, and *Delftia acidovorans* ATCC15668T was made the type species. Similarly, Gosink *et al.* [51] found that the 16S rRNA gene sequence of *Flectobacillus glomeratus* was sufficiently dissimilar to that of the type species, *Flectobacillus* major, that a new genus, *Polaribacter*, was formed. These attempts to clarify the taxonomy of prokaryotes through the inclusion of phylogenetic data are admirable, but at the same time it is problematic since the investigators are unwittingly amalgamating two species concepts. While we can readily determine the degree of similarity and difference between the gene sequences of strains, we have yet to decide how this information can be used as the basis for a taxonomic classification scheme. Until we do, the comparison of phylogenetic data and phenetic data will remain a comparison of apples and oranges.

While few would argue that there are recognizable clusters of organisms in nature that are re-identifiable by

virtue of their shared characteristics, there is considerable debate over whether all organisms can be defined within a single conceptual framework that reflects a 'natural order'. The search for a unifying concept for the classification of organisms has led to more than 20 proposed species concepts. The Biological Species Concept is among the most familiar of these, and it rests on the premise that species can be defined as groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups and implies a closed gene pool of a Mendelian population. This is problematic for the classification of asexually reproducing organisms (prokaryotes, as well as many protist and plant species) but is widely used by zoologists. Within this framework some have argued that asexually reproducing organisms do not form species at all [52]. The inability to classify so much of the biological world using the framework provided by the Biological Species Concept, leads one to conclude that it is not generally useful for understanding the extent and distribution of biological diversity, for it does not provide a common language. On the other hand, the solutions offered by bacterial taxonomists are also not particularly useful. As mentioned above, they advocate that to be considered the same species the genomes of two strains must have more than 70% DNA:DNA relatedness and with 5°C or less ΔT_m. The philosophical basis for this definition is obscure; it also suffers from the fact that it is not transitive, and difficult to apply in practice. Finally, the definition is exceptionally coarse since application of the same definition to eukaryotes would lead to the inclusion of members of distinct taxa into the same species in a way that is ludicrous, for example, humans, orangutans and gibbons would belong to the same species [53].

Conclusions

The incongruencies that arise through commonly used approaches to classify prokaryotes leads one to consider the validity of pluralism and antirealism — that there are multiple, independent, theoretically valid means by which organisms can be classified. If this is true then a given organism can be part of several different species taxa, one for each different species concept, thus permitting the existence of many alternative species taxonomies' [54••]. This view forms the roots of the general lineage concept [54••], and is not inconsistent with ideas on how life evolved on Earth. It simply argues that membership to a 'species' is a relational phenomenon that is entirely dependent on human perspective, and reconciliation of taxonomic groupings that are based on distinct concepts of species may well be a futile exercise. This could be especially true for the classification of prokaryotes because of the complex and varied means by which they have evolved. This complexity was recently reflected in a report from a colloquium sponsored by the American Academy of Microbiology [55••] that stated:

'The natural microbial world can be viewed as a landscape of genes and genome ecology, in which organisms exchange genetic information and co-evolve with one another, shaping themselves and the biosphere over time. Microbial genomic evolution is crafted in microbial communities through the dynamic interplay of mutation, genetic drift, gene transfer, and natural selection.'

So while the classification of prokaryotes can at times seem hopeless, it is likely that the relationships among populations will become better understood as we gain more knowledge of how prokaryotic genomes have evolved, and the ecological processes that have shaped these processes.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. DeLong EF, Wu KY, Prezelin BB, Jovine RV: **High abundance of archaea in antarctic marine picoplankton**. *Nature* 1994, **371**:695-697.
2. Kuske CR, Barns SM, Busch JD: **Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions**. *Appl Environ Microbiol* 1997, **63**:3614-3621.
3. Ward DM, Weller R, Bateson MM: **16S rRNA sequences reveal numerous uncultured microorganisms in a natural community**. *Nature* 1990, **345**:63-65.
4. Miteva VI, Sheridan PP, Brenchley JE: **Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core**. *Appl Environ Microbiol* 2004, **70**:202-213.
5. Marteinsson VT, Hauksdottir S, Hobel CFV, Kristmannsdottir H, Hreggvidsson GO, Kristjansson JK: **Phylogenetic diversity analysis of subterranean hot springs in Iceland**. *Appl Environ Microbiol* 2001, **67**:4242-4248.
6. D'Hondt S, Rutherford S, Spivack AJ: **Metabolic activity of subsurface life in deep-sea sediments**. *Science* 2002, **295**:2067-2070.
7. Edwards KJ, Bond PL, Gihring TM, Banfield JF: **An Archaeal iron-oxidizing extreme acidophile important in acid mine drainage**. *Science* 2000, **287**:1796-1799.
8. Kroes I, Lepp PW, Relman DA: **Bacterial diversity within the human subgingival crevice**. *Proc Natl Acad Sci USA* 1999, **96**:14547-14552.
9. Wheelis ML, Kandler O, Woese CR: **On the nature of global classification**. *Proc Natl Acad Sci USA* 1992, **89**:2930-2934.
10. Woese CR: **Bacterial evolution**. *Microbiol Rev* 1987, **51**:221-271.
11. Woese CR: **Default taxonomy: Ernst Mayr's view of the microbial world**. *Proc Natl Acad Sci USA* 1998, **95**:11043-11046.
12. DeLong EF, Pace NR: **Environmental diversity of bacteria and archaea**. *Syst Biol* 2001, **50**:470-478.
13. Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA: **Microbial ecology and evolution: a ribosomal RNA approach**. *Annu Rev Microbiol* 1986, **40**:337-365.
14. Hugenholtz P, Goebel BM, Pace NR: **Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity**. *J Bacteriol* 1998, **180**:4765-4774.
15. Ludwig W, Bauer SH, Bauer M, Held I, Kirchhof G, Schulze R, Huber I, Spring S, Hartmann A, Schleifer KH: **Detection and**

***in situ* identification of representatives of a widely distributed new bacterial phylum**. *FEMS Microbiol Lett* 1997, **153**:181-190.

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

17. Dunbar J, Barns SM, Ticknor LO, Kuske CR: **Empirical and theoretical bacterial diversity in four Arizona soils**. *Appl Environ Microbiol* 2002, **68**:3035-3045.

18. •• Curtis TP, Sloan WT, Scannell JW: **Estimating prokaryotic diversity and its limits**. *Proc Natl Acad Sci USA* 2002, **99**:10494-10499.
The authors argue that the diversity of prokaryotic communities may be related to the ratio of two measurable variables: the total number of individuals in the community and the abundance of the most abundant members of that community. These variables are used to estimate prokaryotic diversity on large scales.

19. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: The unseen majority**. *Proc Natl Acad Sci USA* 1998, **95**:6578-6583.

20. Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, Catroux G: **DNA extraction from soils: Old bias for new microbial diversity analysis methods**. *Appl Environ Microbiol* 2001, **67**:2354-2359.

21. Polz MF, Cavanaugh CM: **Bias in template-to-product ratios in multitemplate PCR**. *Appl Environ Microbiol* 1998, **64**:3724-3730.

22. Ishii K, Fukui M: **Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR**. *Appl Environ Microbiol* 2001, **67**:3753-3755.

23. Egert M, Friedrich MW: **Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure**. *Appl Environ Microbiol* 2003, **69**:2555-2562.

24. Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J: **Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning**. *Appl Environ Microbiol* 2001, **67**:880-887.

25. Hugenholtz P, Huber T: **Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases**. *Int J Syst Evol Microbiol* 2003, **53**:289-293.

26. Lueders T, Friedrich MW: **Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts**. *Appl Environ Microbiol* 2003, **69**:320-326.

27. Suzuki MT, Giovannoni SJ: **Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR**. *Appl Environ Microbiol* 1996, **62**:625-630.

28. Muyzer G, de Waal EC, Uitterlinden A: **Profiling of complex microbial populations using denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA**. *Appl Environ Microbiol* 1993, **59**:695-700.

29. Kitts CL: **Terminal-restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics**. *Curr Issues Intest Microbiol* 2001, **2**:17-25.

30. Osborn AM, Moore ERB, Timmis KN: **An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics**. *Environ Microbiol* 2000, **2**:39-50.

31. Liu WT, Marsh TL, Cheng H, Forney LJ: **Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA**. *Appl Environ Microbiol* 1997, **63**:4516-4522.

32. Baker GC, Smith JJ, Cowan DA: **Review and re-analysis of domain-specific 16S primers**. *J Microbiol Methods* 2003, **55**:541-555.

33. Dunbar J, Barns SM, Ticknor LO, Kuske CR: **Empirical and theoretical bacterial diversity in four Arizona soils**. *Appl Environ Microbiol* 2002, **68**:3035-3045.

34. •• Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM: **Counting the uncountable: Statistical approaches to estimating microbial diversity**. *Appl Environ Microbiol* 2001, **67**:4399-4406.
The authors compare the utility of various statistical approaches for assessing the diversity of microbial communities.

35. Duarte GF, Rosado AS, Seldin L, Keijzer-Wolters AC, van Elsas JD: **Extraction of ribosomal RNA and genomic DNA from soil for studying the diversity of the indigenous bacterial community**. *J Microbiol Methods* 1998, **32**:21-29.

36. Frostegård A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P: **Quantification of bias related to the extraction of DNA directly from soils**. *Appl Environ Microbiol* 1999, **65**:5409-5420.

37. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ: **Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition**. *Appl Environ Microbiol* 2000, **66**:5488-5491.

38. Hurt RA, Qiu X, Wu L, Roh Y, Palumbo AV, Tiedje JM, Zhou J: **Simultaneous recovery of RNA and DNA from soils and sediments**. *Appl Environ Microbiol* 2001, **67**:4495-4503.

39. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J: **Polyphasic taxonomy, a consensus approach to bacterial systematics**. *Microbiol Rev* 1996, **60**:407-438.

40. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Murry RGE, Stackebrant E, Starr MP, Trüper HG: **Report of the ad hoc committee on reconciliation of approaches to bacterial systematics**. *Int J Syst Bacteriol* 1987, **37**:463-464.

41. Souza V, Rocha M, Valera A, Eguiarte LE: **Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents**. *Appl Environ Microbiol* 1999, **65**:3373-3385.

42. Daubin V, Nancy A, Moran NA, Ochman H: **Phylogenetics and the cohesion of bacterial genomes**. *Science* 2003, **301**:829-832.

43. Stackebrandt E, Goebel BM: **Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology**. *Int J Syst Bacteriol* 1994, **44**:846-849.

44. Zaidi BR, Hinkey LM, Rodrýguez NR, Govind NS, Imam SH: **Biodegradation of toxic chemicals in Guayanilla Bay, Puerto Rico**. *Marine Pollution Bulletin* 2003, **46**:418-423.

45. Kisand V, Wikner J: **Combining culture-dependent and -independent methodologies for estimation of richness of estuarine bacterioplankton consuming riverine dissolved organic matter**. *Appl Environ Microbiol* 2003, **69**:3607-3616.

46. Spear JR, Ley RE, Berger AB, Pace NR: **Complexity in natural microbial ecosystems: The Guerrero Negro experience**. *Biol Bull* 2003, **204**:168-173.

47. Yassin AF, Kroppenstedt RM, Ludwig W: ***Corynebacterium glaucum* sp. nov**. *Int J Syst Evol Microbiol* 2003, **53**:705-709.

48. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM: **The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy**. *Nucleic Acids Res* 2003, **31**:442-443.

49. Fox GE, Wisotzkey JD, Jurtshuk P Jr: **How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity**. *Int J Syst Bacteriol* 1992, **42**:166-170.

50. Wen A, Fegan M, Hayward C, Chakraborty S, Sly LI: **Phylogenetic relationships among members of the *Comamonadaceae*, and description of *Delftia acidovorans* (den Dooren de Jong 1926 and Tamaoka *et al.* 1987) gen. nov., comb. nov**. *Int J Syst Bacteriol* 1999, **49**:567-576.

51. Gosink JJ, Woese CR, Staley JT: ***Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov. and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of '*Flectobacillus glomeratus*' as *Polaribacter glomeratus* comb. nov**. *Int J Syst Bacteriol* 1998, **48**:223-235.

52. Hull DL: **Individuality and selection**. *Annu Rev Ecol Syst* 1980, **11**:311-332.

53. Sibley CG, Ahlquist JE: **DNA hybridization evidence of hominoid phylogeny: results from an expanded data set**. *J Mol Evol* 1987, **26**:99-121.

54. de Queiroz K: **The general lineage concept and the defining** •• **properties of the species category**. In: Species: New Interdisciplinary Essays, RA. Wilson, ed., MIT Press 1999, p. 49-89.
This book contains a number of interesting and provocative essays on the classification of organisms and species concepts.

55. Buckley MR: **Global Genome Question: Microbes as the key to** •• **understanding evolution and ecology**. ASM Press 2004.
This report is based on a colloquium, ''The Global Genome Question: Microbes as the Key to Understanding Evolution and Ecology,'' sponsored by the American Academy of Microbiology. The report describes the potential of genomic approaches to advance our understanding of microbial communities and ecosystems.