npg

ORIGINAL ARTICLE

Diversity rankings among bacterial lineages in soil

Noha H Youssef and Mostafa S Elshahed

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

We used rarefaction curve analysis and diversity ordering-based approaches to rank the 11 most frequently encountered bacterial lineages in soil according to diversity in 5 previously reported 16S rRNA gene clone libraries derived from agricultural, undisturbed tall grass prairie and forest soils ($n = 26,140$, $28\,328$, $31\,818$, $13\,001$ and $53\,533$). The *Planctomycetes*, *Firmicutes* and the δ-*Proteobacteria* were consistently ranked among the most diverse lineages in all data sets, whereas the *Verrucomicrobia*, *Gemmatimonadetes* and β-*Proteobacteria* were consistently ranked among the least diverse. On the other hand, the rankings of α-*Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes* and *Chloroflexi* varied widely in different soil clone libraries. In general, lineages exhibiting largest differences in diversity rankings also exhibited the largest difference in relative abundance in the data sets examined. Within these lineages, a positive correlation between relative abundance and diversity was observed within the *Acidobacteria*, *Actinobacteria* and *Chloroflexi*, and a negative diversity–abundance correlation was observed within the *Bacteroidetes*. The ecological and evolutionary implications of these results are discussed.

Introduction

Owing to its extremely important role in economic sustainability as well as for global nutrient cycling, bacterial diversity in soil has extensively been investigated using culture-independent 16S rRNA gene-based analysis (for example, Borneman and Triplett, 1997; Dunbar *et al.*, 2002; Tringe *et al.*, 2005; Costello and Schmidt, 2006; Schloss and Handelsman, 2006; Roesch *et al.*, 2007; Elshahed *et al.*, 2008). Collectively, these studies have revealed that, on a species level, soil is an extremely diverse habitat, with species richness estimates ranging between 2000 and 52 000 (for example, Tringe *et al.*, 2005; Roesch *et al.*, 2007). In addition to the highly diverse nature of individual soil samples, soil bacterial populations show immense global diversity at the putative species and strain levels. Within the thousands of soil studies conducted so far, no two clone libraries have exactly the same community composition, and exact (100%) sequence matches between the most abundant species and database-deposited sequences are very

rarely encountered (Janssen, 2006; Elshahed *et al.*, 2008).

However, on a higher taxonomic level, soil exhibits a remarkably stable community structure, in which the absolute majority of clones (92% in a recent compilation of 32 different 16S rRNA gene soil clone libraries (Janssen, 2006)) always belongs to nine major bacterial phyla: the *Proteobacteria* (mainly the α-, β- and δ-subdivisions), *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Bacteroidetes*, *Firmicutes*, *Planctomycetes*, *Verrucomicrobia* and *Gemmatimonadetes*, out of an estimated 104 candidate phyla recognized at present according to Hugenholtz taxonomic framework in Greengenes database (July 2008) (DeSantis *et al.*, 2006a, b). This phylum-level community structure stability is remarkable, considering the wide variations in temperature, pH, soil properties, land usage regimes and vegetation among various soils.

We hypothesized that these previously outlined environmental and physicochemical variations in soils are mainly reflected in differences in the community composition (that is, relative abundance and diversity), mainly within these 11 major bacterial lineages (including the three Proteobacterial lineages). Relative abundance values (% representation of a specific lineage in a clone library) and abundance-based rankings are easily obtained and have been reported in many previously described 16S rRNA gene clone library studies (for example, Borneman and Triplett, 1997; Kuske *et al.*, 1997;

Correspondence: MS Elshahed, Department of Microbiology and Molecular Genetics, Oklahoma State University, 1110 S Innovation Way, Stillwater, OK 74074, USA.
E-mail: mostafa@okstate.edu

Nogales *et al.*, 2001; Axelrood *et al.*, 2002; Lipson and Schmidt, 2004; Roesch *et al.*, 2007; Lesaulnier *et al.*, 2008). On the other hand, comparing and ranking the diversity of various lineages within a soil sample and determining whether these relative diversity patterns are affected by soil physicochemical characteristics, soil treatments or historic land usage has never been reported previously.

Ranking phyla according to diversity within a single soil sample has been hampered by the small sample size of most 16S rRNA libraries available at present. Most available soil data sets derived from a single soil sample are relatively small, for example, <1000 clones, and larger data sets are usually a composite of multiple samples examining different treatments within the same ecosystem (Lesaulnier *et al.*, 2008; Talera *et al.*, 2008). Given that some of these 11 lineages sometimes represent less than 1% of the total sequences in a soil clone library (Schloss and Handelsman, 2006; Roesch *et al.*, 2007), using a large data set is imperative to obtain a reasonable size subclone library for diversity estimates in all bacterial lineages. Recently, two studies reported on the bacterial community in five different soils using either near full-length sequences (Elshahed *et al.*, 2008) or short pyrosequencing-generated fragments (Roesch *et al.*, 2007). These five clone libraries are significantly larger than all previously reported 16S rRNA gene soil clone libraries, and thus present an opportunity to examine the relative diversity and rankings of various bacterial lineages in a diverse set of soils. Here, we report on using multiple sample size-unbiased approaches (single diversity indices, rarefaction curve analysis and diversity-ordering approaches) to determine the richness, evenness and diversity of the 11 major bacterial lineages (including the three Proteobacterial lineages) in each of these five data sets and report on the characteristic diversity ranks and diversity–abundance correlations observed, as well as their implications on ecosystem functions.

Materials and methods

Soil data sets
Five data sets, ranging in size between 13 001 and 53 533, were used in this study. Kessler Farm Soil clone library (KFS, $n = 13\,001$) was generated using 16S rRNA gene primers 27F and 1391R from an undisturbed tall grass prairie preserve in central Oklahoma (Elshahed *et al.*, 2008). The other four clone libraries consist of pyrosequencing-generated fragments (average length 106 bp), generated using primer pair 1391F and 1525R (Roesch *et al.*, 2007). The clone libraries originated from a maize field in southern Brazil (Brazil clone library, $n = 26\,140$), a sugarcane field in Florida (Florida clone library, $n = 28\,328$), an agricultural plot at the University of Illinois (Illinois clone library, $n = 31\,818$) and a

boreal forest in Canada (Canada clone library, $n = 53\,533$). Details on the soil characteristics, sampling, DNA extraction, PCR amplification procedures, cloning (if applicable) and sequencing were described in the original manuscripts (Roesch *et al.*, 2007; Elshahed *et al.*, 2008).

Alignment, classification and operational taxonomic unit (OTU) assignment of clones in KFS library were reported previously in detail (Elshahed *et al.*, 2008). In brief, after detecting and excluding potential chimeras using Mallard (Ashelford *et al.*, 2006), the sequences were grouped into phyla, and each phylum was aligned using Greengenes NAST-aligner (DeSantis *et al.*, 2006a, b). Distance matrices were created for each group using the 'create distance matrix' function in Greengenes webserver (DeSantis *et al.*, 2006a, b) and the matrix generated was used to assign all clones to OTUs at a 97% sequence similarity cutoff using DOTUR (Schloss and Handelsman, 2005).

Sequences of each of the four pyrosequencing-generated Brazil, Florida, Illinois and Canada clone libraries were downloaded from Genbank database (accession numbers available in the original manuscript (Roesch *et al.*, 2007)), and sequences from each library were sequentially aligned, and classified using Greengenes aligner and Classifier programs, respectively (DeSantis *et al.*, 2006a, b). Greengenes-based alignments and classifications were simultaneously conducted for all five clone libraries in a span of 2 weeks (April 2008) to exclude or at least minimize the effect of periodic update of the database on the classification results. Aligned sequences belonging to each of the target lineages were extracted from the master alignment file using a perl-script (available upon request). Each lineage-specific file was used to create a distance matrix using the 'create distance matrix' function in Greengenes webserver (DeSantis *et al.*, 2006a, b). The output matrix was used for OTU assignment at 97% sequence similarity cutoffs using DOTUR (Schloss and Handelsman, 2005). DOTUR output files were used as a starting point for rarefaction curve construction and computing various diversity indices used in this study. The numbers of clones, $OTU_{0.03}$, and the relative abundance of each of the 11 bacterial lineages examined in this studies in the five clone libraries analyzed are provided as Supplementary material (Supplementary Table S1).

Diversity estimates
We sought to estimate and compare the richness and evenness aspects, as well as the total diversity for the 11 most abundant bacterial soil lineages in five different clone libraries. Single diversity indices were used to estimate richness and evenness, whereas rarefaction curve analysis and diversity-ordering approaches were used to rank lineages within environments.

Single indices rankings. In our attempt to choose indices for either the richness or the evenness aspects of diversity, we leaned towards indices that are minimally affected or totally unaffected by sample-size variations, as the sample sizes of the lineages compared within each soil as well as across the five soils were different (Supplementary Table S1). We chose the log series index (α) as the richness estimate (Kempton and Taylor, 1974, 1976; Taylor, 1978) (Table 1), as it has been shown that its value is

Table 1 Formulae of various diversity indices used in this study

Index	Scale parameter	Formula	Reference(s)
Single diversity indices			
Complement of Simpson's dominance index	NA	$1 - D = 1 - \sum \left(\frac{n_i(n_i-1)}{N(N-1)} \right)$	Simpson (1949); Magurran (2004)
Log series index (α)	NA	$S/N = \left[\frac{(1-x)}{x} \right] \bullet [-\ln(1-x)]$ $\alpha = \frac{N(1-x)}{x}$	Kempton and Taylor (1974, 1976); Taylor (1978)
Diversity-ordering methods			
I. Information-related group of methods			
1. Renyi generalized entropy	α	$H_\alpha = \frac{\ln\left(\sum_{i=1}^{S} p_i^\alpha\right)}{(1-\alpha)}$	Tóthmérész (1995); Liu *et al.* (2007)
2. Hill's diversity number	A	$N_a = \left(\sum_{i=1}^{S} p_i^a\right)^{1/(1-a)}$	Hill (1973); Liu *et al.* (2007)
3. Daroczy's entropy	α	$H^\alpha = \frac{\left(\sum_{i=1}^{S} p_i^\alpha\right)-1}{(2^{1-\alpha}-1)}$	Dazóczy (1970); Liu *et al.* (2007)
4. Numbers equivalent	β	$S_\beta = \left(\sum_{i=1}^{S} p_i^{\beta+1}\right)^{-1/\beta}$	Patil and Taillie (1979, 1982); Liu *et al.* (2007)
5. Diversity index	β	$\Delta_\beta = \frac{\left(1-\left(\sum_{i=1}^{S} p_i^{\beta+1}\right)\right)}{\beta}$	Patil and Taillie (1979, 1982); Liu *et al.* (2007)
6. Tsallis generalized entropy	q	$H_q^T = \frac{\left(1-\sum_{i=1}^{S} p_i^q\right)}{q-1}$	Tsallis (2002); Liu *et al.* (2007)
II. Expected number of species-related group of methods			
1. Hulbert family of diversity indices	m	$s(m) = \sum_{i=1}^{S} \left(1 - (1-p_i)^m\right)$	Hulbert (1971); Liu *et al.* (2007)
2. Hurlbert–Smith–Grassle index	ω	$\Delta_\omega^{\mathrm{HSG}} = \sum_{i=1}^{S} (1-p_i)(1-(1-p_i)^\omega)$	Patil and Taillie (1982); Liu *et al.* (2007)
III. Intrinsic diversity-related group of methods			
1. Right tail-sum	i	$T_i = \sum_{j=i+1}^{S} p_j$	Patil and Taillie (1979, 1982); Liu *et al.* (2007)
2. Logarithmic dominance	Log i	$L_i = \sum_{j=i+1}^{S} p_j$	Tóthmérész (1995); Liu *et al.* (2007)
3. Majorization	i	$M_i = \sum_{j=1}^{i} p_j$	Solomon (1979); Liu *et al.* (2007)
4. *k* dominance	Log i	$k_i = 100 \times \sum_{j=1}^{i} p_j$	Shaw *et al.* (1983); Liu *et al.* (2007)

Abbreviation: NA, not applicable.
N is the total number of individuals, n_i is the number of individuals of the ith species, S is the observed number of species, p_i is the relative abundance of the ith species and p_j is the relative abundance of species in descending order.

308

relatively independent of sample size variation and is totally unaffected by sample size when the number of individuals in a community (N) is more than 1000 (Magurran, 2004). Out of the 55 lineages compared in this study (11 lineages per soil), 33 had $N > 1000$, 10 had $500 < N < 1000$ and 12 had $N < 500$. As pointed out by Hill *et al.* (2003), even with smaller sample sizes, although α increases with the increase in sample size, it is still more stable to sample-size variation compared with other richness indices, for example, Margalef (Clifford and Stephenson, 1975) and Menhinick (Whittaker, 1977) indices, and is more sensitive to changes in OTU richness. For evenness estimation, the complement of Simpson dominance index D (Simpson, 1949) $(1-D)$ was used as the evenness statistics (Magurran, 2004; Jarvis *et al.*, 2008) (Table 1). Simpson index is known to put more weight on the abundant members of the community and hence is not affected by the variation in sample size.

For each lineage in the five clone libraries, both the log series (α) and the complement of Simpson's dominance index $(1-D)$ were computed as suggested by Magurran (2004). Values for the log series index and the complement of Simpson index for all 55 lineages analyzed in this study are provided as Supplementary material (Supplementary Table S2).

Rarefaction curve rankings. Rarefaction has been viewed as a solution to the sample-size dependence of single diversity indices (Hughes *et al.*, 2001). Communities can be compared using their rarefaction curves, where the community whose rarefaction curve lies above is considered to be more diverse. Rarefaction curves can be used to compare communities with different sample sizes (Hughes *et al.*, 2001) and take into account the two components (richness and evenness) of diversity (Olszewski, 2004). We used DOTUR (Schloss and Handelsman, 2005) to construct rarefaction curves for each of the 11 lineages in all five soils. In each soil, all the rarefaction curves were plotted on the same graph and phyla were ranked from the least diverse (lowermost rarefaction curve) to the most diverse (uppermost rarefaction curve). In cases where the rarefaction curves of two communities crossed, their diversities are incomparable using this method, and they are given a tied rank.

Diversity-ordering rankings. In addition to rarefaction curve-based rankings, we used diversity-ordering methods (12 different methods) to generate diversity profiles (plots of values of diversity indices versus a scale parameter for each community or lineage) that could then be compared graphically with rank lineages in all five different data sets. It has been argued previously that defining diversities in complex communities as a single number using single diversity indices is not practical, as it summarizes the multidimensional concept of community into a single scalar quantity (Tóthmérész,

1995; Liu *et al.*, 2007). The varying degrees of emphasis that the single diversity indices put on rare or abundant species have been regarded as a disadvantage. Owing to the inadequacy of single indices, ecologists in various subfields of ecology have used diversity-ordering approaches to quantify diversity (see Liu *et al.*, 2007 for a review). Diversity-ordering methods, being a parametric family of indices, give a more complete summarization of community diversity, with varying sensitivities or degrees of emphasis on rare versus common species as the scale parameter changes (Tóthmérész, 1995 no. 11; Liu *et al.*, 2007 no. 3). This makes them less dependent on sample size, compared with many of the single diversity indices commonly used in microbial ecology. Moreover, the use of multiple methods of ordering (see below) and their comparison ensure a more accurate and complete assessment of diversity.

Diversity-ordering methods are placed into three groups depending on the properties they measure: information-related methods, expected number of species-related methods and intrinsic diversity-related methods. Table 1 shows the different methods within each group, their scale parameter and the formulae used for the calculation. The scale parameters for the information-related and the expected number of species-related methods are infinite (for a reference for the allowed values of the scale parameters, see Liu *et al.*, 2007 no. 3). This means that, on using indices that belong to the these groups of diversity-ordering methods, even if the profiles did not cross within the range of the scale parameter used for comparison, there is no guarantee that they would not cross at higher values of the scale parameter. However, within the third group of methods (intrinsic diversity-related methods), only a finite number of values need to be calculated and compared (equivalent to the observed number of species (S) or $S-1$). This means that if the profiles did not cross, then the communities are definitely comparable. This makes the intrinsic diversity-related indices the most stringent. Again, as with rarefaction curves, if the diversity profiles of two communities or lineages crossed, they are considered incomparable using this method and given a tied rank. Figure 1 shows an example of comparative diversity of two lineages, the *Planctomycetes* and *Gemmatimonadetes* in Canada soil, using three diversity-ordering indices (one from each group): Renyi generalized entropy index, Hulbert diversity index and right tail-sum method. Although the profiles of the *Planctomycetes* and *Gemmatimonadetes* crossed with the right tail-sum method and hence were incomparable (Figure 1c), they were comparable using the other two methods (Figures 1a and b), and *Planctomycetes* was found to be more diverse than *Gemmatimonadetes* in both cases.

For each of the 12 different diversity-ordering indices, a stepwise approach was taken, where
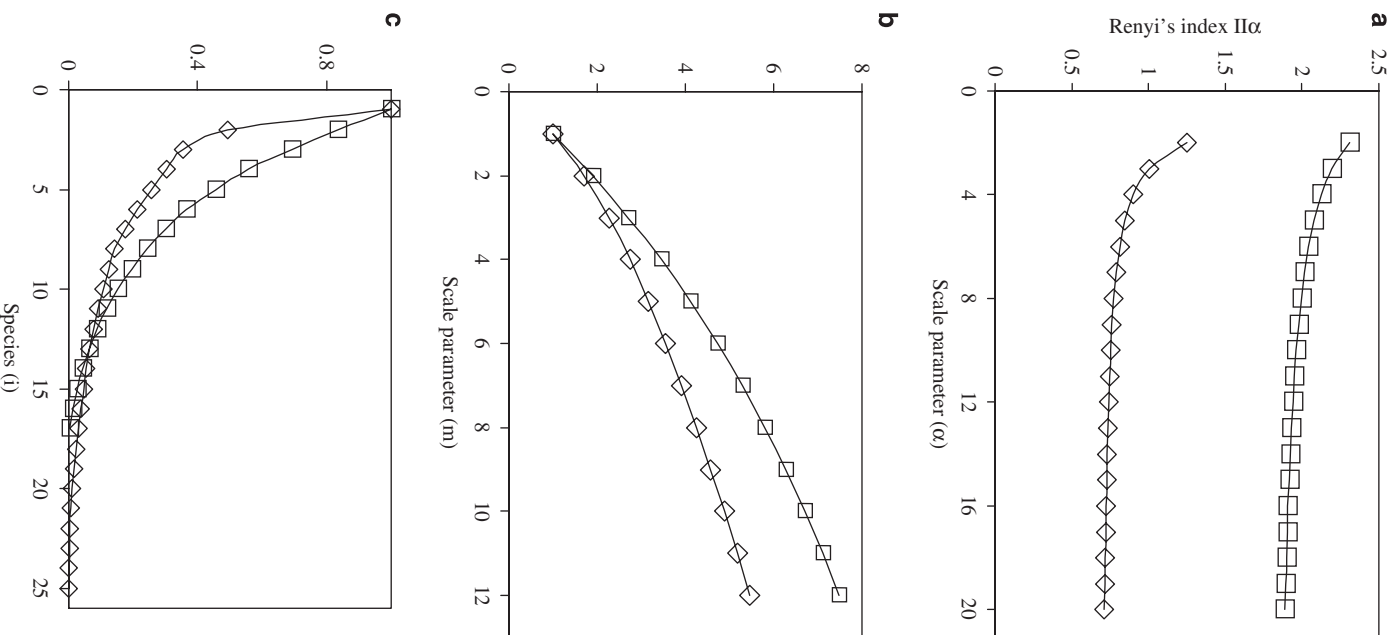
Figure 1 Example of diversity-ordering profiles: a pairwise comparison of two lineages (*Gemmatimonadetes*, open diamonds; and *Planctomycetes*, open squares) within one (Canada clone library) of the five different data sets tested using (**a**) Renyi generalized entropy method as an example of information-related group of methods. (**b**) Hulbert family of diversity indices as an example of expected number of species-related group of methods. (**c**) Right tail-sum method as an example of intrinsic diversity-related group of methods. Similar pairwise comparisons were conducted between the 11 tested phyla in each of the five data sets used in this study to reach a consensus ranking within each soil data set.

Table 2 Diversity rankings of 11 soil bacterial lineages using rarefaction curve analysis and diversity-ordering approaches[a]

Lineage	KFS clone library		Canada clone library		Florida clone library		Illinois clone library		Brazil clone library	
	Rarefaction curve analysis[b]	Diversity-ordering approach[c]	Rarefaction curve analysis[b]	Diversity-ordering approach[c]	Rarefaction curve analysis[b]	Diversity-ordering approach[c]	Rarefaction curve analysis[b]	Diversity-ordering approach[c]	Rarefaction curve analysis[b]	Diversity-ordering approach[c]
Acidobacteria	7	9	4.5	8	9	9	3	2.5	5.5	5
Actinobacteria	10	10	7	11	8	5	8	8.5	4	3
α-*Proteobacteria*	7	6	10	6	4	7	2	1	7	7
Bacteroidetes	3.5	4	8	10.5	2	3	5	6	5.5	6
β-*Proteobacteria*	1.5	3	4.5	7	1	1	4	5	2	2
Chloroflexi	9	7.5	3	2	7	6	7	7	8	8
δ-*Proteobacteria*	5	5	10	10.5	10	10	11	11	10.5	11
Firmicutes	7	7.5	10	5	11	11	10	10	9	10
Gemmatimonadetes	3.5	2	1	1	5	8	6	2.5	1	1
Planctomycetes	11	11	6	3	6	4	9	8.5	10.5	9
Verrucomicrobia	1.5	1	2	4	3	2	1	4	3	4

[a]Ranking are from the least diverse (1) to most diverse (11).
[b]Rankings using rarefaction curve analysis.
[c]Final rankings using 12 different diversity-ordering approaches.

pairwise comparisons between various lineages in the data set were conducted until ranking of all 11 lineages is achieved. A consensus ranking for each group of indices (that is, an information-related indices ranking, an expected number of species indices ranking and an intrinsic diversity indices ranking) was obtained for each clone library. The final rankings (Table 2) represent averages of the three groups rankings. Whenever lineages where incomparable, they were given a tied rank.

To our knowledge, diversity ordering has not been previously used in comparing microbial communities. We believe that this approach might prove extremely valuable in studies where the two compared communities do not share any OTUs, such as ranking various phyla within a data set or comparing archaeal, bacterial and eukaryotic diversity within an ecosystem. It should also prove useful in comparing diversities of ecosystems with little shared OTUs on account of wide variations in environmental conditions.

Diversity–abundance correlations

To investigate the relationship between relative abundances and each of the richness, evenness and total diversity for the 11 lineages compared in this study, we calculated Pearson correlation coefficients between percentage abundance and values of the richness index (α), as well as the complement of Simpson's dominance index ($1-D$) for each of the 11 lineages across the five data sets analyzed. On the other hand, as total diversity is represented as a rank (rarefaction-based and diversity ordering-based), Spearman rank correlation coefficient was used to correlate between relative abundance ranks and total diversity ranks (the average of rarefaction curve and

diversity-ordering rank) for each of the 11 lineages across the five data sets analyzed.

Results

Diversity ranking patterns in soil data sets

The 11 bacterial lineages examined in this study were ranked according to rarefaction curve analysis, as well as diversity-ordering approaches, using information-based methods (six indices), expected number of species-based methods (two indices) and intrinsic diversity-based methods (four indices). Phyla were ranked from one (least diverse) to 11 (most diverse).

The results (Table 2) of the ranking using rarefaction and diversity ordering were broadly in agreement, and indicated that some phylogenetic lineages were collectively classified as highly diverse in clone libraries examined. These lineages are the *Firmicutes* (average rank of 9.4 and 8.7 using rarefaction curve and diversity-ordering approaches, respectively), δ-*Proteobacteria* (average rank 9.3 and 9.5) and *Planctomycetes* (average rank 7.1 and 8.5). Similarly, a second group was collectively classified as the least diverse: The *Gemmatimonadetes* (average rank 3.3 and 2.9), the *Verrucomicrobia* (average rank 2.1 and 3.0) and the β-*Proteobacteria* (average rank 2.6 and 3.6). The five remaining phylogenetic lineages (*Acidobacteria*, *Actinobacteria*, α-*Proteobacteria*, *Chloroflexi* and *Bacteroidetes*) showed no consistent rankings among various data sets, as evident by wide range of rarefaction curve and diversity-ordering rankings in different clone libraries (rarefaction curve rankings ranged from 2.5 to 9, 3 to 11, 1 to 7, 3 to 10.5 and 2 to 8 and diversity ordering ranged from 3 to 9, 4 to 10, 2 to 10, 2 to 8 and 3 to 9 for the *Acidobacteria*, *Actinobacteria*,
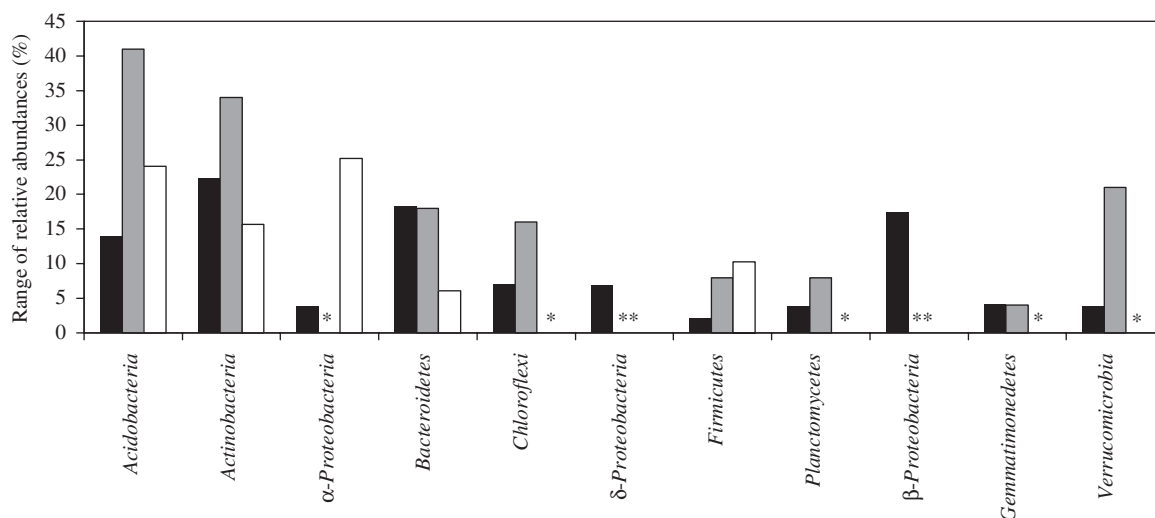

Figure 2 Ranges of relative abundances observed for the 11 lineages in the five data sets examined in this study (black columns), in a compilation of thirty-two 16S rRNA gene clone libraries analyzed by Janssen (2006) (gray columns), and in a quantitative PCR analysis of the relative percentage of six bacterial lineages (*Acidobacteria*, *Actinobacteria*, α-*Proteobacteria*, β-*Proteobacteria*, *Bacteroidetes* and *Firmicutes*) in 69 different soil samples (Fierer *et al.*, 2007) (white columns). *Not reported.

α-*Proteobacteria*, *Bacteroidetes* and *Chloroflexi*, respectively) (Table 2).

Interestingly, lineages that exhibited large variations in rankings between soils also exhibited large difference in relative abundances between different soils in our data set (Figure 2, black columns). The only two exceptions were α-*Proteobacteria*, where relative abundance in various soils exhibited remarkable stability in spite of the changes in diversity rank, and β-*Proteobacteria*, where significant changes in abundance were not accompanied by similar changes in the diversity rank. Furthermore, an examination of relative abundances within a recently compiled data set of 32 16S rRNA gene clone libraries (Janssen, 2006, Figure 2, gray columns), as well as a quantitative PCR analysis of the relative abundance of six bacterial lineages (*Acidobacteria*, *Actinobacteria*, α-*Proteobacteria*, β-*Proteobacteria*, *Bacteroidetes* and *Firmicutes* (Fierer *et al.*, 2007, Figure 2, white columns)) in 69 different soil samples, confirmed the same general trend: Higher differences in relative abundances were observed in lineages with wide variations in diversity rankings, compared with smaller variations in relative abundances in lineages with consistently low- or high-diversity ranks.

We further examined the relationship between relative abundance and richness, evenness and total diversity rank in all lineages, with special emphasis on those that exhibited significant changes in both diversity rankings and abundance in the five libraries tested (*Acidobacteria*, *Actinobacteria*, *Chloroflexi* and *Bacteroidetes*). The correlation between abundance and various richness, evenness and diversity estimates is especially important within these four lineages, as the magnitude of change in community structure (abundance and diversity) is highest in these four phyla. Therefore,

the nature of this correlation would have a greater effect on the overall community structure, compared with other lineages that show little changes in diversity and or abundance in data sets examined. Results (Table 3) indicate that richness, evenness and diversity, in general, exhibited similar patterns of correlation with abundance. Within these four lineages, a strong abundance–diversity correlation was observed within the *Actinobacteria* ($r = 0.7$), a weaker positive correlation was observed within the *Chloroflexi* ($r = 0.58$) and the *Acidobacteria* ($r = 0.41$) and a strong negative correlation was observed within the *Bacteroidetes* ($r = -0.99$).

Discussion

The most important finding of this study is the identification of lineages that are consistently ranked as having high or low diversity in the five data sets analyzed. This observation could provide useful insights regarding our estimates of the contribution of various bacterial lineages to ecosystem functions. According to ecological and evolutionary theory, more diverse communities have higher contribution to ecosystem functions and services than less diverse counterparts (Giovannoni, 2004; Bell *et al.*, 2005). It has been postulated previously that as each species uses slightly different resources and occupy a highly specified niche in the community, more diverse communities provide more services to the ecosystem than counterparts with less diversity (Loreau and Hector, 2001; Bell *et al.*, 2005). From an evolutionary perspective, bacterial communities in soil are highly adapted and constrained by selection, with newly evolving, more ecologically fit cells rapidly evolving and eliminating lesser competitors (Giovannoni, 2004). As such, the communities observed in 16S rRNA gene studies are the victors of an evolutionary process in which less competitive species are continuously being purged from the community. It follows that lineages with higher diversity have higher contribution to ecosystem functioning than lineages with comparable abundance but lower diversity.

Whereas some lineages were consistently ranked as having high or low diversity in all data sets, others exhibited swings in their rankings in different environments (Table 2). Interestingly, these lineages were also those that, in general, exhibited wide ranges of relative abundance in the five clone libraries examined in this study (Figure 2). Whereas diversity rankings are not available and unreliable to compute in smaller data sets available at present, examination of previously available data sets have indeed shown that, in general, members of these lineages exhibited the largest variations in relative abundances in soils previously examined using 16S rRNA gene analysis (Janssen, 2006) and quantitative PCR (Fierer *et al.*, 2007) (Figure 2). This argues that members of these lineages constitute a more dynamic core of micro-

Table 3 Correlation coefficients between relative abundance and richness, evenness or total diversity within soil bacterial lineages[a]

Lineage	Richness[b]	Evenness[c]	Total diversity[d]
Acidobacteria	**0.84**	**0.39**	**0.41**
Actinobacteria	**0.94**	**0.77**	**0.7**
α-*Proteobacteria*	−0.15	−0.17	−0.38
Bacteroidetes	**−0.95**	**−0.8**	**−0.99**
β-*Proteobacteria*	−0.92	−0.61	−0.46
Chloroflexi	**0.98**	**0.76**	**0.58**
δ-*Proteobacteria*	0.01	0.38	0.04
Firmicutes	−0.85	−0.8	−0.78
Gemmatimonadetes	−0.01	−0.35	0.05
Planctomycetes	0.95	0.55	0.24
Verrucomicrobia	0.48	0.49	0.44

[a]Lineages exhibiting wide variations in both diversity and correlations are shown in bold.
[b]Pearson correlation coefficient between log series index (α) and relative abundance.
[c]Pearson correlation coefficient between the complement of Simpson dominance index ($1-D$) and relative abundance.
[d]Spearman rank correlation coefficient between total diversity rank (average of rarefaction and diversity-ordering ranks) and relative abundance rank.

organisms that are more readily affected by and responsive to variations in soil characteristics.

We further examined the correlation between relative abundance and diversity (richness, evenness and total diversity), with special emphasis on lineages exhibiting variations in diversity rankings and relative abundances in our data sets (Table 3). Theoretically, a positive correlation between diversity and abundance within a specific lineage suggests that the observed increase in relative abundance is driven by the recruitment of new bacterial species belonging to this lineage to the ecosystem through immigration, or the elevation of rare, previously undetected members of this specific lineages to higher numbers. Under this scenario, richness, evenness and diversity ranks will be positively correlated with the increase in relative abundance. On the other hand, a negative correlation between abundance and diversity suggests that the increase in relative abundance is driven by the propagation of a single, or relatively few, species to higher numbers, which results in the out-competition of other members of the community or a drop in their numbers to undetectable levels. Under this scenario, the observed richness, evenness and diversity rank will be negatively correlated with the increase in relative abundance. The results obtained in this study suggest that, within the four lineages that exhibited wide variations in both relative abundance and diversity rankings, members of the *Acidobacteria*, *Actinobacteria* and *Chloroflexi* follow the first scenario, whereas members of the *Bacteroidetes* follow the second scenario (Table 2). It is important to note that these correlations are computed using data sets that represent snapshots of five different soil bacterial communities. As such, these reported patterns do not represent dynamics of the recruitment, promotion and demotion of bacterial species within a single ecosystem, but are rather general trends of diversity–abundance correlations in different soil ecosystems.

This study represents a preliminary effort to understand the characteristic diversity ranks and diversity–abundance correlations observed in various bacterial lineages. Similar studies could greatly benefit from the anticipated availability of large (for example, > 10 000 clones) capillary and pyrosequencing soil clone libraries in the future (Tiedje, 2008). Data sets from multiple, diverse soil ecosystems could help in determining whether the pattern obtained in this study represents a global pattern of diversity ranks in various soils. In addition, multiple data sets documenting the response of a single ecosystem to natural and/or anthropogenic manipulations should prove extremely valuable in tracing the diversity–abundance correlations observed at different time points to phylogenetic information regarding specific OTUs in various bacterial lineages. Finally, larger data sets could allow for conducting similar studies on finer taxonomic scales, for example, at the class and order levels).

Members of some of the examined lineages in this study (for example, the α-, β- and δ-*Proteobacteria* and the *Firmicutes*) exhibit a bewildering array of metabolic and physiological capabilities. Therefore, it is conceivable that on account of this extremely diverse metabolic capabilities and multiple ecological roles within some of the 11 specific division or subdivision examined, the patterns of diversity and diversity–abundance correlations could significantly vary when examined at a lower (for example, class or order) taxonomic level.

Acknowledgements

References

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2006). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**: 5734–5741.

Axelrood PE, Chow ML, Radomski CC, McDermott JM, Davies J. (2002). Molecular characterization of bacterial diversity from British Columbia forest soils subjected to disturbance. *Can J Microbioil* **48**: 655–674.

Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. (2005). The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

Borneman J, Triplett EW. (1997). Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* **63**: 2647–2653.

Clifford HT, Stephenson W. (1975). *An Introduction to Numerical Classification*. Academic Press: London, UK.

Costello EK, Schmidt SK. (2006). Microbial diversity in alpine tundra wet meadow soil: novel *Chloroflexi* from a cold, water-saturated environment. *Environ Microbiol* **8**: 1471–1486.

Dazóczy Z. (1970). Generalized information functions. *Inform Contr* **16**: 36–51.

DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM *et al.* (2006a). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acid Res* **34**: D394–D399.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006b). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Dunbar J, Barns SM, Ticknor LO, Kuske CR. (2002). Empirical and theoretical bacterial diversity in four Arizona soils. *Appl Environ Microbiol* **6**: 3035–3045.

Elshahed MS, Youssef NH, Spain AM, Sheik C, Najar FZ, Sukharnikov LO *et al.* (2008). Novelty, uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microbiol* **74**: 5422–5428.

Fierer N, Bradford MA, Jackson RB. (2007). Toward an ecological classification of soil bacteria. *Ecology* **88**: 1354–1364.

Giovannoni S. (2004). Oceans of bacteria. *Nature* **430**: 515–516.

Hill MO. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**: 427–432.

Hill TCJ, Walsh KA, Harris JA, Moffett BF. (2003). Using ecological diversity measure with bacterial communities. *FEMS Microbiol Ecol* **43**: 1–11.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. (2001). Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**: 4399–4406.

Hulbert SH. (1971). The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577–586.

Janssen PH. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* **72**: 1719–1728.

Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S *et al.* (2008). A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proc Natl Acad Sci USA* **105**: 5326–5331.

Kempton RA, Taylor LR. (1974). Log-series and log-normal parameters as diversity determinants of the Lepidoptera. *J Anim Ecol* **43**: 381–399.

Kempton RA, Taylor LR. (1976). Models and statistics for species diversity. *Nature* **262**: 818–820.

Kuske CR, Barns SM, Busch JD. (1997). Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. *Appl Environ Microbiol* **63**: 3614–3621.

Lesaulnier C, Papamichail D, McCorkle S, Ollivier B, Skiena S, Taghavi S *et al.* (2008). Elevated atmospheric CO_2 affects soil microbial diversity associated with trembling aspen. *Environ Microbiol* **10**: 926–941.

Lipson DA, Schmidt SK. (2004). Seasonal changes in an alpine soil bacterial community in the Colorado Rocky Mountains. *Appl Environ Microbiol* **70**: 2867–2879.

Liu C, Whittaker RJ, Ma K, Malcolm JR. (2007). Unifying and distinguishing diversity ordering methods for comparing communities. *Popul Ecol* **49**: 89–100.

Loreau M, Hector A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**: 72–76.

Magurran AE. (2004). *Measuring Biological Diversity*. Blackwell Science Ltd: Oxford, UK.

Nogales BE, Moore RB, Liobet-Brossa E, Roesello-Mora R, Amann R, Timmis K. (2001). Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol* **67**: 1874–1884.

Olszewski TD. (2004). A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *OIKOS* **104**: 377–387.

Patil GP, Taillie C. (1979). An overview of diversity. In: Grassle JF, Patil GP, Smith W, Taillie C (eds). *Ecological Diversity in Theory and Practice*. International Cooperative publishing House: Fairland, pp 3–27.

Patil GP, Taillie C. (1982). Diversity as a concept and its measurement. *J Amer Stat Assoc* **77**: 548–567.

Roesch LFW, Fulthorps RR, Riva A, Casella G, Hadwin AK, Kent AD *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283–290.

Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.

Schloss PD, Handelsman J. (2006). Toward a census of bacteria in soil. *PLoS Comput Biol* **2**: 786–793.

Shaw KM, Lamshead PJD, Platt HM. (1983). Detection of pollution-induced disturbance in marine benthic assemblages with special reference to nematodes. *Mar Ecol Prog Ser* **11**: 195–202.

Simpson EH. (1949). Measurement of diversity. *Nature* **163**: 688.

Solomon DL. (1979). A comparative approach to species diversity. In: Grassle JF, Patil GP, Smith W, Taillie (eds). *Ecological Diversity in Theory and Practice*. International Cooperative publishing House: Fairland, pp 29–35.

Talera S, Jangid K, Ivester AH, Whitman WB, Williams MA. (2008). Microbial community succession and diversity in soils during 77 000 years of ecosystem development. *FEMS Microbiol Ecol* **64**: 129–140.

Taylor LR. (1978). Bates, Williams, Hutchinson-a variety of diversities. In: Mound LA, Warloff N (eds). *Diversity of Insect Faunas*. Blackwell publishing group: Oxford, UK, pp 1–18.

Tiedje JM. (2008). *Microbial Inventory on a Planetary Scale* Presented in 108th general meeting of the American Society for Microbiology, Boston, MA, USA.

Tóthmérész B. (1995). Comparison of different methods for diversity ordering. *J Veg Sci* **6**: 283–290.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554–557.

Tsallis C. (2002). Entropic nonextensivity: a possible measure of complexity. *Chaos Soliton Fract* **13**: 371–391.

Whittaker RH. (1977). Evolution of species diversity in land communities. *Evolutionary Biol* **10**: 1–67.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)