

Bacterial taxa abundance pattern in an industrial wastewater treatment system determined by the full rRNA cycle approach

Eva L. M. Figuerola and Leonardo Erijman*

Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Vuelta de Obligado 2490 (ADN1428) Buenos Aires, Argentina.

Summary

The description of the diversity and structure of microbial communities through quantification of the constituent populations is one of the major objectives in environmental microbiology. The implications of models for community assembly are practical as well as theoretical, because the extent of biodiversity is thought to influence the function of ecosystems. Current attempts to predict species diversity in different environments derive the numbers of individuals for each operational taxonomic unit (OTU) from the frequency of clones in 16S rDNA gene libraries, which are subjected to a number of inherent biases and artefacts. We show that diversity of the bacterial community present in a complex microbial ensemble can be estimated by fitting the data of the full-cycle rRNA approach to a model of species abundance distribution. Sequences from a 16S rDNA gene library from activated sludge were reliably assigned to OTUs at a genetic distance of 0.04. A group of 17 newly designed rRNA-targeted oligonucleotide probes were used to quantify by fluorescence *in situ* hybridization, OTUs represented with more than three clones in the 16S rDNA clone library. Cell abundance distribution was best described by a geometric series, after the goodness of fit was evaluated by the Kolmogorov-Smirnov test. Although a complete mechanistic understanding of all the ecological processes involved is still not feasible, describing the distribution pattern of a complex bacterial assemblage model can shed light on the way bacterial communities operate.

Received 12 July, 2006; accepted 24 February, 2007. *For correspondence. E-mail erijman@dna.uba.ar; Tel. (+54) 11 4783 2871; Fax (+54) 11 4786 8578.

Introduction

Biological municipal and industrial wastewater treatment plants (WWTPs) are one of the most important biotechnological processes in operation worldwide. The success of the treatments relies on the multiple microbial activities and interactions occurring in the context of highly diverse microbial communities. It is widely recognized the merit of describing ecological communities by means of diversity indexes and species abundance patterns, due to the likely influence of biodiversity on the function of ecosystems (Fernandez *et al.*, 1999; Loreau *et al.*, 2001; Briones and Raskin, 2003; Curtis and Sloan, 2004; Horner-Devine *et al.*, 2004; Bell *et al.*, 2005; Magurran, 2005a; O'Mullan and Ward, 2005). Previous estimation of species richness in WWTPs have been based on parametric (Curtis *et al.*, 2002) or non-parametric methods, which derive the numbers of individuals for each operational taxonomic unit (OTU) from the frequency of the clones in 16S rDNA gene libraries. Despite recent methodological progress in the quantitative robustness of multitemplate PCR cloning (Kurata *et al.*, 2004; Acinas *et al.*, 2005), the estimation of abundance and diversity of genes based on PCR are subject to a number of inherent biases and artefacts (Suzuki and Giovannoni, 1996; Polz and Cavanaugh, 1998; Qiu *et al.*, 2001; Thompson *et al.*, 2002; Hugenholtz and Huber, 2003; Kurata *et al.*, 2004; Acinas *et al.*, 2005). In addition, as a clone library is typically a sample of a relatively limited number of individuals taken from a very diverse community, it may fall short of representing the actual abundance distribution even in the absence of PCR bias. In a recent survey of the microbial community in an industrial wastewater bioreactor, in which two sets of primers specific for the eubacteria were used to PCR amplify 16S rRNA gene sequences, Bramucci and colleagues (2003) reported that only the 29 most abundant OTU sequences out of a total 106 OTUs were common to both clone libraries, albeit with very different relative abundances. Analogous results were obtained by Wagner and co-workers (Juretschko *et al.*, 2002), who analysed a total of 96 clones from three clone libraries generated with DNA from activated sludge extracted using three different methods. There are other similarly complex systems that also reflect a lack of agreement between quantitative

probing and the abundance of clones in clone libraries (e.g. Luyten *et al.*, 2006). The implication of these findings is that any estimate of bacterial diversity based on a single 16S rRNA data set has to be interpreted with care.

The full-cycle rRNA approach was introduced about a decade ago as a cultivation-independent solution to the identification and enumeration of microorganisms in complex environments, by direct 16S rDNA sequence retrieval followed by the design of probes specific for these sequences (Amann *et al.*, 1995). Fluorescence microscopy of bacteria hybridized with fluorescent probes targeting 16S rRNA has become the most widely used method for the analysis of multispecies microbial samples. In this context, the use of fluorescence *in situ* hybridization (FISH) for statistically valid quantification is very useful for detailed studies of diversity in activated sludge (Davenport *et al.*, 2000; Coskuner *et al.*, 2005). Full-cycle rRNA analysis of the bacterial community composition from full-scale activated sludge has been performed in only a few WWTPs (Snaird *et al.*, 1997; Juretschko *et al.*, 2002; Eschenhagen *et al.*, 2003). However, these previous studies did not attempt to analyse the abundance pattern in term of a rigorously validated taxonomic hierarchy.

In this work we applied the full-cycle rRNA approach to determine the species abundance of the bacterial community in a full-scale wastewater treatment system receiving pretreated wastewater from a petroleum refinery. After clone sequences from a 16S rDNA library were assigned to OTUs using a consistent genetic distance level, quantitative FISH data of OTUs abundance obtained using newly designed oligonucleotide probes were tested against distribution models. We conclude that the geometric series was an adequate descriptor of the distribution of the most abundant OTUs in this activated sludge.

Results

16S rRNA gene diversity analysis

The DOTUR program (Schloss and Handelsman, 2005) was used to assign the 0.6 kb fragment sequences accurately to OTUs based on a genetic distance of 0.04 between sequences. A total of 39 OTUs were defined, with the top five OTUs containing 62% of the sequences. There were 25 unique 16S rDNA sequences in the clone library, out of the 135 clones sequenced, from which a coverage of 82% was computed using the method of Good (1953). The rarefaction curve did not reach saturation (Fig. 1), suggesting that further cloning would have revealed more OTUs. However, the curve was far from linear, indicating that the most prevalent bacterial groups in the clone library were likely identified. A phylogenetic tree inferred from the 16S rDNA sequences showed that

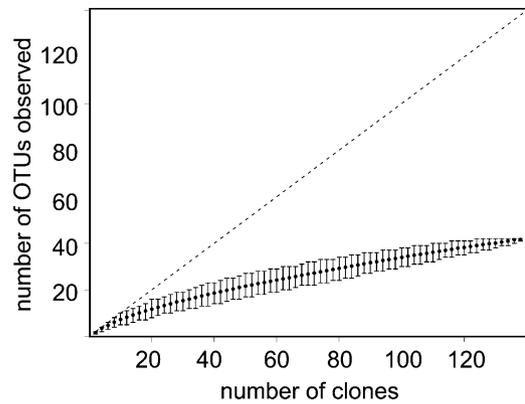


Fig. 1. Rarefaction curve for the 16S rDNA clone library. The expected number of OTUs is plotted versus the number of clones. The error bars represent 95% confidence intervals. The dashed line equals a slope of 1, which is the case if every new examined clone corresponded to a new OTU.

they are distributed over a wide range of bacterial taxonomic groups (Fig. 2). The most represented phylogenetic groups, assigned with the taxonomical hierarchy classifier of the Ribosomal Database Project II (Cole *et al.*, 2005) were the *Betaproteobacteria* with 38.5% of the sequences, followed by *Alphaproteobacteria* with 23.0%. An OTU for which no clear phylogenetic affiliation could be inferred was represented by 9.6% of the clones.

Although nitrification took place with high efficiency, no ammonia-oxidizing bacteria (AOB) were represented in the 16S rRNA gene library, and there was only a single clone that according to a BLASTN search displayed the highest score the sequence of *Nitrobacter* sp. PJN1 (AY055795). However, it contained a single mismatch in the sequence complementary to the *Nitrobacter*-specific probe Nit3, matching the probe that is used as a competitor for Nit3, i.e. a sequence that is characteristic of several strains of *Bradyrhizobium* and *Rhodopseudomonas*. Using the operating conditions given in the *Experimental procedures* and the model developed by Rittman and colleagues (1999) the theoretical ratio of active ammonium oxidizers to active heterotrophs should have constituted approximately 9%. However, FISH analysis using probes Nso1225, Nit3 and S⁻-Ntspa-665-a-A-18 could not detect above background-signals. Ammonia-oxidizing bacteria were only detected by nested PCR-denaturing gradient gel electrophoresis analysis using the AOB-specific CTO primer pair (CTO189f, CTO654r, Kowalchuk *et al.*, 1997) followed by a second round of amplification with eubacterial-directed primers 341f-GC and 534r. The most intense bands from this gel were extracted and sequenced. Though too short to make an acceptable phylogenetic inference (195 bp), they appeared to be related to *Nitrosomonas* (99% similarity to *Nitrosomonas europaea* ATCC 19718, BX321856).

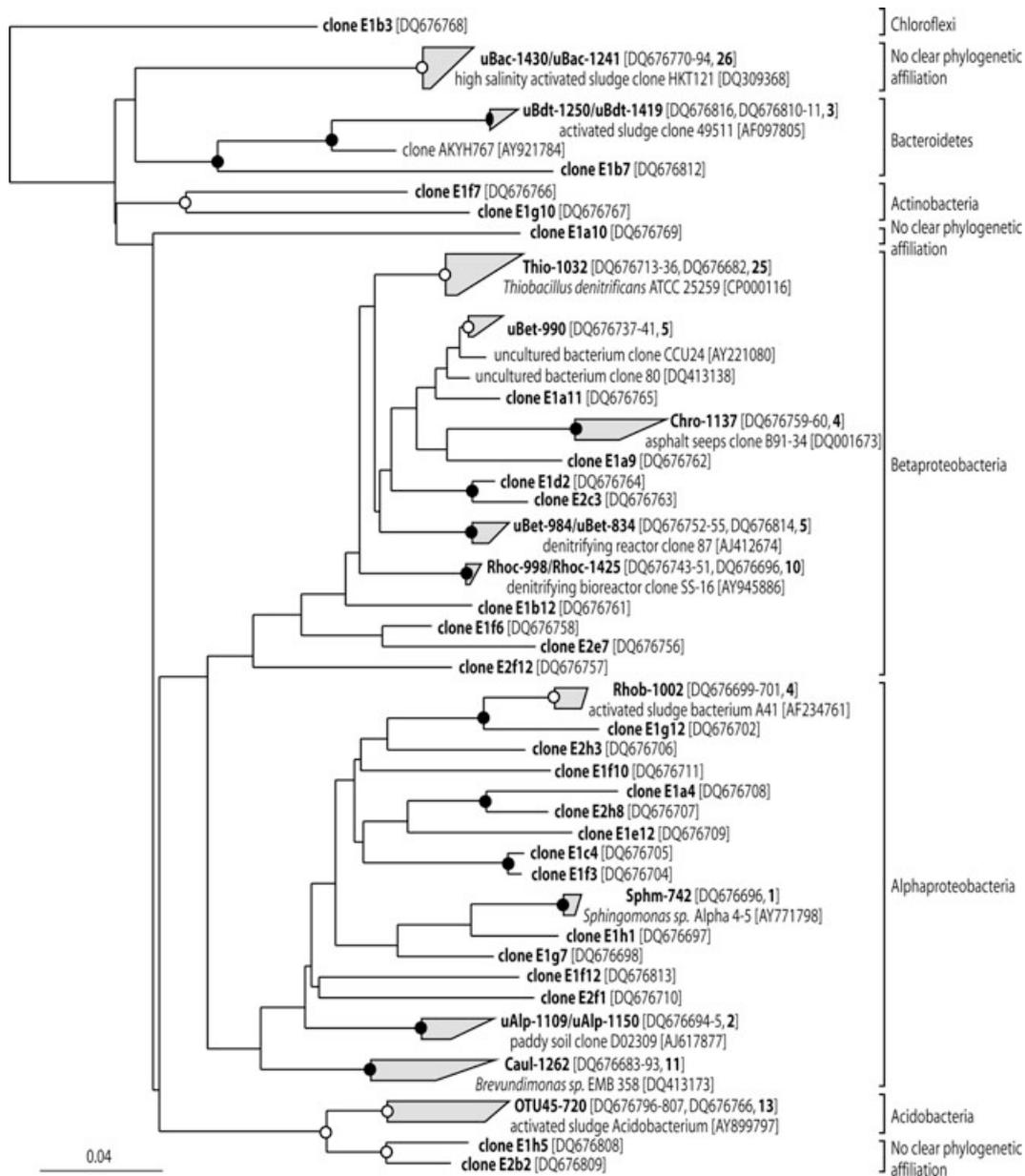


Fig. 2. Dendrogram showing the phylogenetic position of the clones obtained after amplification of the 16S rRNA gene from genomic DNA of the activated sludge. Sequences obtained in this study are indicated in bold. The 16S rDNA sequence of *Thermotoga thermarum* gene (AB039769) was used as outgroup. Filled circles correspond to branches conserved for all methods tested (maximum parsimony and neighbour joining with more than 70% bootstrap, and maximum likelihood). Open circles correspond to branches conserved for at least two of the methods tested. Names in bold indicate the oligonucleotide probes targeting the respective clades. Accession numbers are indicated between brackets. When appropriate, the number after the comma indicates the number of clones in the corresponding clade. The scale bar represents a 0.04 substitution per nucleotide position.

Development of OTU-specific probes and in situ analysis of activated sludge

A series of oligonucleotide probes were designed to target OTUs represented with at least three clones in activated sludge library samples (Table 1). Inspection of matching sequences revealed that the probes targeted signature nucleotide positions in the 16S rRNA of clusters belonging

to distinct phylogenetic affiliation (Fig. 2). After checking for specificity, all probes were used alone or in simultaneous hybridization experiments to quantify the most abundant bacterial groups by FISH.

A pie chart was used to illustrate the bacterial community composition of the activated sludge (Fig. 3). Not unexpectedly, the percentages of cells belonging to the most predominant OTUs do not match their correspond-

Table 1. Oligonucleotide probes used in this study.

Name	Target organisms in the RDP	Sequence	Target site	FA
S ⁻ -Rhob-1002-a-A-18	39 <i>Rhodobacteraceae</i>	ACCATCTCTGGAACCGCG	1002–1019	25
S ⁻ -Ualp-1150-a-A-18	10 unclassified <i>Alphaproteobacteria</i>	GCCACCGGCAGTTCCTACT	1150–1167	15
S ⁻ -Ualp-1109-a-A-18	9 unclassified <i>Alphaproteobacteria</i>	CAACTAGTAGCGAGGGTT	1109–1126	20
S ⁻ -Sphm-0742-a-A-18	27 <i>Sphingomonadales</i>	CTCAGCGTCAACAACGGT	742–759	25
S ⁻ -Bete-1250-a-A-18	2 unclassified <i>Bacteroidetes</i>	TGCCATGTGGCTGCCCAT	1250–1267	20
S ⁻ -Bete-1419-a-A-18	1 unclassified <i>Bacteroidetes</i>	GACTTTAGGTACTCCCGA	1419–1436	15
S ⁻ -Chro-1137-a-A-18	29 <i>Chromatiales</i> + 4 unclassified <i>Gammaproteobacteria</i>	TCTCTAGAGTGTCTAGCC	1137–1154	25
S ⁻ -Caul-1262-a-A-18	10 <i>Caulobacterales</i> + 1 unclassified <i>Alphaproteobacteria</i>	GGCTCACCATCGCTGGGT	1262–1279	35
S ⁻ -Ubac-1430-a-A-18	13 unclassified <i>Bacteria</i>	GCAACCCATTGTATCCGC	1430–1447	0
S ⁻ -Ubac-1241-a-A-18	13 unclassified <i>Bacteria</i>	GTTAGCTCAACCACTTCG	1241–1258	0
S ⁻ -Ubet-0984-a-A-18	16 <i>Betaproteobacteria</i>	CTGGCATGTCAAGCGTAG	984–1001	10
S ⁻ -Ubet-0834-a-A-18	23 <i>Betaproteobacteria</i>	CAAGGAAATGAATCCCG	834–851	15
S ⁻ -Ubet-0990-a-A-21	8 unclassified <i>Betaproteobacteria</i>	GCAGCCTTCCAGACATGTCAA	990–1010	10
S ⁻ -Rhoc-0998-a-A-18	3 unclassified <i>Rhodocyclaceae</i>	TCTCTCAGGATTCCGGA	998–1015	0
S ⁻ -Rhoc-1425-a-A-18	14 unclassified <i>Rhodocyclaceae</i>	ACTACCTACTTCTGGTGA	1425–1442	40
S ⁻ -Thio-1032-a-A-18	27 <i>Thiobacillus</i> + 50 unclassified <i>Betaproteobacteria</i>	CCTGTGTTCCGATTCCCT	1032–1049	60
S ⁻ -OTU45–0720-a-A-20 ^a	103 unclassified <i>Acidobacteria</i>	CAACCCGTCTTACCTCAGG	720–738	35

a. From Juretschko and colleagues (2002).
RDP, Ribosomal Database Project.

ing percentage of clones in the rDNA library. Twelve OTUs accounted for about 87% of FISH-detected cells (Fig. 3A). According to quantitative FISH, the sum of the cells that hybridized with probes targeting cells corresponding to OTUs classified as *Betaproteobacteria* amounted to 38.5% of total cells, whereas 23.4% of total cells hybridized with specific probes in the *Alphaproteobacteria* class. However, the coincidence between these numbers and the percentages of the corresponding clones in the 16S rDNA library is only apparent, as seen from the different apportionment within each class of the *Proteobacteria* (Fig. 3B). There was a good agreement with the percentage of clones belonging to the unclassified group of bacteria comprising the higher number of clones in the 16S rDNA library (26/135) and the percentage of cells hybridizing with the probes S⁻-Ubac-1430-a-A-18 and S⁻-Ubac-1241-a-A-18. In contrast, the phylum *Acidobacteria* in the clone library was represented only by 3.7% of the clones, but cells positive to the probe S⁻-OTU45-0720-a-A-20 amounted to 9.5% of FISH-detected cells (Fig. 3). Three clones affiliated with *Bacteroidetes* accounted for 5.2% of total cells (Fig. 3B).

Fitting to species abundance model

A rapid inspection of the rank/abundance plot showed that the data could be approximated by a linear relationship (Fig. 4, $r^2 = 0.9727$), prompting us to test the geometric model against the observed pattern. Figure 5 shows the plot of cumulative cells abundance determined by FISH and the corresponding expected distribution for a geometric series, with $k = 0.15$. The goodness of fit of the observed distribution was evaluated using the Kolmogorov-Smirnov test statistics. For the sample of 12

taxa probed, g_0 , $g_{0.5}$ and g_1 values were below the 5% critical values (Sokal and Rohlf, 1995), from which we concluded that the difference between the observed and expected distribution was not significant, and therefore the geometric series fit the observed pattern.

Discussion

It is currently recognized that the number of present species, as well as rules governing the community assembly, may have a definite influence on the performance and stability of engineered environments (Fernandez *et al.*, 1999; Briones and Raskin, 2003; Curtis and Sloan, 2004; Horner-Devine *et al.*, 2004; Bell *et al.*, 2005; Magurran, 2005b). The use of activated sludge is well established as a means for degrading wastewater generated by oil refineries. It involves the coordinated activity of a complex community of microorganisms that are highly specialized in the degradation of hydrocarbons and phenolic compounds.

For most systems the actual richness of microbial species cannot be measured directly. Estimates are mostly inferred on the statistical analysis of patterns of species abundances in 16S rRNA gene libraries (Curtis *et al.*, 2002; Lunn *et al.*, 2004; Hong *et al.*, 2006), which are subjected to a number of biases. The main difficulties are related to incomplete sampling, but also arise from the DNA extraction methods, biases and errors introduced by the PCR amplification step for library construction, and unequal cloning efficiency. In contrast, completing the full-cycle rRNA approach allows a more comprehensive analysis through the direct quantification of organisms (Amann *et al.*, 1995). Although the very diverse communities typical of most natural environments would make it

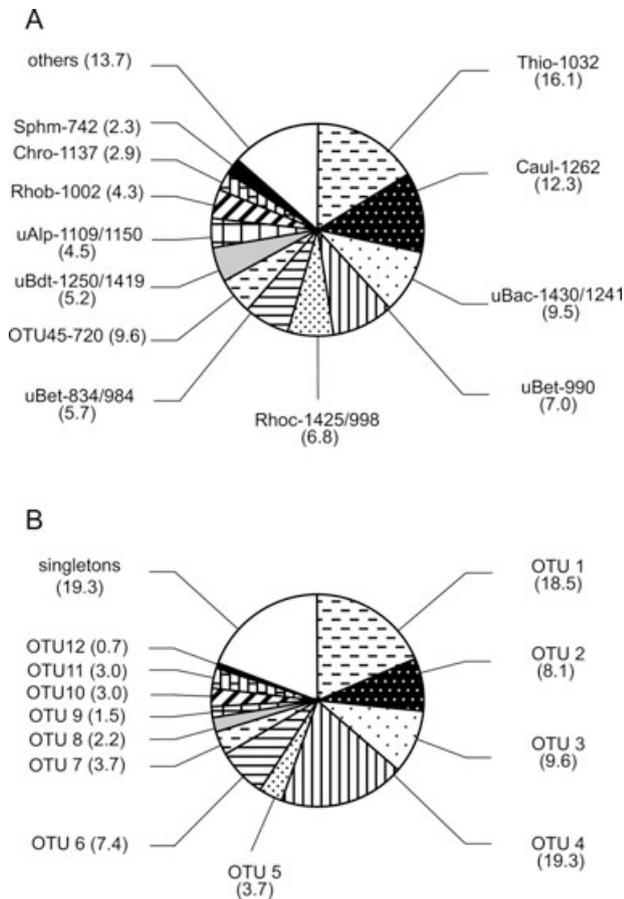


Fig. 3. Pie charts showing the bacterial community composition of the activated sludge. A. Distribution of cells determined by FISH and microscopic counting. The probes used for assignment to the different clades are indicated by their short names, together with the corresponding percentages. B. Operational taxonomic unit abundance as determined from the number of clones in the 16S rRNA clone library. Operational taxonomic units detected with specific probes indicated in A are displayed in the same order and contain the same patterns to facilitate the comparison.

difficult to apply this approach to all ecosystems, previous results from several laboratories converge in that activated sludge possesses relatively limited richness (Curtis *et al.*, 2002; Wagner *et al.*, 2002; Bramucci *et al.*, 2003). Our results agree with previous research showing the *Proteobacteria* as the most abundant group of *Bacteria* in activated sludge (Snaird *et al.*, 1997; Wagner *et al.*, 2002; Bramucci *et al.*, 2003; Eschenhagen *et al.*, 2003; Lozada *et al.*, 2004; Thomsen *et al.*, 2004). The extrapolated minimum number of bacterial species from the clones frequency in the gene library is in good agreement with previous 16S rRNA gene-based diversity surveys of different WWTPs and reactors (reviewed in Wagner *et al.*, 2002). Therefore, activated sludge constitutes a system amenable to analysis by quantitative FISH. Previous

studies showed that approximately 90% of total cells present in activated sludge could be detected with rRNA-targeted probes (Manz *et al.*, 1994; Juretschko *et al.*, 2002). To be entirely satisfactory, quantification of bacterial populations should rely upon the accomplishment of FISH in a statistically valid manner. It has been demonstrated that this task is altogether possible, even with difficult samples like activated sludge flocs, if appropriate controls are performed in conjunction with rigorous statistical analysis (Davenport and Curtis, 2004; Coskuner *et al.*, 2005).

The program DOTUR (Schloss and Handelsman, 2005) was used to assign sequences in the clone library systematically to OTUs at a genetic distance of 0.04. Clustering of sequences into this level of evolutionary distance

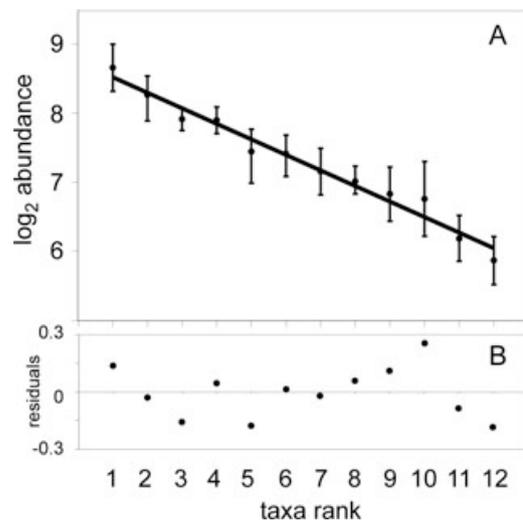


Fig. 4. (A) Rank abundance plots showing the distribution of the 12 most abundant taxa and (B) residuals. Taxa are ranked according to their abundances determined by FISH in descending order. Error bars correspond to 95% confidence intervals.

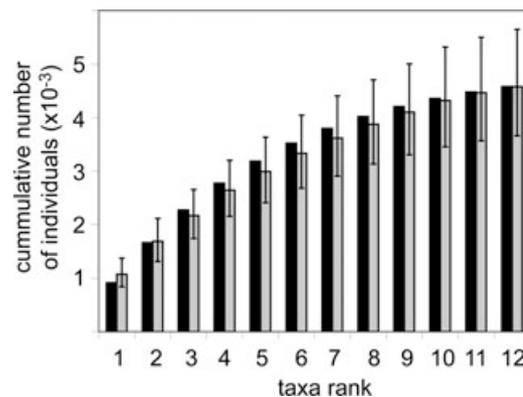


Fig. 5. Cumulative frequency distributions for observed taxa (grey bars) and expected distribution of taxa according to a geometric series, with $k = 0.15$ (black bars). The error bars represent 95% confidence intervals.

allowed us to design signature oligonucleotide probes for all major groups present in the sample.

For the measurement of the bacterial diversity in the WWTP, we have not limited our analysis to the estimation of the more frequently used ecological indexes (Magurran, 2004). Rather, we focused our attention on the more informative patterns of species abundance of the dominant species. Interestingly, we found that the distribution of most abundant bacteria in the investigated industrial activated sludge could be appropriately described by a geometric series. The observed abundance distribution could be rationalized as a resource apportionment model, assuming that the fraction of niche space is proportional to its abundance (May, 1975; Pielou, 1975). An interesting consequence of applying a niche-oriented model to the data is that an ecological meaning can be inferred. In general, geometric series are descriptors of species-poor communities under harsh environmental conditions (May, 1975). There are a multitude of factors operating on the activated sludge process, and therefore interpretation of niche apportionment is not straightforward. The composition of the microbial community is regulated simultaneously by the resources provided by the incoming wastewater, which drive the growth of microorganisms under certain environmental conditions (pH, redox potential, temperature, etc.), and the losses originated by predation, as well as to other sources of cell mortality such as phage infection and competition between members of the community.

Alternatively, the shape of the abundance distribution could be explained according to a neutral community model (Hubbell, 2001), which predicts that a geometric distribution might emerge as a consequence of low immigration rate (m) or from the relatively low size of the source metacommunity (θ). This might also be the case if a competitive advantage of species is introduced in the model (Sloan *et al.*, 2006), considering the high degree of specialization required for the biodegradation of the characteristic wastewater from an oil refinery.

A complete mechanistic understanding of all the ecological processes involved would be desirable, but at the present time unattainable. The description of abundance relationships does not specify a particular mechanism, but constitutes a starting point for the elucidation of the influence of the abiotic and biotic factors operating in the community.

Further work is needed to elucidate whether the observed distribution can withstand changes in species composition and abundance over time. We anticipate that there would be no correspondence between species identity and ranking order, i.e. any species could take any rank in terms of abundance (Tokeshi, 1993), in order to tolerate the existence of functional redundancy necessary for the maintenance of a stable process (Briones and Raskin, 2003).

More data have to be assembled in order to dismiss alternative distributions. For example, considering that the sample contains several tens of taxa, fitting a lognormal distribution with 12 observed taxa that group in only four abundance classes would imply that most of the taxa appear to the left of the veil line, making the fitting uncertain. The taxa abundance distribution might also be affected by the degree of phylogenetic discrimination, which in turn depends on the length of the DNA sequences that were analysed.

The abundance pattern of the less common species, not covered by this approach, might be assembled according to a different law (Tokeshi, 1990; Magurran and Henderson, 2003). Although this may represent a problem for the estimation of species richness, abundance patterns of dominant taxa should not be noticeably influenced (Tokeshi, 1993). However, less abundant members of the community might indeed have key functional roles in the overall process. Example of this situation is the case of nitrifiers. The lack of detection of representative of AOB not only in the rDNA library, but also in FISH is puzzling. A recent study of nitrifying bacteria in intermittently aerated reactors treating high ammonia wastewater has highlighted the limitations of the theoretical model to calculate the AOB fraction when the process is working at high volatile suspended solids concentration (Mota *et al.*, 2005). Additionally, part of this discrepancy could arise from limitations with the specificity of the probes used to target AOB or that yet undiscovered bacteria were responsible for the observed nitrification. As of today, we cannot explain the strong disagreement between the expected and the observed number of nitrifiers. The assessment of the actual diversity of ammonia oxidizer in this system is currently under investigation.

Nevertheless, it is important to stress that whereas the description of metabolic capabilities of surveyed species is beyond the range of this approach, the contribution of less abundant members to ecosystem production cannot and should not be neglected.

Using a 16S rRNA approach it is not possible to make certain that any OTU definition would allow one to describe community structure at a taxonomic level at which physiological traits are shared among the members of the groups, because closely related or even identical 16S rRNA environmental sequences may actually represent ecologically distinct populations (Casamayor *et al.*, 2002; Jaspers and Overmann, 2004; Hahn and Pockl, 2005; Coleman *et al.*, 2006). In order to assess the full extent of microdiversity and their relationship to ecosystem function, more information about the genetic and ecophysiological differences of closely related populations will be required.

By using the full rRNA cycle to describe the species distribution pattern of a complex bacterial assemblage,

this approach represents a starting point for exploring the biological background of taxa abundance patterns. It relies on the detection of the majority of individuals in the sample and it may be applied as an operational tool to examine how the distributions are affected by the operating conditions in engineered systems or more generally by specified environmental variables in natural ecosystems. Due to its quantitative nature, the approach could facilitate further the exploration of relationships between bacterial diversity, habitat heterogeneity and stability within specialized treatment systems or perturbed natural environments. The limitations originating from the threshold for FISH detection will certainly be addressed, exploiting constant improvements in methodology and protocols (Pernthaler *et al.*, 2002; Sekar *et al.*, 2003; Yilmaz *et al.*, 2006), automatic cell counting (Daims *et al.*, 2001; Pernthaler *et al.*, 2003), image analysis (Daims *et al.*, 2006) and flow cytometry, which may enable a more rapid and accurate analysis.

Experimental procedures

Samples

Activated sludge samples were collected from the anoxic and aerated basins of a modified Ludzack–Ettinger process receiving pretreated wastewater from an oil refinery industry. The WWTP treated approximately 3.2 million litres of wastewater per day, with a hydraulic retention time of 36 h. No sanitary sewage was discharged into this WWTP. The carbonaceous influent containing hydrocarbons (average $7.9 \pm 5.7 \text{ mg l}^{-1}$) and phenolic compounds (average $2.6 \pm 0.9 \text{ mg l}^{-1}$) was supplemented with phosphoric acid. The average influent 5 day biological oxygen demand and ammonia nitrogen levels were 186 ± 78 and $95 \pm 31 \text{ mg l}^{-1}$ respectively. The basin pH was maintained at 7.1 ± 0.3 by the addition of NaOH. Process temperature was $34 \pm 3^\circ\text{C}$. The WWTP was operated with a biological solids retention time of 35 days. The level of nitrification was > 98%, with a level of ammonium in the discharge $< 1.7 \text{ mg l}^{-1}$. Although originally designed for total nitrogen removal, the WWTP has not produced significant denitrification due to the low COD/N ratio.

Activated sludge samples were transported to the laboratory at room temperature and stored at -20°C . Samples for FISH analysis were fixed *in situ* with 4% paraformaldehyde, transported to the laboratory at 0°C , and stored at -20°C until analysis.

DNA extraction and library construction

Aliquots of the samples were processed immediately after their arrival at the laboratory. Extraction of DNA from sludge was described previously (Lozada *et al.*, 2006). A fragment of 16S rDNA gene of approximately 1.5 kb in length was amplified using S-D-Bact-0008-a-S-20 and S*-Univ-1492-a-A-18 primers in a 35-cycle PCR. Amplification products were gel-purified, ligated into the pGem T-Easy vector (Promega,

Madison, WI), and transformed into *Escherichia coli* DH10B cells by electroporation. Plasmid templates DNA from 192 transformants were prepared by standard alkaline lysis method. Single extension sequencing reactions with primer S*-Univ-1492-a-A-18 were carried out with BigDye terminator and run on ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA).

Phylogenetic analysis

Only sequences with reads longer than 600 bases were used, and they averaged approximately 700 bp. Each sequence was compared with the sequences in GenBank by BLASTN (Altschul *et al.*, 1990). Chimera sequences detected with the CHECK_CHIMERA software of the Ribosomal Database Project (Cole *et al.*, 2005) were checked manually; five sequences were excluded from further analysis. A total of 135 16S rDNA sequences were aligned by CLUSTALX. Distance matrix (DNADIST) and phylogenetic tree were built by using the NEIGHBOR program, which implement the neighbour-joining method. A bootstrap confidence analysis was performed by generating 100 replicated data sets with the SEQBOOT program and generating the consensus tree with the CONSENSE program. Phylogenetic trees were also constructed with the DNAML (maximum likelihood) and DNAPARS (parsimony) programs (all in the PHYLIP software package version 3.5). Trees were visualized using the ARB program (Ludwig *et al.*, 2004). Sequence assignment to OTUs was performed by using the furthest neighbour algorithm of DOTUR program (<http://www.plantpath.wisc.edu/fac/joh/dotur.html>). Operational taxonomic units were defined using a genetic distance of 0.04 between sequences (Schloss and Handelsman, 2005). *Thermotoga thermarum* gene for 16S rRNA (AB039769) was used as outgroup.

Library coverage

Coverage was calculated using the equation $C = [1 - (n_i/N)] \times 100$, where n_i is the number of single-occurrence OTUs and N is the number of 16S rRNA sequences examined (Good, 1953).

Oligonucleotide probes

Design of oligonucleotide probes was performed with the ARB software program (Ludwig *et al.*, 2004).

The hybridization conditions for new probes were optimized with *E. coli* transformed with plasmids containing the target sequence. Probe sequences developed in this study are listed in Table 1. The probes were named according to Alm and colleagues (1996). Oligonucleotides labelled with the sulfoindocyanine dye Cy3 were purchased from Metabion (Martinsried, Germany). Stringency conditions were determined for all probes, with *E. coli* BL21 cells transformed with clones from the 16S rDNA library containing the minimum number of mismatches with each probe as negative controls (Schramm *et al.*, 2002). The formamide concentration in the hybridization buffer was successively increased (and the NaCl concentration in the wash buffer decreased accordingly) until no fluorescent signal was observed from the

negative controls. Simultaneous hybridization with probes requiring different stringency was carried out by a successive hybridization protocol, as described (Wagner *et al.*, 1994).

Fluorescence in situ hybridization

Sludge flocs were small and irregularly shaped and no filaments could be observed. To facilitate cell counting flocs were disrupted by shaking 500 μ l of fixed activated sludge samples with 200 μ l of 0.5 mm glass beads in a bead-beater (Biospec, OK) during 60 s at 5000 r.p.m. One millilitre of PBS was added in order to dilute the sample. Hybridization was performed as described previously in gelatin-treated glass slides, imprinted with a hydrophobic coating between the wells (Lozada *et al.*, 2004).

Microscopic evaluation

Slides were examined with a Leica DM LB epifluorescence microscope equipped with a 50 W high-pressure mercury lamp (Wetzlar, Germany), and appropriate filter sets for Cy3 and 4,6-diamidino-2-phenylindole (DAPI) fluorescence. Images at a magnification of $\times 1000$ were recorded with a CCD camera. The number of fields of view (FOV) required to obtain a precision level of B was determined with the following formula:

$$n = [(t_{\alpha})^2 (s)^2] / (B)^2$$

where n is the sample size, t_{α} is the critical value obtained from the t -distribution for the appropriate degrees of freedom for a given type error rate α (1-confidence level), and s is the standard deviation (Davenport and Curtis, 2004). Counts were checked for normality using the Anderson–Darling normality test in MINITAB v11 (Minitab, State College, PA), in order to obtain 95% confidence limits, and when appropriate, transformed. When two probes were listed for a single group results were averaged.

Cell counts from DAPI-stained cells (UV excitation, total DAPI counts) were determined with the aid of Gene Tools software (Syngene, Cambridge, UK), using the following settings: analysis type: colony; image type: light colonies. Sensitivity and area were selected to count only targeted cells, and when it was not possible, cells were counted manually. Relative abundance of positive cells for each probe was calculated as percentage of DAPI-stained cells divided by the proportion of EUBmix-labelled cells of all DAPI-positive cells. Using the processed image, we counted the cells that hybridized with the Cy3-labelled specific probe above a specified threshold, determined using the nonsense probe S-D-Bact-0338-a-S-18 (non-eub 338) (Wallner *et al.*, 1993), to control for autofluorescence and non-specific binding. Images of both DAPI and Cy3-fluorescent cells could be stored separately and subsequently superimposed on the monitor screen in different colours. No fewer than 10 microscopic fields were observed, counting at least 5000 cells per microscopic slide.

Fitting to a geometric series

In a geometric series, the abundance of species, ranked from most abundant to least abundant is represented by:

$$n_i = NC_k k(1-k)^{i-1}$$

where k is the proportion of available niche space or resource occupied by each species; n_i the number of individuals in the i th species; N is the total number of individuals and $C_k = [1 - (1-k)^S]^{-1}$ is a constant, so that $\sum n_i = N$; where S is the total number of species.

k was first estimated by iterating the following equation

$$N_{\min}/N = (k/1-k)[(1-k)^S/1-(1-k)^S]$$

where N_{\min} is the number of individuals in the least abundant species, which in this case was taken as 10 cells in 10 observed fields. The calculation process consisted in adjusting k so that N_{\min}/N approached 10/5000, followed by the adjustment of S , iterating until $\sum n_i$ equalled the total number of cells that hybridized with FISH probes.

After the abundance of each taxa targeted by the FISH probes was estimated, the observed and expected abundances were compared using a δ -corrected Kolmogorov-Smirnov test, by using $F_{\delta} = (1-\delta)/(n-2\delta+1)$ as the observed cumulative relative frequency distribution and g_{δ} as the test statistics (Sokal and Rohlf, 1995). As k was obtained from the observed distribution, this constitutes a test of an intrinsic hypothesis. It should be noted that the Kolmogorov-Smirnov formalism results in a conservative test when the obtained frequency distributions are discrete, as is the case in the present study (Sokal and Rohlf, 1995).

Data deposition

Sequences of the small-subunits were deposited in GenBank (accession numbers DQ676682 to DQ676816). Oligonucleotide probes were submitted to probeBase (<http://www.microbial-ecology.net/probebase>).

Acknowledgements

We thank the anonymous reviewers for valuable comments. This work was partly funded by FONCyT (PICT, 2003, No. 01-14218) and CONICET (PIP 5548). E.L.M.F. is a fellow, and L.E. is a career member, of CONICET.

References

- Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**: 8966–8969.
- Alm, E.W., Oerther, D.B., Larsen, N., Stahl, D.A., and Raskin, L. (1996) The oligonucleotide probe database. *Appl Environ Microbiol* **62**: 3557–3559.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Amann, R.L., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L., and Lilley, A.K. (2005) The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

- Bramucci, M., Kane, H., Chen, M., and Nagarajan, V. (2003) Bacterial diversity in an industrial wastewater bioreactor. *Appl Microbiol Biotechnol* **62**: 594–600.
- Briones, A., and Raskin, L. (2003) Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr Opin Biotechnol* **14**: 270–276.
- Casamayor, E.O., Pedros-Alio, C., Muyzer, G., and Amann, R. (2002) Microheterogeneity in 16S ribosomal DNA-defined bacterial populations from a stratified planktonic environment is related to temporal changes and to ecological adaptations. *Appl Environ Microbiol* **68**: 1706–1714.
- Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., *et al.* (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**: D294–D296.
- Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.
- Coskuner, G., Ballinger, S.J., Davenport, R.J., Pickering, R.L., Solera, R., Head, I.M., and Curtis, T.P. (2005) Agreement between theory and measurement in quantification of ammonia-oxidizing bacteria. *Appl Environ Microbiol* **71**: 6325–6334.
- Curtis, T.P., and Sloan, W.T. (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr Opin Microbiol* **7**: 221–226.
- Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99**: 10494–10499.
- Daims, H., Ramsing, N.B., Schleifer, K.H., and Wagner, M. (2001) Cultivation-independent, semiautomatic determination of absolute bacterial cell numbers in environmental samples by fluorescence *in situ* hybridization. *Appl Environ Microbiol* **67**: 5810–5818.
- Daims, H., Lucker, S., and Wagner, M. (2006) DAIME, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**: 200–213.
- Davenport, R.J., and Curtis, T.P. (2004) Quantitative fluorescence *in situ* hybridisation (FISH): statistical methods for valid cell counting. In *Molecular Microbial Ecology Manual*. Kowalchuk, G.A., de Bruijn, F.J., Head, I.M., Akkermans, A.D.L., and van Elsas, J.D. (eds). Dordrecht, the Netherlands: Kluwer Academic, pp. 1487–1516.
- Davenport, R.J., Curtis, T.P., Goodfellow, M., Stainsby, F.M., and Bingley, M. (2000) Quantitative use of fluorescent *in situ* hybridization to examine relationships between mycolic acid-containing actinomycetes and foaming in activated sludge plants. *Appl Environ Microbiol* **66**: 1158–1166.
- Eschenhagen, M., Schuppler, M., and Roske, I. (2003) Molecular characterization of the microbial community structure in two activated sludge systems for the advanced treatment of domestic effluents. *Water Res* **37**: 3224–3232.
- Fernandez, A., Huang, S.Y., Seston, S., Xing, J., Hickey, R., Criddle, C., and Tiedje, J. (1999) How stable is stable? Function versus community composition. *Appl Environ Microbiol* **65**: 3697–3704.
- Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.
- Hahn, M.W., and Pockl, M. (2005) Ecotypes of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Appl Environ Microbiol* **71**: 766–773.
- Hong, S.H., Bunge, J., Jeon, S.O., and Epstein, S.S. (2006) Predicting microbial species richness. *Proc Natl Acad Sci USA* **103**: 117–122.
- Horner-Devine, M.C., Carney, K.M., and Bohannon, B.J. (2004) An ecological perspective on bacterial biodiversity. *Proc Biol Sci* **271**: 113–122.
- Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ, USA: Princeton University Press.
- Hughenoltz, P., and Huber, T. (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int J Syst Evol Microbiol* **53**: 289–293.
- Jaspers, E., and Overmann, J. (2004) Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysologies. *Appl Environ Microbiol* **70**: 4831–4839.
- Juretschko, S., Loy, A., Lehner, A., and Wagner, M. (2002) The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *Syst Appl Microbiol* **25**: 84–99.
- Kowalchuk, G.A., Stephen, J.R., De Boer, W., Prosser, J.I., Embley, T.M., and Woldendorp, J.W. (1997) Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class Proteobacteria in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Appl Environ Microbiol* **63**: 1489–1497.
- Kurata, S., Kanagawa, T., Magariyama, Y., Takatsu, K., Yamada, K., Yokomaku, T., and Kamagata, Y. (2004) Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Appl Environ Microbiol* **70**: 7545–7549.
- Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., *et al.* (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**: 804–808.
- Lozada, M., Itria, R.F., Figuerola, E.L.M., Babay, P.A., Gettar, R.T., de Tullio, L.A., and Erijman, L. (2004) Bacterial community shifts in nonylphenol polyethoxylates-enriched activated sludge. *Water Res* **38**: 2077–2086.
- Lozada, M., Figuerola, E.L., Itria, R.F., and Erijman, L. (2006) Replicability of dominant bacterial populations after long-term surfactant-enrichment in lab-scale activated sludge. *Environ Microbiol* **8**: 625–638.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Lunn, M., Sloan, W.T., and Curtis, T.P. (2004) Estimating bacterial diversity from clone libraries with flat rank abundance distributions. *Environ Microbiol* **6**: 1081–1085.
- Luyten, Y.A., Thompson, J.R., Morrill, W., Polz, M.F., and Distel, D.L. (2006) Extensive variation in intracellular sym-

- biont community composition among members of a single population of the wood-boring bivalve *Lyrodus pedicellatus* (Bivalvia: Terebinidae). *Appl Environ Microbiol* **72**: 412–417.
- Magurran, A.E. (2004) *Measuring Biological Diversity*. Oxford, UK: Blackwell Publishing.
- Magurran, A.E. (2005a) Ecology: linking species diversity and genetic diversity. *Curr Biol* **15**: R597–R599.
- Magurran, A.E. (2005b) Biological diversity. *Curr Biol* **15**: R116–R118.
- Magurran, A.E., and Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.
- Manz, W., Wagner, M., Amann, R., and Schleifer, K.-H. (1994) *In situ* characterization of the microbial consortia active in two wastewater treatment plants. *Water Res* **28**: 1715–1723.
- May, R.M. (1975) Patterns of species abundance and diversity. In *Ecology and Evolution of Communities*. Cody, M.L., and Diamond, J.M. (eds). Cambridge, MA, USA: Harvard University Press, pp. 81–120.
- Mota, C., Head, M.A., Ridenoure, J.A., Cheng, J.J., and de Los Reyes, F.L., III (2005) Effects of aeration cycles on nitrifying bacterial populations and nitrogen removal in intermittently aerated reactors. *Appl Environ Microbiol* **71**: 8565–8572.
- O'Mullan, G.D., and Ward, B.B. (2005) Relationship of temporal and spatial variabilities of ammonia-oxidizing bacteria to nitrification rates in Monterey Bay, California. *Appl Environ Microbiol* **71**: 697–705.
- Pernthaler, A., Pernthaler, J., and Amann, R. (2002) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68**: 3094–3101.
- Pernthaler, J., Pernthaler, A., and Amann, R. (2003) Automated enumeration of groups of marine picoplankton after fluorescence *in situ* hybridization. *Appl Environ Microbiol* **69**: 2631–2637.
- Pielou, E.C. (1975) *Ecological Diversity*. New York, NY, USA: Wiley Interscience.
- Polz, M.F., and Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.
- Qiu, X.Y., Wu, L.Y., Huang, H.S., McDonel, P.E., Palumbo, A.V., Tiedje, J.M., and Zhou, J.Z. (2001) Evaluation of PCR-generated chimeras: mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl Environ Microbiol* **67**: 880–887.
- Rittmann, B.E., Lapidou, C.S., Flax, J., Stahl, D.A., Urbain, V., Harduin, H., *et al.* (1999) Molecular and modeling analyses of the structure and function of nitrifying activated sludge. *Water Sci Technol* **39**: 51–59.
- Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.
- Schramm, A., Fuchs, B.M., Nielsen, J.L., Tonolla, M., and Stahl, D.A. (2002) Fluorescence *in situ* hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environ Microbiol* **4**: 713–720.
- Sekar, R., Pernthaler, A., Pernthaler, J., Warnecke, F., Posch, T., and Amann, R. (2003) An improved protocol for quantification of freshwater Actinobacteria by fluorescence *in situ* hybridization. *Appl Environ Microbiol* **69**: 2928–2935.
- Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., and Curtis, T.P. (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.
- Snaird, J., Amann, R., Huber, I., Ludwig, W., and Schleifer, K.-H. (1997) Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. *Appl Environ Microbiol* **63**: 2884–2896.
- Sokal, R.R., and Rohlf, F.J. (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. New York, NY, USA: New York W.H. Freeman.
- Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.
- Thompson, J.R., Marcelino, L.A., and Polz, M.F. (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* **30**: 2083–2088.
- Thomsen, T.R., Nielsen, J.L., Ramsing, N.B., and Nielsen, P.H. (2004) Micromanipulation and further identification of FISH-labelled microcolonies of a dominant denitrifying bacterium in activated sludge. *Environ Microbiol* **6**: 470–479.
- Tokeshi, M. (1990) Niche apportionment or random assortment: species abundance patterns revisited. *J Anim Ecol* **59**: 1129–1146.
- Tokeshi, M. (1993) Species abundance patterns and community structure. *Adv Ecol Res* **24**: 111–186.
- Wagner, M., Amann, R., Kampfer, P., Assmus, B., Hartmann, A., Hutzler, P., *et al.* (1994) Identification and *in situ* detection of Gram-negative filamentous bacteria in activated sludge. *Syst Appl Microbiol* **17**: 405–417.
- Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., and Daims, H. (2002) Microbial community composition and function in wastewater treatment plants. *Antonie Van Leeuwenhoek* **81**: 665–680.
- Wallner, G., Amann, R., and Beisker, W. (1993) Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**: 136–143.
- Yilmaz, L.S., Okten, H.E., and Noguera, D.R. (2006) Making all parts of the 16S rRNA of *Escherichia coli* accessible *in situ* to single DNA oligonucleotides. *Appl Environ Microbiol* **72**: 733–744.