Open

ORIGINAL ARTICLE

Unraveling assembly of stream biofilm communities

Katharina Besemer[1,2], Hannes Peter[3], Jürg B Logue[3], Silke Langenheder[3], Eva S Lindström[3], Lars J Tranvik[3] and Tom J Battin[1,2]

[1]*Department of Limnology, Vienna University, Vienna, Austria;* [2]*WasserCluster Lunz, Lunz, Austria and* [3]*Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden*

Microbial biofilms assemble from cells that attach to a surface, where they develop into matrix-enclosed communities. Mechanistic insights into community assembly are crucial to better understand the functioning of natural biofilms, which drive key ecosystem processes in numerous aquatic habitats. We studied the role of the suspended microbial community as the source of the biofilm community in three streams using terminal-restriction fragment length polymorphism and 454 pyrosequencing of the 16S ribosomal RNA (rRNA) and the *16S rRNA* gene (as a measure for the active and the bulk community, respectively). Diversity was consistently lower in the biofilm communities than in the suspended stream water communities. We propose that the higher diversity in the suspended communities is supported by continuous inflow from various sources within the catchment. Community composition clearly differed between biofilms and suspended communities, whereas biofilm communities were similar in all three streams. This suggests that biofilm assembly did not simply reflect differences in the source communities, but that certain microbial groups from the source community proliferate in the biofilm. We compared the biofilm communities with random samples of the respective community suspended in the stream water. This analysis confirmed that stochastic dispersal from the source community was unlikely to shape the observed community composition of the biofilms, in support of species sorting as a major biofilm assembly mechanism. Bulk and active populations generated comparable patterns of community composition in the biofilms and the suspended communities, which suggests similar assembly controls on these populations.

Introduction

Microbial biofilms develop from primary cells that attach to a surface, where they form microcolonies that eventually coalesce into matrix-enclosed communities (Battin *et al.*, 2007). Biofilm formation has been extensively studied in laboratory and medical systems that are typically composed of mono- or polycultures (Costerton *et al.*, 1995; Hall-Stoodley *et al.*, 2004). Such systems, while useful to test basic concepts in microbiology, contrast the massive microbial diversity generally encountered in natural ecosystems (Sogin *et al.*, 2006; Newton *et al.*, 2011). In numerous aquatic ecosystems, surface-attached biofilms assemble from the microbial diversity contained in the overlying water. According to metacommunity theory (Leibold *et al.*, 2004; Holyoak *et al.*, 2005), local (abiotic environment, biotic interactions) and regional (dispersal) processes regulate the assembly of local communities. By viewing biofilms as microbial landscapes, their community assembly can be studied according to metacommunity ecology theory (Battin *et al.*, 2007). Mechanistic insight into community assembly is crucial to better understand the functioning of biofilms, which drive key ecosystem processes in streams (Singer *et al.*, 2010; Peter *et al.*, 2011).

Available knowledge on biofilm community assembly in nature is scarce and largely based on molecular fingerprinting techniques. For instance, using denaturing gradient gel electrophoresis, Jackson *et al.* (2001) and later Lyautey *et al.* (2005) studied successional changes in lake and river biofilms. Essentially, their findings suggest that biofilm assembly is not a random process, and that certain bacterial groups contribute more to biofilm formation than others. A conceptual model proposed by Jackson *et al.* (2001) suggests elevated bacterial diversity during initial biofilm formation and decreasing diversity as biofilm growth progressed as a result of the combined effects of niche availability and competition. Besemer *et al.* (2007) compared community succession in stream biofilms and found consistent differences between the biofilm and stream water communities,

---

which indicate the existence of a specific biofilm community.

On the basis of these previous findings, we hypothesize that the assembly of a local biofilm community is not a mere reflection of the source community suspended in the overlying stream water. The compositions of the biofilm and the suspended communities are thus anticipated to differ. We argue that stream water transports bacteria from multiple sources within the catchment, whereas biofilms, according to the species sorting perspective in metacommunity theory (Leibold et al., 2004; Holyoak et al., 2005), specifically select for certain taxonomic groups. We also hypothesized that the diversity of the suspended community may exceed the diversity in biofilms, as various sources within the catchment continuously feed the community suspended in the stream water. We are aware that niche diversification could, nevertheless, support a high diversity in biofilms (Jackson et al., 2001; Besemer et al., 2007). Fingerprinting methods as used in these earlier studies are, however, limited in their ability to detect and quantify rare species (Blackwood et al., 2007; Bent and Forney, 2008). In this study, we used a dual approach to explore possible mechanisms of biofilm community assembly in three headwater streams within the same catchment. We applied terminal-restriction fragment length polymorphism (T-RFLP), which we supplemented with 454 pyrosequencing to gain deeper insight into community assembly. We analyzed both the 16S ribosomal RNA (rRNA; as a measure for the active fraction of a community) and the 16S rRNA gene (for the bulk community) to test whether the active members of the suspended microbial community differ from the inactive members in their ability to contribute to biofilm formation.

Materials and methods

Biofilm growth and sampling procedure
Streambed (hyporheic) biofilms were grown on initially sterile, sintered, borosilicate glass beads (2 mm diameter) deployed in three headwater streams. Beads were exposed for colonization from the suspended microbial community for 3 weeks during snowmelt in April, when terrestrial–aquatic connectivity was high. The streams are located in Fiby Urskog (N 59° 53′ 7″ E 17° 20′ 43″), a protected forest area close to Uppsala, Sweden. One of the streams (referred to as 'outflow stream' hereafter) is the outflow of lake Fibysjön. Downstream, it merges with a small humic-rich ditch (referred to as 'humic stream' hereafter) that drains a forest, into a confluence (referred to as 'confluence' hereafter). Water chemistry was largely similar in all three streams, except for the concentration of dissolved organic carbon, which was, on average, 75.2 mg C L^{-1} in the humic, 34.0 mg C L^{-1} in the outflow and 32.9 mg C L^{-1} in the confluence (Supplementary Table 1).

Glass beads were packed into nets (1 mm mesh size) that were cased in perforated pipes (diameter: 5 cm, length: 20 cm). Triplicate pipes were installed in the thalweg (30 cm above bottom) of the respective stream parallel to the main flow direction to allow continuous flow through of the bead packages. During the 3-week colonization period, we sampled stream water seven times for the analysis of the suspended community. Samples were filtered onto sterile 0.2 µm filters (GSWP filter, Millipore, Solna, Sweden) and frozen (−80 °C). Beads with biofilms were sampled after 3 weeks. Aliquots were suspended in sterile (autoclaved and 0.2 µm filtered) water and sonicated (10 min, 40 W output; Branson Sonifier, Danbury, CT, USA) to detach cells. Suspended cells were concentrated on sterile filters (0.2 µm GSWP filter, Millipore) and stored (−80 °C) pending for further processing.

Nucleic acid extraction and reverse transcription
Nucleic acids were extracted from biofilms and suspended communities using the PowerSoil DNA Isolation Kit (MoBio, Carlsbad, CA, USA) and the Easy-DNA Kit (Invitrogen, Paisley, UK) omitting the RNase step (Logue and Lindström, 2010). Although the PowerSoil DNA Isolation Kit is designed to extract DNA only, the resulting DNA and RNA yields were higher than those obtained with the Easy-DNA Kit; we therefore used the PowerSoil DNA Isolation Kit also for RNA.

Reverse transcription of RNA into complementary DNA was performed as described by Logue and Lindström (2010). Briefly, an aliquot of the nucleic acid extract was subjected to DNA digestion with DNase I (Invitrogen) for 15 min at room temperature following the manufacturer's recommendations. Absence of DNA was verified by PCR of the DNA digests as described below. RNA was transcribed at 42 °C for 50 min using SuperScript II reverse transcriptase and random primer oligonucleotides (Invitrogen), followed by an enzyme inactivation step at 70 °C for 15 min. Samples without reverse transcriptase served as negative controls.

T-RFLP analysis
The PCR primers used for T-RFLP analysis were the hexachlorofluorescein-labeled bacteria-specific primer 27F (5′-AGRGTTTGATCMTGGCTCAG-3′) and the universal primer 519R (5′-GWATTACCGC GGCKGCTG-3′). Each 50 µl PCR mixture contained both primers at 0.4 µmol l^{-1} (Invitrogen), each deoxynucleoside triphosphate at 0.2 mmol l^{-1} (Invitrogen), 75 µg bovine serum albumin (New England BioLabs, Ipswich, UK), MgCl$_2$ at 3.5 mmol l^{-1}, 1.5 U of DyNAzyme II DNA polymerase and the recommended PCR buffer (Finnzymes, Espoo, Finland). The amplification protocol consisted of an initial denaturation step of 94 °C for 3 min, 25 cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 45 s, extension at 72 °C for 1 min and a final extension

step at 72 °C for 10 min. Each PCR was run in triplicates and subsequently pooled. PCR products were cleaned applying the QIAquick PCR Purification kit (Qiagen, Hilden, Germany) and quantified using agarose gel electrophoresis in combination with the Low DNA Mass Ladder (Invitrogen).

The fluorescently labeled PCR products were digested separately with the restriction enzymes *Hae*III and *Hinf*I (New England BioLabs). Restriction digests were performed according to Logue and Lindström (2010). The product was subjected to capillary electrophoresis in an ABI 3730XL DNA Analyzer (Uppsala Genome Center, Uppsala, Sweden) using the size marker GS 500 Rox (Applied Biosystems, Foster City, CA, USA). The electropherograms were analyzed using the Peak Scanner software (Applied Biosystems). The relative contribution of the respective operational taxonomic units (OTUs) to the community was estimated as peak height divided by the cumulative peak height of the given sample.

454 pyrosequencing

To reduce the number of samples for 454 pyrosequencing, equal amounts of extracted or transcribed DNA of the suspended communities from the seven sampling dates were pooled to yield time-integrated samples for each active and bulk community from the three streams. Multiplex amplicon sequencing was then performed on the six biofilm samples and the six time-integrated suspended community samples. The V3 and V4 regions of bacterial *16S rRNA* genes were amplified using the fusion primers 341F (5′-CCTACGGGNGG CWGCAG-3′) and 805R (5′-GACTACHVGGGTATCT AATCC-3′), containing the 454 FLX adaptors and a sample-specific multiplex identifier (Andersson *et al.*, 2008). Each 50 µL PCR mixture contained each primer at 0.5 µmol l^{-1}, each deoxynucleoside triphosphate at 0.25 mmol l^{-1} (Invitrogen), MgCl$_2$ at 1.5 mmol l^{-1}, 1.25 U of Phusion High-Fidelity DNA Polymerase and the recommended PCR buffer (Finnzymes). Triplicate PCR products for each sample were pooled, purified using the QIAquick Gel Extraction Kit (Qiagen) and quantified using gel electrophoresis and the Low DNA Mass Ladder (Invitrogen). Equal amounts of the barcoded PCR products were mixed and submitted to the KTH Biotechnology Sequencing Center (Stockholm, Sweden) for pyrosequencing on a 454 GS20 FLX platform.

The obtained pyrosequencing data were denoised using the software package AmpliconNoiseV1.0 (Quince *et al.*, 2011). Pyrosequencing flowgrams with an exact match to the primer and multiplex identifier sequences were preclustered with PyroNoise (AmpliconNoiseV1.0) to remove pyrosequencing noise. PCR single base errors were corrected using SeqNoise (AmpliconNoiseV1.0), a sequence-based clustering method, which performs the alignment of the sequences. The Perseus algorithm was used to check for chimeras with an intercept of $\alpha = -7.5$ and coefficient of $\beta = 0.5$ (Quince *et al.*, 2011). This procedure reduced the originally 229 026 flowgrams to 118 612 reads. The denoised reads were clustered to OTUs, with a complete linkage algorithm on a 97% sequence identity level. The taxonomic affiliation of the OTUs was determined using a naïve Bayesian rRNA Classifier (Wang *et al.*, 2007) and a confidence threshold of 80%.

Data analysis

Similarity matrices of community compositions based on T-RFLP and 454 pyrosequencing data were calculated using the presence/absence-based Sørensen index and the relative abundance-based Horn index. These similarity indices were chosen because they are independent from alpha-diversity and therefore consistent with valid beta-diversity indices (Jost, 2007). Nonmetric multidimensional scaling (nMDS) analysis was performed on the similarity matrices to visualize patterns of community composition. Similarity matrices obtained for the *rRNA* gene-based (referred to as bulk community hereafter) and the rRNA-based (referred to as active community hereafter) communities were compared using Mantel's matrix randomization test (Mantel, 1967) with Pearson's correlation and 999 permutations. Diversities were estimated applying indices of the Hill family (Hill, 1973), namely, richness and the number equivalents of the Shannon entropy. Data analysis was performed with PAST (Hammer *et al.*, 2001) and R 2.13.0 (R Development Core Team, 2011).

Using the 454 pyrosequencing data, we performed a random sampling procedure to estimate the probability that a biofilm community represented a random subsample of the respective suspended source community in the stream water. Each tested sample pair consisted of a biofilm and a suspended community, either bulk or active, from the same stream, respectively. OTUs were sampled from the suspended community with replacement until the number of OTUs in this randomly assembled community equaled the richness of the respective biofilm community. This procedure was repeated to yield 1000 random subsamples of each suspended community. The probability of the biofilm community to fall within the distribution of these random subsamples was calculated as the percentage of the distances of the random subsamples to their centroid, which were as high or higher than the distance of the biofilm community to the centroid. The biofilm community data set was reduced to OTUs, which occurred also in the respective suspended community, thereby increasing the chance of the biofilm community to resemble the suspended community. The estimated differences between the biofilm community and random subsamples of the suspended community can therefore be regarded as conservative.

Rarefaction curves for the 454 pyrosequencing data were computed using the AmpliconNoise software package. Rank-abundance curves were constructed from relative OTUs abundances obtained form 454 pyrosequencing data. Linear regression models were fitted to each curve after log transformation of the rank and abundance data. The slopes of these regression models were used as a simple descriptive statistic of community structure (Ager *et al.*, 2010) and were compared using Student's *t*-test. The 'true richness' of the communities was estimated by Bayesian fitting of the OTUs abundances obtained by 454 pyrosequencing to the Sichel distribution (Sichel, 1974) using the Diversity Estimation software according to Quince *et al.* (2008). The Sichel distribution was chosen as the best model to describe OTU abundances based on deviance information criterion calculation (Spiegelhalter *et al.*, 2002).

Results

Community composition

A total of 141 and 126 OTUs were found by T-RFLP analysis with the enzymes *Hae*III and *Hinf*I, respectively. OTUs from both enzymatic digestions were combined for further analysis. nMDS analyses of both the presence/absence-based Sørensen and the relative abundance-based Horn similarity matrices revealed clear differences between biofilm and suspended communities (Figures 1a and b). Although considerable variation existed among the suspended communities from all streams and among the different sampling times, biofilm and suspended communities did not overlap. Biofilm communities from all three streams were similar, and no relation with their respective suspended counterpart could be observed. These patterns were congruent for the bulk (*16S rRNA* gene based) and the active (16S rRNA based) communities, even though differences in community compositions of bulk and active communities are apparent from the nMDS analysis. Mantel's test confirmed significant correlations between the similarity matrices of the bulk and the active communities (Sørensen index: $r = 0.82$, $P < 0.01$, $n = 23$; Horn index: $r = 0.79$, $P < 0.01$, $n = 23$).

The denoised 454 pyrosequencing data set consisted on average of 9884 ± 1321 reads per sample, which clustered into 7512 OTUs at a 97% sequence similarity level. The sequence data are available at the NCBI Sequence Read Archive under the accession number SRX099353. A total of 4899 (that is, 65%) of the detected sequences were singletons. In all, 6270 OTUs occurred only in the suspended community, 556 OTUs only in biofilm and 686 OTUs were shared by both communities. Applying a confidence threshold of 80% to the Bayesian classifier, 99.86% of all reads were classified as bacteria, 0.01% were classified as Archaea and 0.12% failed to be classified to any domain. nMDS
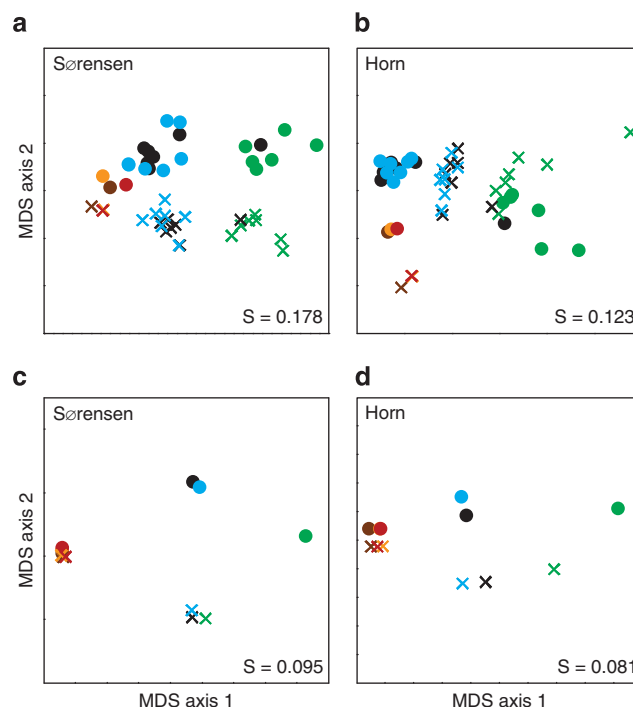

Figure 1 nMDS analysis of the microbial community compositions estimated by T-RFLP (**a**, **b**) and 454 pyrosequencing (**c**, **d**), calculated from the presence/absence-based Sørensen index (**a**, **c**) and the abundance-based Horn index (**b**, **d**). Kruskal's standardized stress values (S) below 0.2 indicated acceptable representation of the calculated similarities. Circles represent the bulk (*16S rRNA* gene based), crosses the active (16S rRNA-based) community compositions, brown the biofilm community humic stream, orange the biofilm community outflow stream, red the biofilm community confluence stream, green the suspended community humic stream, blue the suspended community outflow stream and turquoise the suspended community confluence stream.

analyses on 454 pyrosequencing data yielded similar patterns of community compositions as T-RFLP data, showing no resemblance between biofilm and suspended communities from the same stream (Figures 1c and d). Analysis of the *16S rRNA* gene and 16S rRNA gave accordant patterns, as confirmed by Mantel's correlations (Sørensen index: $r = 0.98$, $P < 0.05$, $n = 6$; Horn index: $r = 0.93$, $P < 0.01$, $n = 6$).

To test for species sorting as a possible mechanism of biofilm assembly, we compared the biofilm communities with random subsamples of the suspended communities that might result from purely stochastic immigration to an empty habitat patch from a source community. The bulk and the active biofilm communities of all three streams differed significantly from the random assemblages produced (probability of the biofilm community to fall within the distribution of the random subsamples, $P < 0.001$; Figure 2).

Microbial biodiversity

For T-RFLP data, OTU richness was generally higher in the suspended than in the biofilm communities, whereas the number equivalents of the Shannon
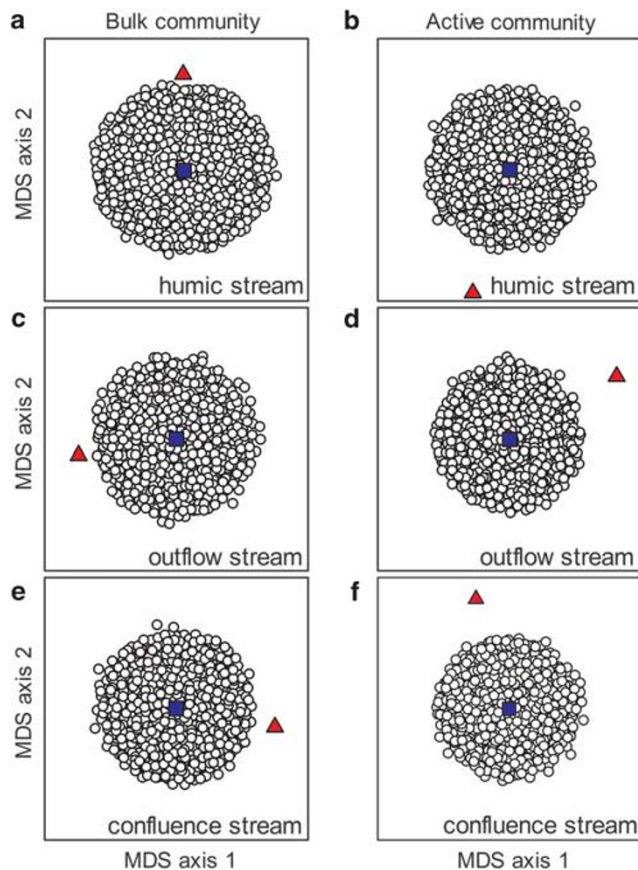
Figure 2 nMDS analysis visualizing the results of a random sampling procedure to estimate the probability that the biofilm communities represented random samples of their respective suspended source communities. A total of 1000 random subsamples of the suspended communities were assembled for each sample pair. White circles represent the random subsamples of the suspended communities, red triangle the biofilm community and blue cube the suspended community; humic stream (**a**, **b**), outflow stream (**c**, **d**), confluence stream (**e**, **f**), bulk (**a**, **c**, **e**) and active (**b**, **d**, **f**) communities.

entropy did not show any clear patterns (Figure 3a). The active fraction exhibited similar richness and Shannon entropy estimates as the bulk communities without showing a consistent difference.

Bacterial OTU richness estimates by 454 pyrosequencing were 3–7 times higher in the suspended than in the respective biofilm communities. The number equivalents of the Shannon entropy estimates were 4–22 times higher in the suspended than in the biofilm communities (Figure 3b). Both measures indicated higher diversity in the bulk community than in the active community, with the exception of the Shannon entropy of the biofilm in the outflow stream. To assess the importance of rare species to the observed patterns of diversity and to compare results from 454 pyrosequencing with T-RFLP analysis, a threshold of 0.2% contribution to the community was applied to the 454 pyrosequencing data. The threshold was chosen because 0.2% was the percentage, which was represented by the lowest T-RFLP peaks considered. Obtained patterns
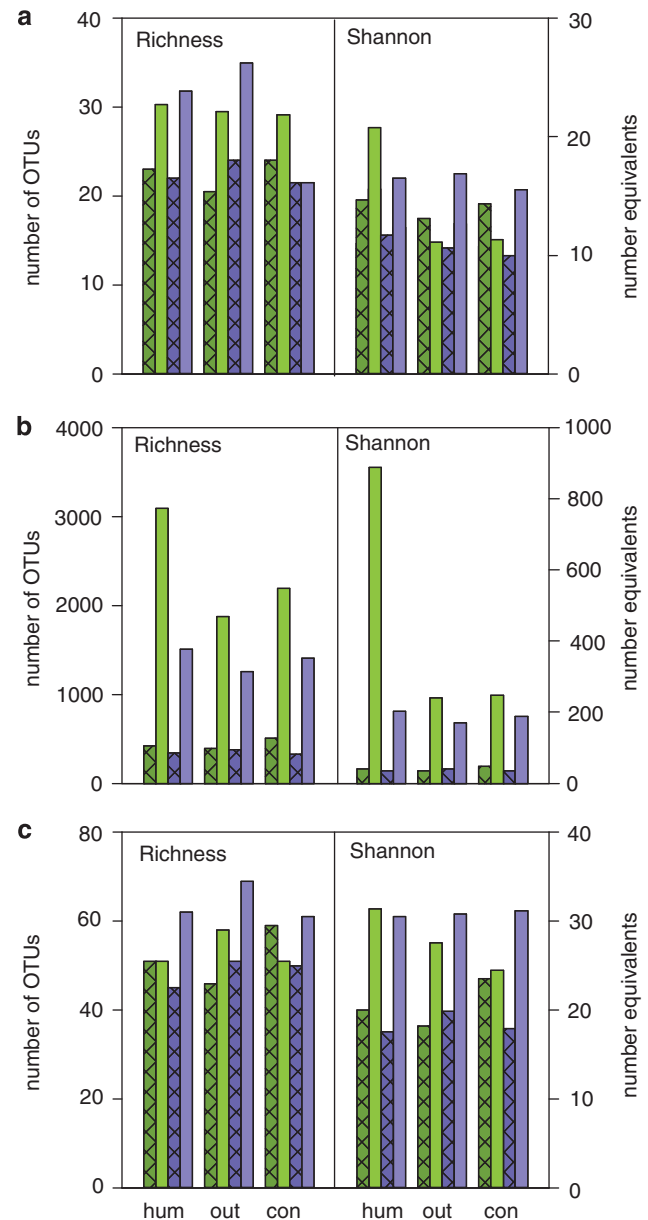

Figure 3 Microbial diversity in biofilm and suspended community, as estimated by T-RFLP (**a**) and 454 pyrosequencing (**b**), calculated as richness and the number equivalents of the Shannon entropy. A threshold of 0.2% contribution to the community was applied to the 454 pyrosequencing data to compare results from 454 pyrosequencing with T-RFLP analysis (**c**). con, confluence stream; hum, humic stream; out, outflow stream. Cross-hatched bars represent the biofilm community, solid bars the suspended community, green bars the bulk community and blue bars the active community.

and diversity estimates were in the same order of magnitude as values estimated by T-RFLP; on average, one T-RFLP-based OTU corresponded to two OTUs as defined by 454 pyrosequencing (Figure 3c). The reduced 454 pyrosequencing data set failed to show clear differences in diversity between the suspended and biofilm communities and between active and bulk community, respectively. Instead, the reduced 454 pyrosequencing data
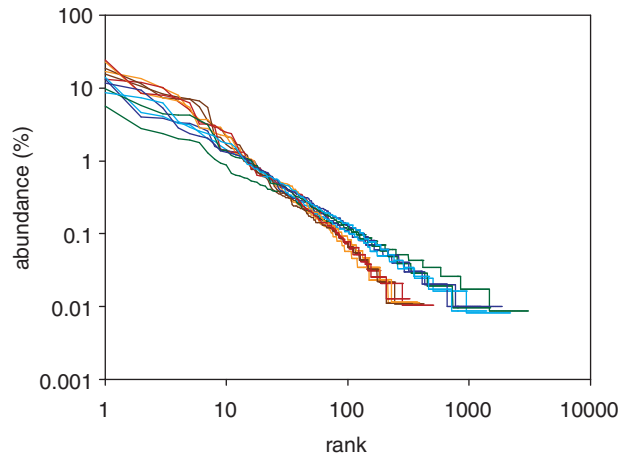
Figure 4 Rank-abundance curves of biofilm and suspended communities for relative abundances obtained from 454 pyrosequencing data. Curves are displayed in log–log scale for clarity. Colors are same as in Figure 1.

Figure 5 'True diversity' estimates (medians with 95% confidence interval) for the biofilm and suspended communities, calculated by fitting Sichel distribution curves to the abundance distributions obtained from the 454 pyrosequencing data. Colors are same as in Figure 1.

correlated with the T-RFLP data (richness: Pearson's $r = 0.81$, $P < 0.01$; Shannon entropy: Pearson's $r = 0.67$, $P < 0.05$).

Rarefaction curves did not reach an asymptote, indicating a significant amount of undetected diversity, especially for the suspended communities (Supplementary Figure 1). The rank-abundance distributions showed a strong dominance of a few OTUs and a long tail of rare OTUs (Figure 4). The dominance of the most abundant OTUs was higher in the biofilms, and the number of rare OTUs was higher in the suspended communities. Accordingly, the slopes of the regression models fitted to the rank-abundance curves ($r^2 > 0.95$, $P < 0.001$ for all models; Supplementary Table 2) differed significantly between suspended and biofilm communities (*t*-test, $P < 0.001$, $n = 6$). Rank-abundance curves of bulk and active communities exhibited no significant difference. Computed values of 'true richness' ranged from 526 to 1347 in biofilms and from 2854 to 6512 in the suspended communities (Figure 5). Richness of the bulk community was consistently higher than of the respective active fraction.

Taxonomic composition
Overall, 3603 OTUs (that is, 48% of all OTUs), representing 79% of all reads, could be assigned to a class at a confidence threshold of 80%. Biofilm OTUs were allocated to 29 classes belonging to 14 phyla; OTUs of the suspended community were allocated to 48 classes of 24 phyla. Those classes contributing most to the observed diversity were present in the biofilm and suspended communities, although in several cases the distribution of their relative abundance indicated a preference for one of the two life forms (Figure 6). *Betaproteobacteria* accounted for more than one-third and one-fourth of the reads in the biofilm and the suspended
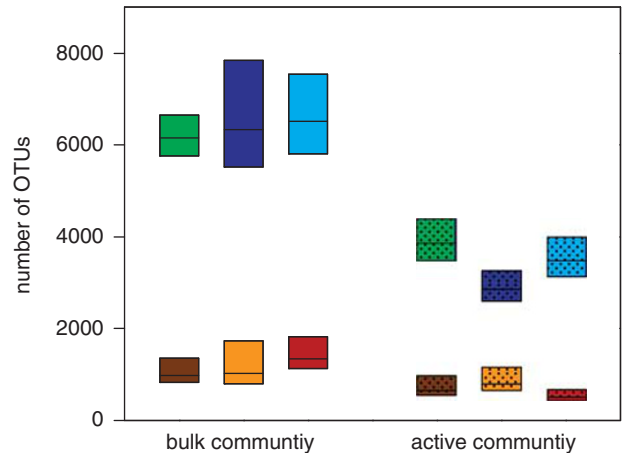
community, respectively. *Actinobacteria, Sphingobacteria* and *Alphaproteobacteria* contributed similarly to communities, whereas *Flavobacteria, Gammaproteobacteria* and *Bacilli* were relatively more abundant in biofilms than in the suspended communities. *Chlamydiae, Deltaproteobacteria* and members of the *OD1* group were relatively more abundant in the suspended than in the biofilm communities; a number of chloroplasts of eukaryotic algae were found in the suspended communities (Figure 6).

Generally, bulk and active communities showed similar taxonomic compositions. *Alpha-, Beta-, Gammaproteobacteria* and *Bacilli* occurred at higher relative abundance in the active than in the bulk biofilm community, *Flavobacteria, Sphingobacteria and Actinobacteria* at lower relative abundance. In the suspended communities, *Alphaproteobacteria* constituted a higher percentage of the active community, whereas *Flavobacteria, Actinobacteria, Chlamydiae* and the *OD1* group were less abundant in the active than in the bulk community.

In total, 1606 OTUs (21% of all OTUs) were classified to the genus level, representing 52% of all reads. The three most common genera were *Acidovorax* (1 OTU), *Flavobacterium* (56 OTUs) and *Polynucleobacter* (7 OTUs) in both the biofilm and suspended communities, although in different order. Together, they contributed between 33% and 41% to the individual biofilm communities and between 13% and 21% to the suspended communities (Table 1). A nMDS analysis of a Horn similarity matrix, including only the OTUs of these three most common genera, revealed patterns similar to those obtained from the whole communities, showing a clear separation of biofilm and suspended communities, as well as bulk and active communities (Supplementary Figure 2). Few genera showed different abundances in the bulk and active
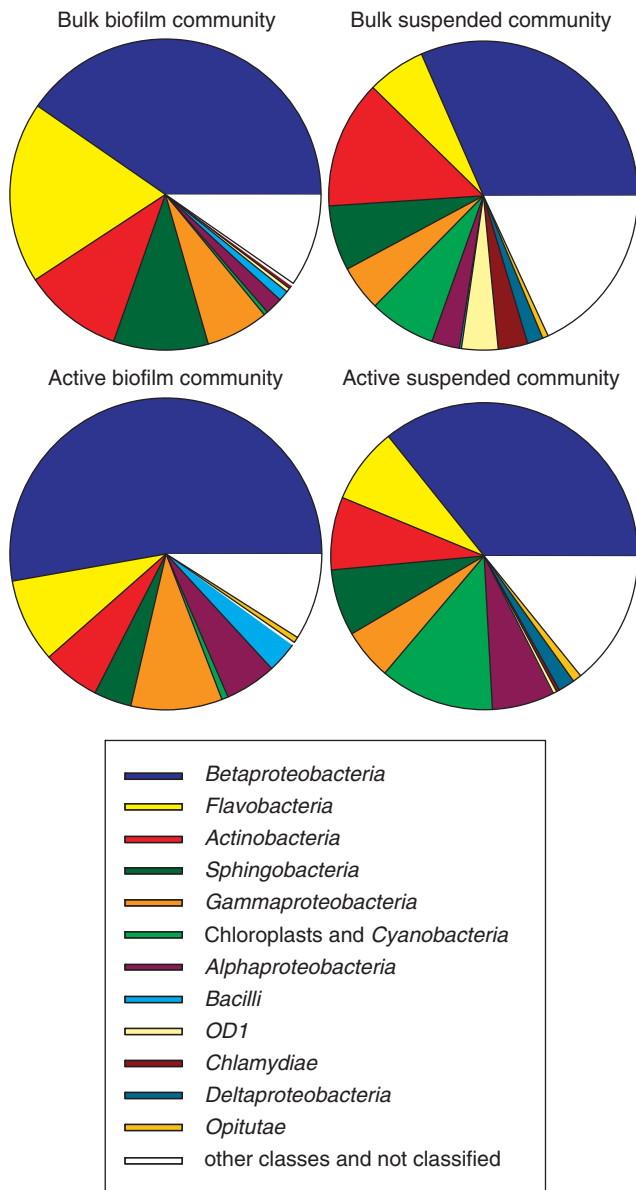
Figure 6 Relative abundances of the most important phylogenetic classes in the biofilm and suspended communities. Each pie chart represents the pooled data from all the three investigated streams.

communities. *Arcicella* constituted 5% of the bulk biofilm community but was not among the 10 most abundant genera in the active community. *Pseudomonas* was among the most abundant genera only in the active biofilm community. Two of the most common genera in the suspended community were identified as chloroplasts of eukaryotic algae.

Discussion

Methodological constraints have until recently hindered the accurate measurement of microbial diversity (Lunn *et al.*, 2004; Quince *et al.*, 2008),

and diversity estimates for the microbial communities in streams and rivers remain scarce (Vishnivetskaya *et al.*, 2011). Our estimates derived from 454 pyrosequencing data are the first to provide comprehensive insights into the microbial diversity contained in stream water and streambed biofilms. The 'true richness' estimates for the suspended communities are comparable to reports from soils, whereas the lower richness of the biofilms are similar to values from the ocean, as computed by Quince *et al.* (2008).

Diversity, as derived by 454 pyrosequencing, was consistently lower in the biofilm communities than in the suspended communities. This agrees with the generally lower slopes of the rank-abundance curves for the suspended than for the biofilm communities. Pommier *et al.* (2010) suggested that low-slope rank abundance distributions for bacterial communities in coastal waters resulted from the mixing of terrestrial and deep-water taxa. This would be in accordance with our hypothesis that various sources of bacterial species within the catchment support a diverse community suspended in the stream water. The occurrence of typical soil bacteria, such as members of the *Deltaproteobacteria* and *OD1* division (Spring *et al.*, 2000; Harris *et al.*, 2004; Elshahed *et al.*, 2005), in the suspended communities supports this notion. This is also in line with results from a recent meta-analysis of published environmental sequences showing that the taxonomic profiles of freshwater and terrestrial habitats widely overlap (Tamames *et al.*, 2010), which makes particularly sense for headwaters where the integration with the landscape is most pronounced (Battin *et al.*, 2008). Headwater streams might thus be considered as important terrestrial–aquatic links that collect bacterial diversity from the surrounding landscape into a source community that potentially seeds the benthic biofilms.

We integrated the samples of the suspended stream water communities over time for 454 pyrosequencing analysis to represent the full diversity of the microbes, potentially seeding the biofilm. Considering the short residence time of the stream water with the suspended bacteria and assuming that the temporal dynamics of the suspended community was higher than captured by our sampling scheme, we likely missed some of the suspended diversity. Accordingly, the finding of higher diversity in the suspended community as compared with the biofilms may be conservative. However, we studied relatively young rather than mature biofilms, where diversity may not have reached its maximum yet (Jackson *et al.* 2001).

The dominance of relatively few OTUs and a long tail of rare OTUs is typical for rank-abundance curves of microbial communities (Schwalbach *et al.*, 2004; Pommier *et al.*, 2010). Such rank-abundance curves have been postulated to be composed of a set of abundant taxa, performing most ecosystem functions, and of a seed bank containing rare taxa

Table 1 List of the ten most common genera in biofilm and suspended communities

Bulk biofilm community (%)	Active biofilm community (%)	Bulk suspended community (%)	Active suspended community (%)
Flavobacterium (16.3)	*Acidovorax* (21.9)	*Polynucleobacter* (9.3)	*Acidovorax* (9.3)
Acidovorax (12.0)	*Polynucleobacter* (8.1)	*Acidovorax* (4.6)	*Flavobacterium* (5.7)
Polynucleobacter (9.5)	*Flavobacterium* (7.5)	*Flavobacterium* (4.0)	*Polynucleobacter* (4.0)
Arcicella (5.2)	*Polaromonas* (6.7)	*OD1 genera incertae sedis* (2.9)	*Cryptomonadaceae* chloroplast (3.6)
Alkanindiges (3.6)	*Alkanindiges* (5.6)	*Sediminibacterium* (2.5)	*Bacillariophyta* chloroplast (2.8)
Polaromonas (3.6)	*Pseudomonas* (2.1)	*Cryptomonadaceae* chloroplast (1.5)	*Sediminibacterium* (2.3)
Sediminibacterium (2.2)	*Sediminibacterium* (1.7)	*Bacillariophyta* chloroplast (1.4)	*Duganella* (0.7)
Methylophilus (1.8)	*Beijerinckia* (1.2)	*Arcicella* (1.0)	*Verrucomicrobia subdivision 3 genera incertae sedis* (0.7)
Pedobacter (0.9)	*Comamonas* (1.1)	*Methylophilus* (0.8)	*Methylophilus* (0.6)
Fluviicola (0.8)	*Herbaspirillum* (1.0)	*TM7 genera incertae sedis* (0.6)	*Arcicella* (0.6)

Numbers indicate the average relative abundance of the genus in the communities from all the three streams.

(Pedrós-Alió, 2006). This rare biosphere has since been reported to contain a large proportion of active taxa (Jones and Lennon, 2010) and to be subjected to environmental controls (Andersson *et al.*, 2010; Campbell *et al.*, 2011). Other studies found that rare phylotypes tend to stay rare, arguing against the seed bank hypothesis (Galand *et al.*, 2009; Kirchman *et al.*, 2010). If the abundant OTUs were actively growing while a large part of rare OTUs was inactive, we would expect steeper slopes in the rank-abundance curves of the active community than of the bulk community. However, the rank-abundance curves of bulk and active communities were indistinguishable in the present study, indicating that at least a certain fraction of the rare OTUs was active. These rare but active populations may be controlled by top–down forces or competition; however, they have the potential to increase in abundance, which supports the idea that microbial rank-abundance curves may be highly dynamic (Jones and Lennon, 2010).

The fact that the composition of the suspended communities differed to some extent among the three investigated streams while the biofilm communities were similar is evidence that biofilm assembly did not simply reflect differences in the source communities. Furthermore, simulated biofilm communities from random sampling of the respective suspended community demonstrate that stochastic dispersal from the source community was unlikely to shape the observed community composition of the biofilms. This supports our hypothesis that species sorting has a certain role in the assembly of the biofilm community. Previous work showed that sorting, as induced by fine-scale hydrodynamic niche differentiation, rather than mass effects, was a potential mechanism of stream biofilm community assembly (Besemer *et al.*, 2009).

The interplay of niche availability and competition has been suggested to drive the patterns of bacterial biodiversity in biofilms (Jackson *et al.*, 2001) and may induce species sorting. Our results suggest that species sorting resulted in different relative abundances of dominant taxa and in the

presence/absence of rare taxa, rather than the complete replacement of the dominant groups. The observation that the most abundant genera occurred in both suspended and biofilm communities is surprising, given the clear separation of the two groups in the nMDS analysis. This apparent contradiction can be partly explained by our finding that variance in the dominant genera *Acidovorax spp.*, *Flavobacterium spp.* and *Polynucleobacter spp.* alone yielded similar community composition patterns as derived from the complete community. This indicates that community diversification below the genus level contributes to the observed separation of the suspended and biofilm communities.

Bulk and active populations, although clearly different from each other, generated comparable patterns of community composition among the biofilms and the suspended communities. This suggests that similar mechanisms control the assembly of these populations. Although the most abundant genera were active in both the biofilm and suspended communities, others showed opposing patterns. For instance, *Proteobacteria* occurred in the active fraction either in similar or in higher percentages than in the bulk community, whereas members of the *Bacteroidetes* phylum and *Actinobacteria* contributed less to the active community. Similar differences in the distribution of active taxa have been reported from lakes (Jones and Lennon, 2010). We found no evidence that the apparently active populations in the suspended community contributed more to biofilm formation than the less active populations. For instance, *Alphaproteobacteria*, which were active in the suspended community, were less abundant in the biofilms than in the suspended community.

To initiate biofilm formation, bacteria need to be able to attach to surfaces or to co-aggregate (Rickard *et al.*, 2003, 2004). This ability might have favored the proliferation of certain groups of *Betaproteobacteria*, which were found to dominate biofilm communities in this as well as in earlier studies (Schweitzer *et al.*, 2001; Araya *et al.*, 2003). Interestingly, the most abundant genus *Acidovorax*

(*Betaproteobacteria*) consisted of only one OTU at a 97% sequence similarity level. This suggests that *Acidovorax* may be a highly competitive generalist, potentially involved in early biofilm formation in these streams. Bacteria related to Acidovorax have been found to be among the first colonizers of diatom microaggregates (Knoll *et al.*, 2001). Furthermore, *Bacilli* and the *Gammaproteobacteria*, preferentially found in the active biofilm communities, contain well-known biofilm-forming species, such as *Bacillus subtilis, Pseudomonas aeruginosa, Vibrio cholerae, Escherichia coli* (Hall-Stoodley *et al.*, 2004; Branda *et al.*, 2005). Members of these groups have also been shown to auto- and co-aggregate (Rickard *et al.*, 2003).

Members of the Bacteroidetes phylum (*Flavobacteria* and *Sphingobacteria*) occurred predominantly in the bulk, although not in the active, biofilm communities. This might indicate comparably low activity of these groups, resulting from more favorable growth conditions for these bacteria during early biofilm formation. Particularly *Flavobacteria* are known to degrade biopolymers, such as cellulose, from dead plant material (Kirchman, 2002) as it is often flushed into streams during the onset of the snowmelt.

454 pyrosequencing and T-RFLP analysis generated comparable patterns of community composition, indicating that a fingerprinting method targeting the most abundant OTUs may generate reliable patterns of community composition. High-throughput sequencing methods are, however, imperative to obtain reliable estimates of bacterial diversity. T-RFLP analysis failed to reveal clear diversity patterns, and those patterns inferred from 454 pyrosequencing data vanished when an artificial threshold, mimicking a typical T-RFLP resolution, was applied. The number of OTUs detected by a low-resolution method, such as T-RFLP, may depend on the rank-abundance curve rather than on the actual richness of the community (Bent and Forney, 2008), and it has been argued that such methods do not provide reliable depiction of diversity patterns (Blackwood *et al.*, 2007).

In summary, our findings indicate that species sorting is an important mechanism involved in the assembly of benthic biofilm communities from the source community in the stream water. Our results also suggest that putatively active and inactive populations contributed comparably to the observed patterns of community composition in both the biofilms and their suspended counterparts.

References

Ager D, Evans S, Li H, Lilley AK, Van Der Gast CJ. (2010). Anthropogenic disturbance affects the structure of bacterial communities. *Environ Microbiol* **12**: 670–678.

Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PloS ONE* **3**: e2836.

Andersson AF, Riemann L, Bertilsson S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.

Araya R, Tani K, Takagi T, Yamaguchi N, Nasu M. (2003). Bacterial activity and community composition in stream water and biofilm from an urban river determined by fluorescent *in situ* hybridization and DGGE analysis. *FEMS Microbiol Ecol* **43**: 111–119.

Battin TJ, Kaplan LA, Findlay S, Hopkinson CS, Marti E, Packman AI *et al.* (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nat Geosci* **1**: 95–100.

Battin TJ, Sloan WT, Kjelleberg S, Daims H, Head IM, Curtis TP *et al.* (2007). Microbial landscapes: new paths to biofilm research. *Nat Rev Microbiol* **5**: 76–81.

Bent SJ, Forney LJ. (2008). The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J* **2**: 689–695.

Besemer K, Singer G, Hödl I, Battin TJ. (2009). Bacterial community composition of stream biofilms in spatially variable-flow environments. *Appl Environ Microbiol* **75**: 7189–7195.

Besemer K, Singer G, Limberger R, Chlup AK, Hochedlinger G, Hödl I *et al.* (2007). Biophysical controls on community succession in stream biofilms. *Appl Environ Microbiol* **73**: 4966–4974.

Blackwood CB, Hudleston D, Zak DR, Buyer JS. (2007). Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. *Appl Environ Microbiol* **73**: 5276–5283.

Branda SS, Vik A, Friedman L, Kolter R. (2005). Biofilms: the matrix revisited. *Trends Microbiol* **13**: 20–26.

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108**: 12776–12781.

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM. (1995). Microbial biofilms. *Annu Rev Microbiol* **49**: 711–745.

Elshahed MS, Najar FZ, Aycock M, Qu C, Roe BA, Krumholz LR. (2005). Metagenomic analysis of the microbial community at Zodletone Spring (Oklahoma): insights into the genome of a member of the novel candidate division OD1. *Appl Environ Microbiol* **71**: 7598–7602.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106**: 22427–22432.

Hall-Stoodley L, Costerton JW, Stoodley P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2**: 95–108.

Hammer Ø, Harper DAT, Ryan PD. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electronica* **4**: 9.

Harris JK, Kelley ST, Pace NR. (2004). New perspective on uncultured bacterial phylogenetic division OP11. *Appl Environ Microbiol* **70**: 845–849.

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

Holyoak M, Leibold MA, Holt RD. (2005). *Metacommunities: Spatial Dynamics and Ecological Communities*. University of Chicago Press: Chicago.

Jackson CR, Churchill PF, Roden EE. (2001). Successional changes in bacterial assemblage structure during epilithic biofilm development. *Ecology* **82**: 555–566.

Jones SE, Lennon JT. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* **107**: 5881–5886.

Jost L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427–2439.

Kirchman DL. (2002). The ecology of Cytophaga–Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.

Kirchman DL, Cottrell MT, Lovejoy C. (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol* **12**: 1132–1143.

Knoll S, Zwisler W, Simon M. (2001). Bacterial colonization of early stages of limnetic diatom microaggregates. *Aquat Microb Ecol* **25**: 141–150.

Leibold M, Holyoak M, Mouquet N, Amarasekare P, Chase J, Hoopes M *et al.* (2004). The metacommunity concept: a framework for multi scale community ecology. *Ecol Lett* **7**: 601–613.

Logue JB, Lindström ES. (2010). Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *ISME J* **4**: 729–738.

Lunn M, Sloan WT, Curtis TP. (2004). Estimating bacterial diversity from clone libraries with flat rank abundance distributions. *Environ Microbiol* **6**: 1081–1085.

Lyautey E, Jackson CR, Cayrou J, Rols JL, Garabétian F. (2005). Bacterial community succession in natural river biofilm assemblages. *Microb Ecol* **50**: 589–601.

Mantel N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.

Peter H, Ylla I, Gudasz C, Romaní AM, Sabater S, Tranvik LJ. (2011). Multifunctionality and diversity in bacterial biofilms. *PloS ONE* **6**: e23225.

Pommier T, Neal PR, Gasol JM, Coll M, Acinas SG, Pedrós-Alió C. (2010). Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat Microb Ecol* **61**: 221–233.

Quince C, Curtis TP, Sloan WT. (2008). The rational exploration of microbial diversity. *ISME J* **2**: 997–1006.

Quince C, Lanzen A, Davenport R, Turnbaugh P. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.

Rickard A, McBain A, Ledder R, Handley P, Gilbert P. (2003). Coaggregation between freshwater bacteria within biofilm and planktonic communities. *FEMS Microbiol Lett* **220**: 133–140.

Rickard AH, McBain AJ, Stead AT, Gilbert P. (2004). Shear rate moderates community diversity in freshwater biofilms. *Appl Environ Microbiol* **70**: 7426–7435.

Schwalbach MS, Hewson I, Fuhrman JA. (2004). Viral effects on bacterial community composition in marine plankton microcosms. *Aquat Microb Ecol* **34**: 117–127.

Schweitzer B, Huber I, Amann R, Ludwig W, Simon M. (2001). Alpha- and Betaproteobacteria control the consumption and release of amino acids on lake snow aggregates. *Appl Environ Microbiol* **67**: 632–645.

Sichel H. (1974). On a distribution representing sentence-length in written prose. *J R Stat Soc Ser A* **137**: 25–34.

Singer G, Besemer K, Schmitt-Kopplin P, Hödl I, Battin TJ. (2010). Physical heterogeneity increases biofilm resource use and its molecular diversity in stream mesocosms. *PloS ONE* **5**: e9988.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. (2002). Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol* **64**: 583–639.

Spring S, Schulze R, Overmann J, Schleifer KH. (2000). Identification and characterization of ecologically significant prokaryotes in the sediment of freshwater lakes: molecular and cultivation studies. *FEMS Microbiol Rev* **24**: 573–590.

Tamames J, Abellán J, Pignatelli M, Camacho A, Moya A. (2010). Environmental distribution of prokaryotic taxa. *BMC Microbiol* **10**: 85.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, AT.

Vishnivetskaya TA, Mosher JJ, Palumbo AV, Yang ZK, Podar M, Brown SD *et al.* (2011). Mercury and other heavy metals influence bacterial community structure in contaminated Tennessee streams. *Appl Environ Microbiol* **77**: 302–311.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.