

ORIGINAL RESEARCH

Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river

Geneviève Bricheux^{1,3}, Loïc Morin⁴, Gwenaël Le Moal^{2,3}, Gérard Coffe^{1,3}, Damien Balestrino^{2,3}, Nicolas Charbonnel^{2,3}, Jacques Bohatier^{2,3} & Christiane Forestier^{2,3}

¹Laboratoire "Microorganismes: Génome et Environnement", Clermont Université, Université Blaise Pascal, BP 10448, F-63000, Clermont-Ferrand, France

²Laboratoire "Microorganismes: Génome et Environnement", Clermont Université, Université d'Auvergne, BP 10448, F-63000, Clermont-Ferrand, France

³CNRS, UMR 6023, LMGE, F-63177, Aubiere, France

⁴Institut de Génétique et Microbiologie, UMR, CNRS, 8621 PRES, Paris Sud, France

Keywords

Aquatic communities, biofilm, pyrosequencing.

Correspondence

Christiane Forestier, Laboratoire "Microorganismes: Génome et Environnement", Clermont Université, Université d'Auvergne, BP 10448, F-63000 Clermont-Ferrand. Tel: +33 (0) 473 177 994; Fax: +33(0) 473 178 370; E-mail: christiane.forestier@udamail.fr

Funding Information

This work was partially supported by grant AFSSSET EST-2007-6.

Received: 12 November 2012; Revised: 1 February 2013; Accepted: 13 February 2013

MicrobiologyOpen 2013; 2(3): 402–414

doi: 10.1002/mbo3.80

Abstract

Despite the recent and significant increase in the study of aquatic microbial communities, little is known about the microbial diversity of complex ecosystems such as running waters. This study investigated the biodiversity of biofilm communities formed in a river with 454 SequencingTM. This river has the particularity of integrating both organic and microbiological pollution, as receiver of agricultural pollution in its upstream catchment area and urban pollution through discharges of the wastewater treatment plant of the town of Billom. Different regions of the small subunit (SSU) ribosomal RNA gene were targeted using nine pairs of primers, either universal or specific for bacteria, eukarya, or archaea. Our aim was to characterize the widest range of rDNA sequences using different sets of polymerase chain reaction (PCR) primers. A first look at reads abundance revealed that a large majority (47–48%) were rare sequences (<5 copies). Prokaryotic phyla represented the species richness, and eukaryotic phyla accounted for a small part. Among the prokaryotic phyla, Proteobacteria (beta and alpha) predominated, followed by Bacteroidetes together with a large number of nonaffiliated bacterial sequences. Bacillariophyta plastids were abundant. The remaining bacterial phyla, Verrucomicrobia and Cyanobacteria, made up the rest of the bulk biodiversity. The most abundant eukaryotic phyla were annelid worms, followed by Diatoms, and Chlorophytes. These latter phyla attest to the abundance of plastids and the importance of photosynthetic activity for the biofilm. These findings highlight the existence and plasticity of multiple trophic levels within these complex biological systems.

Introduction

Current knowledge of the microbial communities in rivers is scarce whereas these communities are hot spots of exchanges between bacteria from human sources and aquatic ecosystems (Vilanova et al. 2004; Martins da Costa et al. 2006; Moura et al. 2007; Galvin et al. 2010) and are natural receptacles for pollutant molecules from waste water treatment plants (organic and microbiological pollution), individual industries (chemical pollution), or

agricultural activity (pesticides) (Brown and van Beinum 2009). Many studies on the impact of pollutants have been made on free microbial communities, seldom assessing the effects on the periphyton compartment. In the lotic environment, autotrophic (algae and cyanobacteria) and heterotrophic (bacteria, fungi, and protozoa) communities are mainly present as biofilms. Owing to their benthic nature, these communities are much more exposed than plankton communities. They can therefore provide more information about the effects of pollution

and are excellent models for observing the health of rivers, because of their high sensitivity to environmental changes, their short life cycle, their diversity, and their ubiquity in aquatic environments (Lavoie *et al.* 2004).

These communities cannot be adequately studied by traditional culture-based techniques. As a result, a variety of molecular fingerprinting methods have been developed to assess the diversity of microbial communities such as ribosomal spacer analysis, terminal restriction fragment length polymorphism (t-RFLP), denaturing gradient gel electrophoresis (DGGE), 16S rRNA clone libraries, or fluorescence in situ hybridization (FISH) (Nocker *et al.* 2007; Rajendhran and Gunasekaran 2011).

The results obtained with these methods are likely to be incomplete as they do not capture the whole complexity of the communities and mostly address the bacterial compartment of the biofilms. High-throughput next-generation sequencing, such as 454 pyrosequencing (Margulies *et al.* 2005) of the small subunit (SSU) rRNA genes, offers an alternative, in which detailed community structure can be achieved in combination with fairly high taxonomic resolution (Sogin *et al.* 2006; Andersson *et al.* 2008). This approach has so far been used to explore human microbiomes (Costello *et al.* 2009; Turnbaugh *et al.* 2009; Koren *et al.* 2010; Lazarevic *et al.* 2010; Ling *et al.* 2010) and environmental samples such as soil (Roesch *et al.* 2007; Lauber *et al.* 2009) and oceans (Sogin *et al.* 2006; Qian *et al.* 2010). A few investigations have been performed on environmental water samples from either waste or drinking water (Hong *et al.* 2010; Lee *et al.* 2010; Liu *et al.* 2012). Studies on biofilms have demonstrated the preponderance of bacteria affiliated to the subclass of beta-proteobacteria and Cytophaga-Flavobacterium in bacterial communities in anthropized environments characterized by high pollution (organic and metallic) (Brummer *et al.* 2000, 2003). Recently, Besemer demonstrated the predominance of beta-proteobacteria in biofilms and suspended communities in water streams, as well as an overall lower diversity in the biofilm communities compared with those in suspended stream water (Besemer *et al.* 2012). Eukaryotes are known to be integrated inside natural aquatic biofilms, but available biodiversity data are poorly documented (Dorigo *et al.* 2007, 2010; Fechner *et al.* 2010).

In this study, we used extensive 454 pyrosequencing analysis of different regions of SSU rRNA genes of biofilm communities formed in the flow of a small river receiving urban effluent from a wastewater treatment plant. To characterize the largest range of rDNA sequences, we used a set of nine couples of primers hybridizing to different regions of the SSU rRNA genes. Eukaryotes being closely associated with bacteria in these communities, the eukaryotic population was also investigated.

Materials and Methods

Biofilm sampling

Biofilms were formed on several glass substrates, which were placed nearby within the stream of the Jauron River, and incubated for 19 days in April 2009. Jauron is a small river located downstream of the wastewater effluent discharge of Billom, a town of about 5000 inhabitants, 25 kms south of Clermont-Ferrand, France (GPS location: 45°44'19.36"N, 3°20'20.903E, 340 m).

Microscopic observations

Biofilms were rinsed with cacodylate buffer, fixed for 60 min with a solution of 4% glutaraldehyde with 0.15% ruthenium red (in 0.2 mol/L cacodylate buffer), rinsed for 5–10 min in 0.2 mol/L cacodylate buffer, and then post-fixed in 1% OsO₄ (in 0.1 mol/L cacodylate buffer) for 1 h. They were then transferred directly to cold 50% ethanol and dehydrated through a series of increasing concentrations of cold ethanol. After drying, the samples were mounted on stubs and coated with gold in a sputter coater and observed with a Jeol 6060-low vacuum microscope.

DNA extraction, PCRs, and pyrosequencing

Biofilm samples were collected from the surface of several glass substrate slides scattered across the sampling place. The biofilms formed at the surface of the slides after 19 days of incubation were mixed and resuspended in mineral water by vortexing and sonication. The resulting biofilm suspension was then centrifuged to collect the pellets. The DNA was extracted according to a chemical and enzymatic DNA extraction protocol using the Macherey-Nagel Nucleospin tissue kit after mechanical disruption of the cells by a Precellys[®] apparatus (Bertin Technologies, Montigny le Bretonneux, France). The bulk community DNA sequences were amplified with nine pairs of primers (Table 1) covering the different regions of the SSU ribosomal genes: two pairs of universal primers were used for both eukaryotes and prokaryotes (U1 and U2), one pair was used to detect archaea (A), four pairs were specific for prokaryotes (B1, B2, B3, and B4), and two pairs were specific for eukaryotes (Euc1 and Euc2). Primers were chosen according to previous analyses showing that biodiversity richness is linked to the primer sequences (Medlin *et al.* 1988; Reysenbach *et al.* 1994; Dassarma and Fleischmann 1995; Vetriani *et al.* 1999; Baker *et al.* 2003; Huse *et al.* 2008). Only one primer was adapted for this study (343FB) (Hansen *et al.* 1998).

Forward primers contained the Roche 454 pyrosequencing adapter FLX A plus MID (Multiplex identifier) and

Table 1. Sequences of primers used for pyrosequencing.

Primer name	Sequence (5'–3')	Target	References
1053F	GCATGGCYGYCGTCAG	Universal SSU rRNA gene (U1)	Dassarma and Fleischmann 1995; Reysenbach <i>et al.</i> 1994;
1510R	GGTTACCTTGTACGACTT		
1053F	GCATGGCYGYCGTCAG	Universal SSU rRNA gene (U2)	Dassarma and Fleischmann 1995; Reysenbach <i>et al.</i> 1994;
1391R	GACGGGCGGTGWGTRCA		
1098FA	GGCAACGAGCGMGACCC	Archaeal SSU rRNA gene (A)	Reysenbach <i>et al.</i> 1994; Reysenbach <i>et al.</i> 1994;
1510R	GGTTACCTTGTACGACTT		
528FE	CGGTAATCCAGCTCC	Eukaryotic SSU rRNA gene (Euc1)	Medlin <i>et al.</i> 1988; Medlin <i>et al.</i> 1988;
1193RE	GGGCATMACDGACCTGTT		
7FE	ACCTGGTTGATCCTGCCAG	Eukaryotic SSU rRNA gene (Euc2)	Vetriani <i>et al.</i> 1999; Dassarma and Fleischmann 1995;
529RE	ACCGCGGCKGCTGGC		
8FB	AGAGTTTGATCCTGGCTCAG	Bacterial SSU rRNA gene (B1)	Reysenbach <i>et al.</i> 1994; Dassarma and Fleischmann 1995;
534RB	ATTACCGGGCTGCTGGC		
517FB	GCCAGCAGCCGCGGTAA	Bacterial SSU rRNA gene (B2)	Reysenbach <i>et al.</i> 1994; Huse <i>et al.</i> 2008;
1046RB	CGACAGCCATGCANACCT		
343FB	TACGGRAGGCAGCAG	Bacterial SSU rRNA gene (B3)	This study Huse <i>et al.</i> 2008;
1046RB	CGACAGCCATGCANACCT		
1099FB	GYAACGAGCGCAACCC	Bacterial SSU rRNA gene (B4)	Reysenbach <i>et al.</i> 1994; Reysenbach <i>et al.</i> 1994
1510R	GGTTACCTTGTACGACTT		

the specific designed RNA sequence, while the reverse primers contained the FLX B adapter plus the same MID plus the designed RNA sequence. Polymerase chain reactions (PCRs) were performed with 50 ng of template DNA and 20 pmol of each forward and reverse primer, 100 μ L of mix supplemented with 1.25 U of Phusion High-Fidelity DNA Polymerase Finnzyme (Thermo Scientific Fisher, Villebon-sur-Yvette, France), using a GeneAmp PCR system 2700 thermocycler (Applied Biosystems[®], Life Technologies, Villebon-sur-Yvette, France). The two-step PCR conditions were as follows: 95°C for 1 min, 30 cycles of denaturation (95°C; 10 sec), annealing (60°C; 30 sec), and extension (72°C; 30 sec), followed by the final elongation (72°C; 10 min). The DNA was then quantified using a DNA quantitation kit fluorescence assay (Sigma-Aldrich, St Quentin Fallavier, France) and DNAs from the different reactions were mixed in equimolar proportions. Amplicons were then subjected to pyrosequencing using a Genome Sequencer 454 Titanium GS FLX (Eurofins, MWG/Operon, Germany).

Pyrosequencing data analysis

PyroSuite

This program (available upon request) was developed to combine algorithms from several packages in order to (i) compute indices and rarefaction curves to determine whether the data set is representative of the sample biodiversity (<http://www.mothur.org/>); (ii) compare the sampled sequences against a local dedicated bank for automated identification (Blastn <http://blast.ncbi.nlm.nih.gov> or Blat <http://genome.ucsc.edu>); and (iii) study their

clusterization in terms of relative abundance and annotated phylogenetic trees (Phylip (<http://www.phylip.com/>)).

PyroSuite includes several subroutines, of which the following are the two most important.

PyroSeqToAlignedSeq

This handles the Mothur Software (Schloss 2009), reading the commercial sequence fasta file, selecting a valid subset of the sequence stretches and producing an alignment fasta file through several algorithms described in the “Costello stool analysis” (Costello *et al.* 2009). Reads from commercial files were shortened when more than two undetermined bases (N) were found and Phred quality scores were $Q \leq 30$. Sequences with more than eight repeats or smaller than 100 bp were eliminated, as well as primer sequences with mismatches. Sequence boundaries were chosen and SILVA (Pruesse *et al.* 2007) was used as the reference alignment in the process according to the analyzed amplicons data set (Bacteria, Archae, Eukarya, or Universal). The subroutine also computes rarefaction curves.

IdentifySeq

This subroutine uses different information sources: taxonomic files (GenBank), a dedicated annotated databank fasta file extracted from the updated archive of Genbank, EMBL, and DDBJ, and a subset of the alignment fasta file produced by PyroSeqToAlignedSeq. The sequences of this last file were processed with BLAST or BLAT against the annotated bank for identification, producing two files containing the BLAST result and ID of each sequence.

Also produced are two files, which can identify and calculate the relative importance of redundant sequences.

Results

Microscopy analysis

Scanning electron microscopy observations performed with 19-day-old biofilm samples revealed heavily coated surfaces with a high number of diatoms at the interface between the biofilm and free running water (Fig. 1A). Rod-shaped bacteria were visible on the surface of diatoms (Fig. 1B) but most of these microorganisms were embedded in heavy extracellular material.

454 pyrosequencing strategy

The 454 pyrosequencing of biofilm DNA extract produced 26,423 raw reads. After the removal of 3994 low-quality

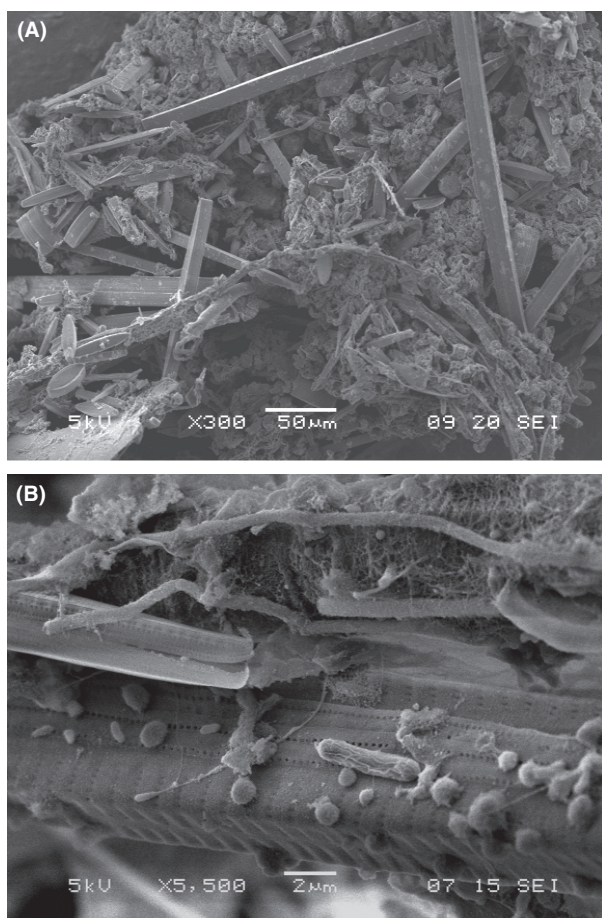


Figure 1. Scanning electron microscopy (SEM) images representative of biofilms grown for 19 days in riverine water on glass slides. (A) original magnification 300 \times , Bar = 50 μ m, (B) original magnification 5500 \times , Bar = 2 μ m.

reads (15.1%), 21,998 reads with a read length of 400 ± 50 bp (mean \pm standard deviation) were used for further analyses.

Rarefaction curves were produced for each pair of primers, with an operational taxonomic units (OTU) 3% species cutoff (Fig. 2). Two sets clearly appeared. In the first, the curves tended to be asymptotic and included U1, U2, and B4 primer pairs, while in the second OTU numbers increased with the sequence numbers (B1, B2, B3, Euc1, and Euc2 primer pairs), indicating that the sequencing effort was not sufficient in these cases. The second set was correlated with the weak number of sequences for B3, Euc1, and Euc2 primer pairs. The A primer pair was designed to recover archaeal sequences in the sample; sequence analysis of the resulting amplicons revealed the presence of bacterial sequences, to the exclusion of any archaeal sequence, indicating that a PCR mismatch annealing led to bacterial amplification for the forward archaeal specific primer. Further PCRs with different specific archaeal primers failed to recover any archaeal sequence (data not shown), thus confirming this result. The B3, Euc1, and Euc2 primer pairs were designed to recover about 700-bp-long amplicons. This length is the upper limit of the Titanium pyrosequencing technology, and may explain why the output of sequences was highly reduced. Likewise, as we made an equimolar mix of each ampli-

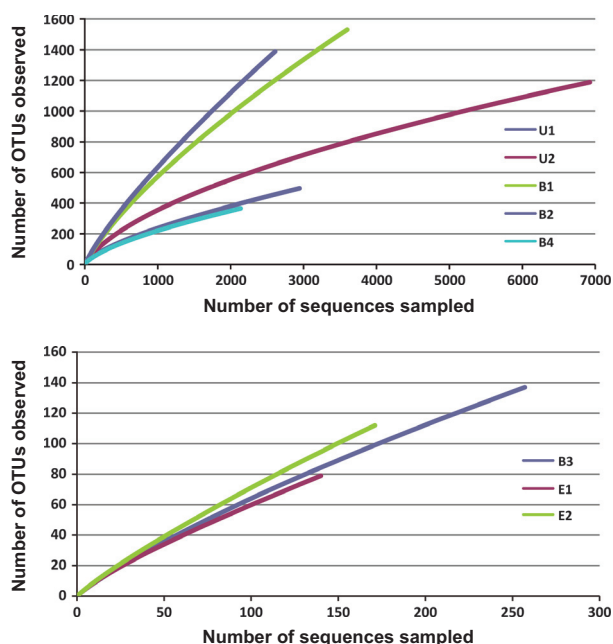


Figure 2. Rarefaction analysis of the amplicons obtained with the different pairs of primers at a level of 97% 16S rRNA similarity.

con, we can predicate that longer sequences were less numerous in the final mix.

Diversity indices, calculated with Mothur (Schloss 2009) are shown in Table 2. Evenness index was the highest and Coverage estimation the lowest for the B3 pair of primers, in correlation with the low sequence number and the growing rarefaction curves mentioned above. However, Euc1 and Euc2 pairs of primers displayed a medium evenness index and a good Coverage estimation, despite the low number of sequences recovered.

Number of amplicons obtained according to the set of primers

Figure 3 displays the diversity of sequences at the class level in relation to the primer pair that was used. The two pairs of universal primers (U1 and U2) were localized at the end of the SSU ribosomal gene, involving the V7 and V8 domains plus the V9 domain for U1. The U2 primer pair (V7–V8) gave rise to about three times as many amplicons as the U1 primer pair (V7–V9). A proportion

Table 2. Diversity and evenness indices of the SSU rDNA gene sequences from the pyrosequencing analysis according to the primers used.

Primers	NS	Sobs (0.03)	Chao1	Shannoneven	Shannon	Coverage
U1(V7–V9)	3058	284	642 (512;847)	0.63	3.58 (3.52;3.65)	0.95
U2(V7–V8)	7099	993	2163 (1917;2475)	0.71	4.91 (4.87;4.96)	0.92
Euc1(V4–V6)	192	37	125 (67;295)	0.60	2.16 (1.89;2.43)	0.84
Euc2(V1–V3)	258	71	184 (119;338)	0.78	3.31 (3.16;3.46)	0.87
B1(V1–V3)	5298	1523	3718 (3340;4175)	0.79	5.79 (5.74;5.85)	0.85
B2(V4–V6)	3287	846	2019 (1755;2361)	0.78	5.25 (5.19;5.31)	0.87
B3(V3–V6)	454	137	619 (389;1058)	0.86	4.22 (4.02;4.41)	0.58
B4(V7–V9)	2352	264	552 (449;713)	0.65	3.64 (3.56;3.73)	0.92
Total	21,998					

NS, number of sequences.

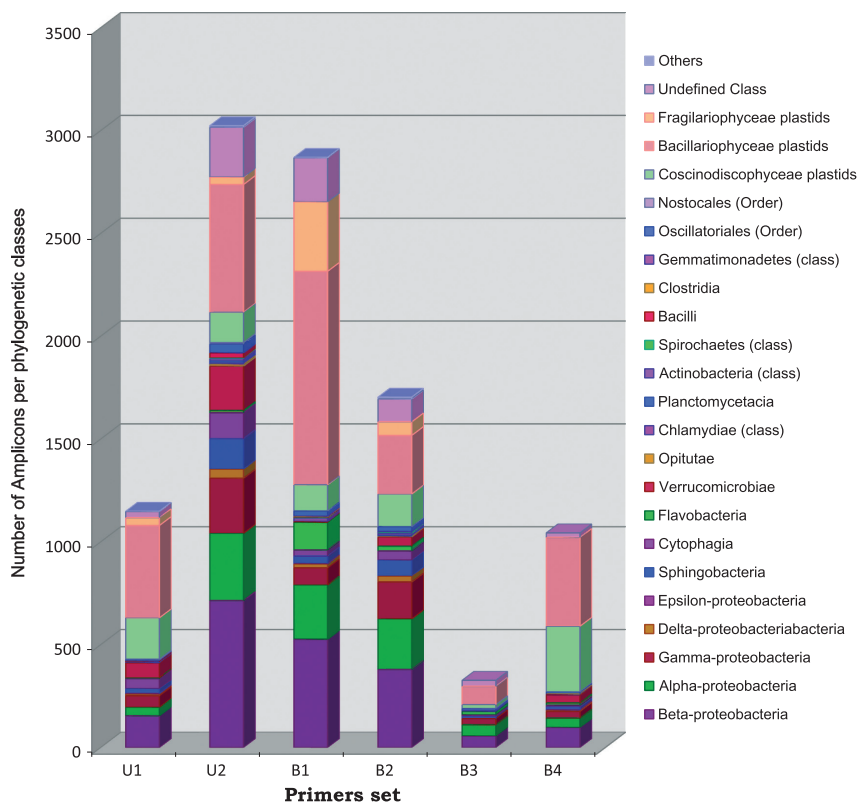


Figure 3. Taxonomic distribution of sequences assigned to prokarya according to primer sets.

of 12.4% of the reads were classified as unidentified sequences for the U2 primer pair, and 5.7% for the U1 primer pair. Universal primers specifically recognized some taxa, such as Chlamydiae, Chloroflexi, and Chlorophycota plastids, but no eukaryotic sequence was detected. The four pairs of primers specific for bacteria (B1, B2, B3, and B4) recognized different variable domains of the molecule from V1 to V9. B1 primer pair (V1–V3) produced about twice as many amplicons as the B2 primer pair (V4–V6), about three times as many amplicons as the B4 primer pair (V7–V9), and about ten times as many amplicons as B3 primer pair (V3–V6). More than half of the B1 amplicons corresponded to the chloroplast sequences of diatoms. The B1 primer pair also produced more sequences than the universal primers among the Bacillariophyta chloroplasts. However, phyla such as Nitrospirae were recognized only by the B4 primer pair, while *Deinococcus-Thermus* was recognized only by the B2 primer pair. Primer pairs B3 and B4

revealed a weak diversity, dominated by chloroplastic sequences for primer pair B4. The B1 primer pair produced 11% unidentified sequences, as against 10.2%, 17%, and 4.5%, respectively, for primers B2, B3, and B4. The presence of eukaryotic sequences was confirmed by the simultaneous use of two eukaryote specific primers (Euc1 and Euc2). No member of the domain Archaea was detected, although 144 unique sequences were initially retrieved using primer pairs designed to detect archaea.

Composition of the biofilm communities

After alignment of the 21,998 reads, 13,804 unique sequences derived from all pairs of primers were retained for OTU clustering and a final number of 4264 OTUs was obtained, on the basis of a 3% distance interval for species grouping (Table S1). These unique sequences were then grouped according to different taxonomic levels within

Table 3. Table summarizing the distribution of each phylum according to the pair primers used for amplification.

Phylum/primers	U1 (V7–V9)	U2 (V7–V8)	A (V7–V9)	B1 (V1–V3)	B2 (V4–V6)	B3 (V3–V6)	B4 (V7–V9)	Euc1	Euc2
Proteobacteria	267-23,24	1367-45,21	18-12,5	900-31,36	839-49,21	142-41,04	184-17,07		
Bacteroidetes	81-7,05	331-10,95		210-7,32	161-9,44	37-10,69	39-3,62		
Verrucomicrobia	77-6,70	247-8,17	1-0,69	5-0,17	54-3,17		38-3,53		
Cyanobacteria	9-0,78	50-1,65	2-1,39	26-0,91	24-1,41	10-2,89	10-0,93		
Firmicutes	5-0,44	24-0,79		6-0,21			4-0,37		
Actinobacteria	3-0,26	11-0,36		16-0,56	9-0,53	1-0,29	2-0,19		
Planctomycetes	1-0,09	10-0,33			6-0,35				
Chlamydiae		5-0,17							
Nitrospirae		2-0,07			2-0,12				
Spirochetes		4-0,13		2-0,07	2-0,12				
Acidobacteria					2-0,12	1-0,29			
Gemmatimonadetes		1-0,03		1-0,03	7-0,41	3-0,87			
Chloroflexi		2-0,07							
<i>Deinococcus-Thermus</i>					3-0,18				
Bacillariophyta plastids	675-58,75	793-26,22	123-85,42	1504-52,40	521-30,56	131-37,86	785-72,82		
Bacillariophyta								3-3,30	160-71,75
Euglenida plastids		5-0,17		1-0,03	3-0,18				
Chlorophycota plastids		3-0,10							
Chlorophycota								7-7,69	6-2,69
Eustigmatophyceae plastids					3-0,18				
Streptophyta plastids	1-0,09	3-0,10		2-0,07					
Annelida								69-75,82	49-21,97
Arthropoda								1-1,10	
Chytridiomycota								2-2,20	
Ascomycota								1-1,10	
Nemata								4-4,40	
Mollusca									10-0,45
Rotifera									1-0,45
Undefined Phylum	30-2,61	166-5,49		197-6,86	69-4,05	21-6,07	16-1,48	4-4,40	6-2,69
Total	1149	3024	144	2871	1705	346	1078	91	223

The composition of each phylum is given in term of number of different amplicons and percentage.

Table 4. Table summarizing the distribution of each class according to the pair primers used for amplification. The composition of each class is given in term of number of different amplicons with a confident score of >80%.

Classes/primers	U1(V7–V9)	U2(V7–V8)	A (V7–V9)	B1(V1–V3)	B2(V4–V6)	B3(V3–V6)	B4(V7–V9)	Euc1	Euc2
Beta-proteobacteria	155-13,49	715-23,64	8-5,56	527-18,36	380-22,29	56-16,18	98-9,09		
Alpha-proteobacteria	41-3,57	327-10,81	7-4,86	264-9,20	246-14,43	55-15,90	45-4,17		
Gamma-proteobacteria	56-4,87	268-8,86	3-2,08	83-2,89	179-10,50	30-8,67	33-3,06		
Delta-proteobacteria	9-0,78	44-1,46		19-0,66	29-1,70		5-0,46		
Epsilon-proteobacteria	2-0,17			2-0,07			3-0,28		
Sphingobacteria	24-2,09	148-4,89		36-1,25	78-4,57	12-3,47	15-1,39		
Cytophagia	48-4,18	126-4,17		30-1,05	44-2,58	7-2,02	7-0,65		
Flavobacteria	4-0,35	11-0,36		134-4,67	23-1,35	16-4,62	11-1,02		
Verrucomicrobiae	73-6,35	216-7,14	1-0,69	5-0,17	45-2,64		38-3,53		
Opiritae		10-0,33			3-0,18				
Chlamydiae (class)		5-0,17							
Planctomycetacia	1-0,09	10-0,33			6-0,35				
Actinobacteria (class)	3-0,26	11-0,36		16-0,56	9-0,53	1-0,29	2-0,19		
Spirochetes (class)		4-0,13		2-0,07	2-0,12				
Bacilli	5-0,44	24-0,79		3-0,10			3-0,26		
Clostridia				3-0,10			1-0,09		
Gemmatimonadetes (class)		1-0,03		1-0,03	7-0,41	3-0,87			
Nitrospira (class)		2-0,07			2-0,12				
Acidobacteria (class)					2-0,12				
Spartobacteria					1-0,06				
Hadobacteria					3-0,18				
Oscillatoriales (order)	7-0,61	42-1,38	2-1,39	260,91	23-1,35	10-2,89	10-0,93		
Nostocales (order)	2-0,17	8-0,27			1-0,06				
Coscinodiscophyceae plastids	201-17,49	147-4,86	49-34,03	127-4,43	159-9,33	19-5,49	319-29,59		
Bacillariophyceae plastids	449-39,08	621-20,54	62-43,06	1038-36,17	285-16,72	89-25,72	432-40,07		
Bacillariophyceae								2-2,20	148-66,37
Fragilariophyceae plastids	38-3,31	36-1,19	12-8,33	338-11,78	66-3,87	19-5,49	34-3,15		
Fragilariophyceae								1-1,10	12-5,38
Conjugatophyceae plastids	1-0,09	1-0,03		2-0,07					
Trebouxiophyceae plastids		2-0,07							
Klebsormidiophyceae plastids		1-0,03							
Ulvophyceae plastids		1-0,03							
Ulvophyceae								6-6,59	5-2,24
Florideophyceae plastids					1-0,06				
Florideophyceae								1-1,10	1-0,45
Chryomonada									1-0,45
Holotricha								1-1,01	
Phyllopharyngea								1-1,10	
Nassophorea									1-0,45
Litostomatea									1-0,45
Maxillopoda								1-1,10	
Chromadorea								2-2,20	
Chytridiomycetes								2-2,20	
Hemiascomycetes								1-1,10	
Ichthyosporea								1-1,10	
Polychaeta									1-0,45
Bivalvia									1-0,45
Monogononta									1-0,45
Clitellata								69-75,82	49-21,97
Undefined class	30-2,61	243-8,04		214-7,45	111-6,51	29-8,38	22-2,04	3-3,30	2-0,90

each set of amplicons. Tables 3 and 4 show the result at the phylum and class levels. Whatever the primer used, Bacillariophyta plastids largely predominated, followed by Proteobacteria and Bacteroidetes. Verrucomicrobia were as abundant as Bacteroidetes when detected with universal and B4 primer pairs, but much less so with the B1 and B2 primer pairs, and were not detected with the B3 primer pair. Cyanobacteria, while not as abundant as the phyla cited above, were also significantly present. The other phyla were either poorly represented, or detected with few primer pairs.

At the class level, Bacillariophyceae plastids prevail over Coscinodiscophyceae plastids, and largely over araphid pennate diatom plastids, except with the B1 primer pair for which araphid pennate diatom plastids dominate Coscinodiscophyceae plastids. Other plastid sequences were poorly represented. Among the proteobacteria, beta-proteobacteria largely predominate, followed by alpha- and gamma-proteobacteria; a few Delta-proteobacteria were also detected, but not with the B3 primer pair. Among the Bacteroidetes, Cytophagia and Sphingobacteria were the most common, while Flavobacteria were poorly abundant, except with the B1 primer pair for which they predominated over the other classes. Cyanobacteria were mainly Oscillatoriales (Phormidium). The most diversified eukaryotic phyla included Annelid worms and Stramenopiles, followed by Chlorophytes. The microalgae reflect the abundance of plastids and the importance of photosynthetic activity within biofilm. The remaining eukaryotic phyla were poorly diversified, but included a variety of protistan phyla-like ciliates, Cercozoa, Euglenozoa, and Foraminifera, but also fungi, nematodes, rotifers, arthropods, and mollusks.

Grouping sequences obtained with the U2 and B2 pairs of primers according to the abundance of each amplicon (i.e., including redundant sequences) are shown in Fig. 4A and B, respectively. Blast results were manually tested for species identifications when redundancy was significant (>5 identical amplicons). Both figures displayed a similar overall composition, with a very high number of unique sequences (47–48%), and a high proportion of Bacillariophyta plastid sequences (30–35%). Bacillariophyta plastid sequences included the three major diatom classes: a large majority (29%) of raphid pennate diatoms (Bacillariophyceae), few nonraphid pennate diatoms (Fragilariophyceae), and centric diatoms (Coscinodiscophyceae). Less abundant taxa included those that were detected with both U2 and B2 sets of primers, such as the filamentous phototrophic cyanobacteria Phormidium, and the genera Rhodobacter and Rhodoferrax, for which most species are phototrophic. In contrast, Leptothrix, sheathed bacteria that are able to oxidize Fe (II) and Mn (II), and Verrucomicrobium

were detected with the sole B2 primers while Arenimonas was obtained with the U2 primers.

The analysis of eukaryotic sequences obtained from the Euc2 primers showed a predominance of diatoms (66%), followed by annelids (23%). The remaining sequences were unique. Annelids predominated in data obtained from Euc1 primers (75%) (Table 3).

An overall analysis combining the information given by the different primer pairs was not possible because no single region of rDNA can differentiate among all bacteria or eukaryotic cells (Liu et al. 2008; Kim et al. 2011b).

Discussion

Biofilms are spatially structured communities of microbes with a complex web of symbiotic interactions. Different studies pointed out a larger heterogeneity at the local scale (due to local aggregates, availability of nutrients, etc.) than at the global scale, for which larger samples are homogenized. In water, microbial populations are influenced by fluids circulation which may impact the local populations drastically, giving rise to a somewhat similar diversity/structure on the large and global scales and a much more divergent picture at the local scale. This study presents the analysis of a biofilm community formed on several glass slides placed into a single point of a river that was initially investigated by DGGE (Bricheux et al. 2013). Massive bar-coded pyrosequencing was applied to perform a comprehensive analysis of the composition of the bacterial and eukaryotic communities of biofilms from this anthropized riverine water. This technology is a powerful method for investigating the diversity of complex ecosystems compared to conventional cloning analyses whereas not exempted of some bias or limitations (Tedersoo et al. 2010; Pinto and Raskin 2012). As every recent technologies, improvement will be done, notably with the arrival of ~3000-base single-molecule reads of Pacific Biosciences (Eid et al. 2009; Ribeiro et al. 2012). To detect most of the organisms in this ecosystem, we used several primer pairs, either universal or specific for prokaryotic, eukaryotic, and archaeal cells, which targeted the ribosomal DNA gene region, owing to the richness of the corresponding databases. The number of sequences recovered by each primer pair was highly variable and illustrates the well-known amplification bias introduced by primer selection (Polz and Cavanaugh 1998; Baker et al. 2003; Baker and Cowan 2004; Acinas et al. 2005; Rajendhran and Gunasekaran 2011). Our results corroborate previous findings showing that bacterial V1–V3 domains better illustrate species richness and diversity (see Table 2 Chao1 and Shannon estimates) (Youssef et al. 2009; Kim et al. 2011a), and that a shorter amplicon produces greater richness (for example, U1 vs. U2) (Engelbrektsen et al. 2010).

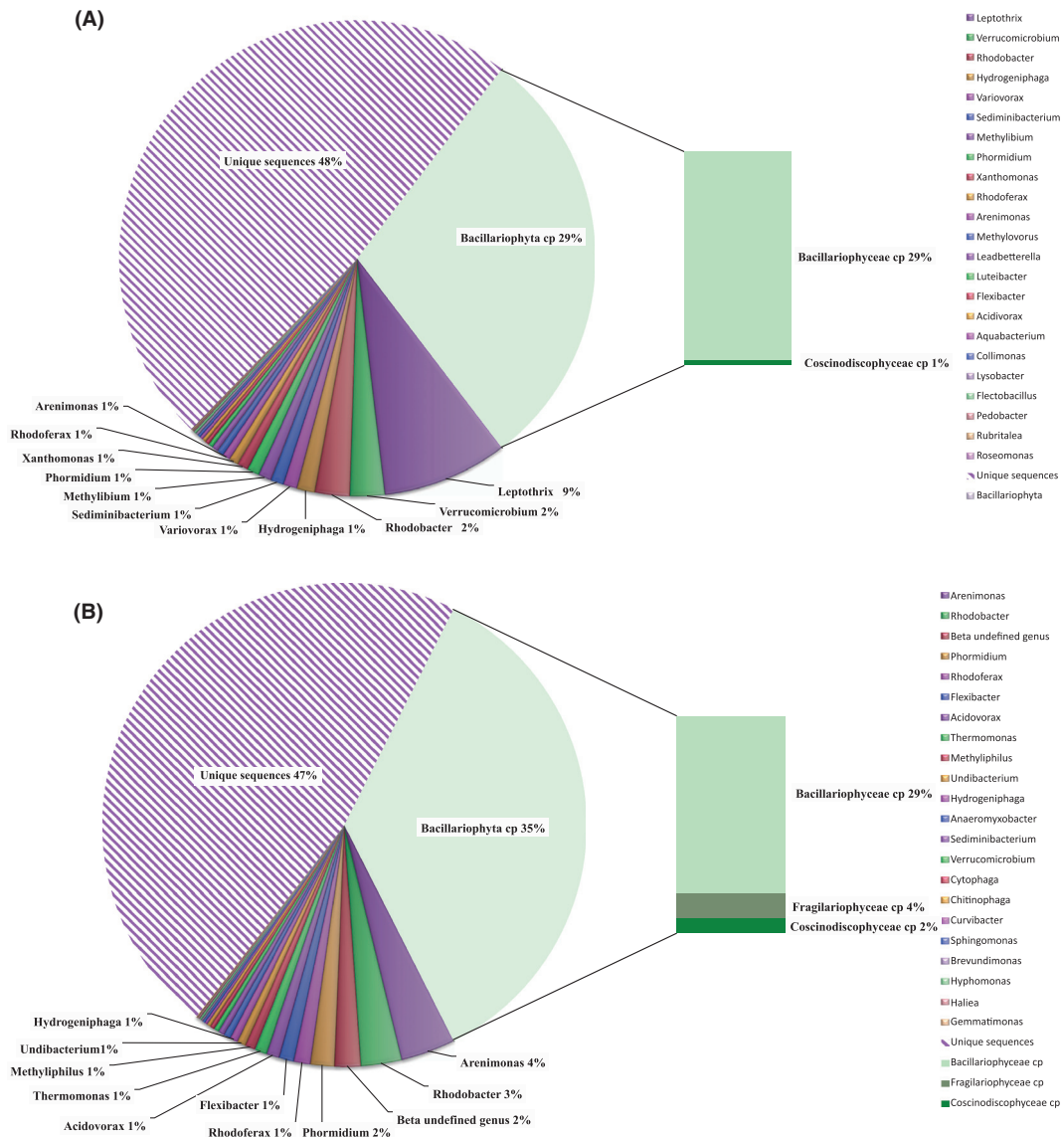


Figure 4. Taxonomic distribution and abundance observed with U2 and B2 primers (A and B, respectively). The composition and percentage of Bacillariophyta are shown in detail. Only results $\geq 1\%$ are given.

Bacterial and universal primer pairs gave rise to the highest number of sequences (6786–7567). In contrast, few amplicons were obtained with eukaryotic specific primers Euc1 and Euc2 (268–422, see Table 2), which could explain why no eukaryotic sequence was detected with universal primers. However, the Euc1 and Euc2 pairs of primers displayed a medium Shannon evenness index and a good Coverage estimation, despite the low number of sequences retrieved. It is likely therefore that the low sequence number reflects eukaryotic abundance in the sample and is not only related to the technical limitation due to the length of the amplicons.

Altogether, 4264 OTUs were detected, demonstrating the considerable diversity of anthropized riverine water biofilms, as observed in reports on other complex environments (Sogin et al. 2006; Huber et al. 2007; Roesch et al. 2007; Rusch et al. 2007; Andersson et al. 2009). Our results indicated a good diversity among the prokaryotes with a predominance (if diatom chloroplastic reads are excluded) of Proteobacteria (beta- and alpha-) and with Bacteroidetes and Verrucomicrobia as subdominant phyla. The same culture-independent technology has shown similar distributions in biofilm communities of water distribution systems, albeit with a much lower total number of OTUs (133 and

208) (Hong *et al.* 2010), and in wastewater treatment plant samples (Sanapareddy *et al.* 2009). Proteobacteria were also predominant in samples from an anoxic basin of a wastewater treatment plant (Chouari *et al.* 2010), from a drinking water treatment plant (Kwon *et al.* 2011), and from freshwater collected in a river (Vaz-Moreira *et al.* 2011). Comparison of these data with those of previous studies on river biofilms using different methods is awkward because of the different levels of resolution. However, using FISH and confocal laser scanning microscopy, Manz *et al.* (1999) showed that lotic biofilms were mostly composed in their initial formation steps of beta-Proteobacteria, followed by an increase in both alpha-proteobacteria and bacteria affiliated to the Cytophaga-Flavobacterium group. The predominance of beta-Proteobacteria and of the Cytophaga-Flavobacterium (former name of Bacteroidetes) cluster was also observed in biofilms formed in urban rivers as shown by FISH and DDGE analysis of 16S SSU ribosomal gene fragments (Brummer *et al.* 2000; Araya *et al.* 2003). Thus, the prevalence of Proteobacteria in anthropized riverine water biofilms could simply be the result of their predominance or could be related to the greater ability of these bacteria to initiate biofilm formation by adhering to surfaces. It is noteworthy that the most abundant bacterial groups detected by pyrosequencing in soil samples from several sites across the western hemisphere were also Proteobacteria and Bacteroidetes (Roesch *et al.* 2007). Interestingly, in biofilms of Swedish freshwater, beta-proteobacteria represent more than one third of the reads, followed by Flavobacteria (Besemer *et al.* 2012) whereas in our biofilm, the beta-proteobacteria were followed by the alpha-proteobacteria. Although we did not determine the bacterial composition of soil around our aquatic sampling sites, we can postulate that these wastewater effluent biofilms receive microorganisms both from the wastewater treatment plant (activated sludge) effluent and from the soil environment by water shedding. The presence in our samples of Rhodospirillum rubrum, which are able to degrade phenoxyherbicides (Ehrig *et al.* 1997), suggests that these communities could have been previously in contact with such pollutants. Potential human pathogens were also identified in all samples by culture (data not shown), mostly Enterobacteriaceae (*Klebsiella* and *Enterobacter*), *Aeromonas*, and *Pseudomonas*. Most of these isolates were resistant to antibiotics, with 75% of them producing cephalosporinases and 48% showing resistance to quinolones (data not shown). It is likely that sewage treatment processes are unable to avoid the dissemination of such resistant bacteria into the environment, resulting therefore in their contact with water- or soil-borne microorganisms.

In this study, no member of the domain Archaea was detected, although 144 sequences were initially retrieved using primer pairs designed to detect archaea. These

amplicons actually corresponded mainly to bacterial sequences, predominantly plastids. One primer (1510R) from the A pair was not archaeal specific and could be responsible in part for this mispriming (Baker and Cowan 2004). Archaea have been shown to be present in significant amounts in some marine plankton, soil, lakes, and river (Schleper and Nicol 2010). However, our results are in agreement with the previous report of Manz *et al.* (1999) on lotic biofilms in which they used using FISH technology.

The main feature in the overall composition of the biofilms studied was a very high number of rare sequences (<5 copies, 46–47%), which is per se informative. This number is clearly linked to the sequencing effort and is affected by PCR bias, but the relative amplicon abundance was significant for the overall biofilm composition. Rare sequences could represent microorganisms that are not physiologically active or having only a minor influence on biofilm physiology. However, it has been shown that a relatively low-abundance organism (around 0.3% of all cells) can be responsible for significant fractions of metabolite fluxes (Musat *et al.* 2008). Besemer *et al.* (2012) compared the rank-abundance curves of bulk and active communities with stream water biofilms and reported that they were indistinguishable. As these communities are composed of a few dominant OTUs and a large fraction of rare OTUs, it is likely that a certain fraction of the rare OTUs was active. However, these microorganisms may also constitute a potential adaptive resource for biofilm communities in response to environmental changes, allowing the communities to respond with the selective accelerated growth of adapted microorganisms. This part of the communities warrants further investigation, with regard to its potential biological processes.

Another important feature of the biofilms studied was the high proportion of Bacillariophyta plastid sequences. The presence of such sequences indicates that the physiology of the biofilm was strongly oriented toward phototrophy, and that eukaryotes played a major role in this adaptation. Bacillariophyta plastid sequences included raphid pennate, nonraphid pennate, and centric diatoms. Pennate diatoms generally harbor one, two, or four chloroplasts, whereas centric diatoms may contain many (Bourrelly 1981). This could explain why so few diatoms (only 212 eukaryotic diatom sequences were identified with primer set Euc2, of which 157 were from the *Navicula* species) were able to generate numerous plastid sequences (1771 sequences with primer set B2). Using an automated ribosomal intergenic spacer analysis, Fechner *et al.* (2010), in agreement with our results, showed that the most prevalent sequence type in natural freshwater biofilm was associated with the diatom *Navicula*, mostly

during the early summer season. Other phototrophic microorganisms also significantly contribute to biofilm composition: Chlorophyta, Phormidium Cyanobacteria and Rhodobacter, Rhodoferrax Proteobacteria.

Conclusion

Surface associated bacterial populations play a critical role in the ecosystem of rivers because they have greater activity than planktonic bacterial populations. This study, showing a potentially important photosynthetic activity, represents a preliminary assessment that will be useful in further investigations of naturally occurring seasonal changes or modifications during pollutant discharge recovering.

Acknowledgments

We thank Christelle Blavignac for her help in microscopy analysis and Christophe Portelli for his technical assistance. This study was partially supported by grant AFSSET EST-2007-6.

Conflict of Interest

None declared.

References

- Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 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.
- Andersson, A. F., M. Lindberg, H. Jakobsson, F. Backhed, P. Nyren, and L. Engstrand. 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE* 3:e2836.
- Andersson, A. F., L. Riemann, and S. Bertilsson. 2009. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* 4:171–181.
- Araya, R., K. Tani, T. Takagi, N. Yamaguchi, and M. Nasu. 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.
- Baker, G. C., and D. A. Cowan. 2004. 16 S rDNA primers and the unbiased assessment of thermophile diversity. *Biochem. Soc. Trans.* 32:218–221.
- Baker, G. C., L. A. Tow, and D. A. Cowan. 2003. PCR-based detection of non-indigenous microorganisms in 'pristine' environments. *J. Microbiol. Methods* 53:157–164.
- Besemer, K., H. Peter, J. B. Logue, S. Langenheder, E. S. Lindstrom, L. J. Tranvik, et al. 2012. Unraveling assembly of stream biofilm communities. *ISME J.* 6:1459–1468.
- Bourrelly, P. 1981. Les algues d'eau douce. Vol. II. Les Algues Jaunes et Brunes, Chrysophycées, Phaeophycées, Xanthophycées et Diatomées, rev. ed. N. Boubée et Cie, Paris. 517 pp.
- Bricheux, G., G. le Moal, C. Hennequin, G. Coffe, F. Donnadiou, C. Portelli, et al. 2013. Characterization and evolution of natural aquatic biofilm communities exposed in vitro to herbicides. *Ecotoxicol. Environ. Saf.* 88:126–134.
- Brown, C. D., and W. van Beinum. 2009. Pesticide transport via sub-surface drains in Europe. *Environ. Pollut.* 157: 3314–3324.
- Brummer, I. H., W. Fehr, and I. Wagner-Dobler. 2000. Biofilm community structure in polluted rivers: abundance of dominant phylogenetic groups over a complete annual cycle. *Appl. Environ. Microbiol.* 66:3078–3082.
- Brummer, I. H., A. Felske, and I. Wagner-Dobler. 2003. Diversity and seasonal variability of beta-Proteobacteria in biofilms of polluted rivers: analysis by temperature gradient gel electrophoresis and cloning. *Appl. Environ. Microbiol.* 69:4463–4473.
- Chouari, R., D. le Paslier, P. Daegelen, C. Dauga, J. Weissenbach, and A. Sghir. 2010. Molecular analyses of the microbial community composition of an anoxic basin of a municipal wastewater treatment plant reveal a novel lineage of proteobacteria. *Microb. Ecol.* 60:272–281.
- Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight. 2009. Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697.
- Dassarma, S., and E. F. Fleischmann. 1995. *Archea: a laboratory manual - halophiles*. Cold Spring Harbour Laboratory Press, New York.
- Dorigo, U., C. Leboulanger, A. Berard, A. Bouchez, J. F. Humbert, and B. Montuelle. 2007. Lotic biofilm community structure and pesticide tolerance along a contamination gradient in a vineyard area. *Aquat. Microb. Ecol.* 50:91–102.
- Dorigo, U., A. Berard, F. Rimet, A. Bouchez, and B. Montuelle. 2010. In situ assessment of periphyton recovery in a river contaminated by pesticides. *Aquat. Toxicol.* 98:396–406.
- Ehrig, A., R. Müller, and W. Babel. 1997. Isolation of phenoxo herbicide-degrading Rhodoferrax species from contaminated building material. *Acta Biotechnol.* 17:351–356.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138.
- Engelbrekton, A., V. Kunin, K. C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, et al. 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J.* 4:642–647.

- Fechner, L. C., F. Vincent-Hubert, P. Gaubert, T. Bouchez, C. Gourlay-France, and M. H. Tusseau-Vuillemin. 2010. Combined eukaryotic and bacterial community fingerprinting of natural freshwater biofilms using automated ribosomal intergenic spacer analysis. *FEMS Microbiol. Ecol.* 74:542–553.
- Galvin, S., F. Boyle, P. Hickey, A. Vellinga, D. Morris, and M. Cormican. 2010. Enumeration and characterization of antimicrobial-resistant *Escherichia coli* bacteria in effluent from municipal, hospital, and secondary treatment facility sources. *Appl. Environ. Microbiol.* 76:4772–4779.
- Hansen, M. C., T. Tolker-Nielsen, M. Givskov, and S. Molin. 1998. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol. Ecol.* 26:141–149.
- Hong, P. Y., C. Hwang, F. Ling, G. L. Andersen, M. W. Lechevallier, and W. T. Liu. 2010. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Appl. Environ. Microbiol.* 76:5631–5635.
- Huber, J. A., D. B. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A Butterfield, et al. 2007. Microbial population structures in the deep marine biosphere. *Science* 318:97–100.
- Huse, S. M., L. Dethlefsen, J. A. Huber, D. M. Welch, D. A. Relman, and M. L. Sogin. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* 4:e1000255.
- Kim, M., M. Morrison, and Z. Yu. 2011a. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* 84:81–87.
- Kim, M., M. Morrison, and Z. Yu. 2011b. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* 84:81–87.
- Koren, O., A. Spor, J. Felin, F. Fak, J. Stombaugh, V. Tremaroli, et al. 2010. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl Acad. Sci.* 108(Suppl. 1):4592–4598.
- Kwon, S., E. Moon, T. S. Kim, S. Hong, and H. D. Park. 2011. Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. *Microbes Environ.* 26:149–155.
- Lauber, C. L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75:5111–5120.
- Lavoie, I., W. Vincent, R. Pienitz, and J. Painchaud. 2004. Benthic algae as indicators of agricultural pollution in the streams and rivers of southern Québec (Canada). *Aquat. Ecosyst. Health Manage.* 7:43–58.
- Lazarevic, V., K. Whiteson, D. Hernandez, P. Francois, and J. Schrenzel. 2010. Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics* 11:523.
- Lee, T. K., T. van Doan, K. Yoo, S. Choi, C. Kim, and J. Park. 2010. Discovery of commonly existing anode biofilm microbes in two different wastewater treatment MFCs using FLX Titanium pyrosequencing. *Appl. Microbiol. Biotechnol.* 87:2335–2343.
- Ling, Z., J. Kong, F. Liu, H. Zhu, X. Chen, Y. Wang, et al. 2010. Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis. *BMC Genomics* 11:488.
- Liu, Z. Z., T. Z. Desantis, G. L. Andersen, and R. Knight. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* 36:11.
- Liu, R., Z. Yu, H. Guo, M. Liu, H. Zhang, and M. Yang. 2012. Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. *Sci. Total Environ.* 435–436:124–131.
- Manz, W., K. Wendt-Potthoff, T. R. Neu, U. Szewzyk, and J. R. Lawrence. 1999. Phylogenetic composition, spatial structure, and dynamics of lotic bacterial biofilms investigated by fluorescent in situ hybridization and confocal laser scanning microscopy. *Microb. Ecol.* 37:225–237.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- Martins da Costa, P., P. Vaz-Pires, and F. Bernardo. 2006. Antimicrobial resistance in *Enterococcus* spp. isolated in inflow, effluent and sludge from municipal sewage water treatment plants. *Water Res.* 40:1735–1740.
- Medlin, L., H. J. Elwood, S. Stickel, and M. L. Sogin. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71:491–499.
- Moura, A., I. Henriques, R. Ribeiro, and A. Correia. 2007. Prevalence and characterization of integrons from bacteria isolated from a slaughterhouse wastewater treatment plant. *J. Antimicrob. Chemother.* 60:1243–1250.
- Musat, N., H. Halm, B. Winterholler, P. Hoppe, S. Peduzzi, F. Hillion, et al. 2008. A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc. Natl Acad. Sci. USA* 105:17861–17866.
- Nocker, A., M. Burr, and A. K. Camper. 2007. Genotypic microbial community profiling: a critical technical review. *Microb. Ecol.* 54:276–289.
- Pinto, A. J., and L. Raskin. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS ONE* 7:e43093.
- Polz, M. F., and C. M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* 64:3724–3730.
- Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, et al. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.

- Qian, P. Y., Y. Wang, O. O. Lee, S. C. Lau, J. Yang, F. F. Lafi, et al. 2010. Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. *ISME J.* 5:507–518.
- Rajendhran, J., and P. Gunasekaran. 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol. Res.* 166:99–110.
- Reysenbach, A. L., G. S. Wickham, and N. R. Pace. 1994. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl. Environ. Microbiol.* 60:2113–2119.
- Ribeiro, F., D. Przybylski, S. Yin, T. Sharpe, S. Gnerre, A. Abouelleil, et al. 2012. Finished bacterial genomes from shotgun sequence data. *Genome Res.* doi: 10.1101/gr.141515.112.
- Roesch, L. F., R. R. Fulthorpe, A. Riva, G. Casella, A. K. Hadwin, A. D. Kent, et al. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1:283–290.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooshef, et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5:e77.
- Sanapareddy, N., T. J. Hamp, L. C. Gonzalez, H. A. Hilger, A. A. Fodor, and S. M. Clinton. 2009. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. *Appl. Environ. Microbiol.* 75:1688–1696.
- Schleper, C., and G. W. Nicol. 2010. Ammonia-oxidising archaea—physiology, ecology and evolution. *Adv. Microb. Physiol.* 57:1–41.
- Schloss, P. D. 2009. A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE* 4:e8230.
- Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, et al. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl Acad. Sci. USA* 103:12115–12120.
- Tedersoo, L., R. H. Nilsson, K. Abarenkov, T. Jairus, A. Sadam, I. Saar, et al. 2010. 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol.* 188:291–301.
- Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, et al. 2009. A core gut microbiome in obese and lean twins. *Nature* 457:480–484.
- Vaz-Moreira, I., C. Egas, O. C. Nunes, and C. M. Manaia. 2011. Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample. *Antonie Van Leeuwenhoek* 100:245–257.
- Vetriani, C., H. W. Jannasch, B. J. Macgregor, D. A. Stahl, and A. L. Reysenbach. 1999. Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* 65:4375–4384.
- Vilanova, X., A. Manero, M. Cerda-Cuellar, and A. R. Blanch. 2004. The composition and persistence of faecal coliforms and enterococcal populations in sewage treatment plants. *J. Appl. Microbiol.* 96:279–288.
- Youssef, N., C. S. Sheik, L. R. Krumholz, F. Z. Najar, B. A. Roe, and M. S. Elshahed. 2009. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl. Environ. Microbiol.* 75:5227–5236.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Data from the pyrosequencing. Diversity: sum of amplicons different found in a given category (Phylum, Class...). Abundance: sum of amplicons found in a given category (Phylum, Class...).