**Materials and Methods**

**DNA and RNA extraction, PCR and high-throughput amplicon sequencing**

Total RNA and DNA were extracted from four samples as previously described (Pitkänen et al., 2013) with some minor modifications. Briefly, the AllPrep DNA/RNA Mini Kit (Qiagen GmbH, Hilden, Germany) was used to extract total nucleic acid. RNA was further purified using Ambion TURBO DNA-free DNase kit (Life Technologies, Grand Island, NY). The concentration and purity of RNA and DNA were determined using Qubit RNA and dsDNA HS assay kits and the Qubit 2.0 Fluorometer (Life Technologies). cDNA was generated using random hexamer primed Superscript III system for RT-PCR (Life Technologies). Samples (cDNA and DNA) were stored at -20°C until used for next generation sequencing. Specifically, we used barcoded 16S rRNA gene targeting primers (i.e., 515F and 806R) (Caporaso et al., 2011) and sequenced the targeted product (i.e., 291 bp) in both directions using an Illumina MiSeq PE250 approach.

**Next generation sequencing data preprocessing and analysis**

Sequence reads (16S rDNA- and 16S rRNA-based) were processed and analyzed using Mothur software (Schloss et al., 2009). Sequence reads that did not fit the following criteria were discarded from further analyses: did not form contigs, deviated considerably from the expected PCR size product, identified as chimeras, had ambiguous bases, and had homopolymers greater than 7 bases long. Sequence reads were grouped at a 97 % similarity and the consensus sequences were then identified using Mothur and the Silva (Quast et al., 2013) database as a reference. Excel was used to determine the overall relative abundance of representative sequences at different taxonomic levels (e.g., class, order, family, genus). Sequences were analyzed using Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>) and RDP classifier (Wang et al., 2007) further confirm their phylogenetic affiliation and to classify sequences at a low taxonomic level (genus and species) whenever possible.

**Results**

High throughput sequencing databases were generated using Illumina MiSeqPE250. Two independent libraries targeting bacterial 16S ribosomal RNA genes (rDNA) and transcripts (rRNA) were developed to describe total bacterial community composition and metabolically active bacterial members, respectively.

A total of 99,199 and 111,471 sequences from two biofilm samples were used for rDNA and rRNA libraries, respectively (Table 1). Proteobacteria was the most abundant phyla in both DNA and RNA libraries. Other dominant rDNA classes found were Bacteroidia, Cloacamonae, Clostridia, Flavobacteria, and Spirochaetes, whereas only two classes such as Cloacamonae and Spirochaetes were found in RNA libraries. Overall, DNA libraries showed more diverse populations at genus level taxonomy than RNA ones (Table 1), suggesting the high metabolic activity of the dominant bacteria.

*Geobacter* spp. within a class of delta-proteobacteria were the most numerically abundant member in both DNA and RNA libraries. Compared to the DNA libraries, *Geobacter* was more dominant in the RNA libraries. In the MMXC biofilm sample, the abundance of *Geobacter* was over 93% and consequently only few other abundant members were found. In the pH-MXC biofilm sample, there were relatively high diverse populations in the bacterial community as the relative abundance of *Geobacter* decreased. Unlike the MMXC sample, about 10% and 15% of sequences were composed of the rare members (i.e., less than 20 sequences) in DNA and RNA libraries, respectively. Besides *Geobacter*, the bacterial community in the DNA library was mostly composed of *Azospirillum*, Comamonadaceae, Cloacamonaceae, *Desulfovibrio*, *Dysgonomonas*, Enterobacteriaceae, *Myroides*, *Petrimonas*, and *Pseudomonas*. Conversely, the relative abundance of these bacteria except for *Azospirillum* and Cloacamonaceae was almost negligible in the RNA library, suggesting metabolically inactive bacterial members.

TABLE 1. Distribution of bacterial 16S rDNA and 16S rRNA

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Genus | MMXC | pH-MXC |
|  |  | DNA(n=51016) | RNA(n=48183) | DNA(n=74648) | RNA(n=36823) |
| Alpha-Proteobacteria  | *Azospirillum**Bradyrhizobiaceae\***Hyphomicrobiaceae\***Telmatospirillum* | ---- | ---- | 620 (1.3%)5152247 | 1275 (3.5%)-35- |
| Beta-Proteobacteria | *Alcaligenaceae\***Comamonadaceae\***Rhodocyclaceae\** | 32-21 | --21 | 366528 (1.1%)41 | 3928- |
| Delta-Proteobacteria | *Desulfovibrio**Geobacter* | 13747362 (93%) | -73185 (98%) | 598 (1.2%)33724 (70%) | 15828807 (78%) |
| Gamma-Proteobacteria | *Aeromonas**Enterobacteriaceae\***Pseudomonas* | 1571633 (3.2%)- | --- | 2921017 (2.1%)2350 (4.9%) | --97 |
| Bacteroidia  | *Dysgonomonas**Petrimonas**Proteiniphilum* | -366- | --- | 645 (1.3%)738 (1.5%)206 | --- |
| Cloacamonae | *Cloacamonaceae\** | - | - | 1632 (3.4%) | 1074 (2.9%) |
| Clostridia | *Anaerovorax**Fusibacter**Ruminococcaceae\** | 252723 | --- | 403064 | --- |
| Flavobacteria | *Myroides* | - | - | 508 (1.1%) | - |
| Spirochaetes | *Treponema* | 242 | 198 | 216 | - |

- (not found or less than 20 sequences)

\* Family

References

1. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl , 4516–4522.
2. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new window. Nucl. Acids Res. 41 (D1), D590-D596.
3. Pitkänen, T., Ryu, H., Elk, M., Hokajärvi, A.-M., Siponen, S., Vepsäläinen, A., Räsänen, P., Santo Domingo, J.W., 2013. Detection of fecal bacteria and source tracking identifiers in environmental waters using rRNA-based RT-qPCR and rDNA-based qPCR assays. Environ. Sci. Technol. 47, 13611-13620.
4. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing Mothur: open-source, platform-independent, community- supported software for describing and comparing microbial communities. Appl Environ Microbiol 75, 7537-7541.
5. Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 73(16), 5261-5267.