**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 231,381 and 255,013 sequences were used for rDNA and rRNA libraries, respectively. Proteobacteria was the most abundant phyla in both DNA (about 40-75%) and RNA (50-98%) libraries followed by a member of Spirochaetes class. Other dominant rDNA classes found were Bacteroidia, Clostridia, Elusimicrobia, and Syergistia, whereas significantly lower abundance was found in RNA libraries. Overall, DNA libraries showed more diverse populations at genus level taxonomy than RNA ones (Tables 1 and 2), suggesting the high metabolic activity of the dominant bacteria (e.g., *Geobacter* and *Treponema*).

Members of *Geobacter* and *Treponema* genera were the most numerically abundant bacteria in both DNA and RNA libraries, and their relative abundance is closely associated each other. *Geobacter* within a class of delta-proteobacteria was more dominant than other species, but in Sample 3 its relative dominance (34%) was close to *Treponema* population (29%). Compared to the DNA libraries, *Geobacter* was much more dominant in the RNA libraries. *Treponema* within a class of Spirochaetes was the second most dominant member in both DNA and RNA libraries. Specifically, in the RNA library, the abundance of *Treponema* (40%) increased significantly in Sample 4 as *Geobacter* population decreased (Table 2). Besides these two dominant bacteria, based on the DNA library, the bacterial community was mostly composed of *Acidaminococcus, Aminiphilus, Bacteroides, Desulfovibrio, Elusimicrobium, and Pseudomonadaceae* (Table 1). In contrast, the relative abundance of these bacteria was almost negligible in RNA libraries due to the predominance of two *Geobacter* and *Treponema* genera (Table 2).

TABLE 1. Distribution of bacterial 16S rDNA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Genus | Sample1(n=79655) | Sample2(n=33908) | Sample3(n=60457) | Sample4(n=57361) |
| Alpha-Proteobacteria  | *Agrobacterium**Telmatospirillum* | 66138 | 116109 | 41317 | 158219 |
| Beta-Proteobacteria | *Achromobacter**Comamonadaceae\***Rhodocyclaceae\** | 228-152 | 249-169 | 450-92 | 30972201 |
| Delta-Proteobacteria | *Desulfovibrio**Geobacter**Pelobacter* | 45658096 (73%)128 | 1084 (3.2%)16870 (50%)66 | 4232 (7.0%)20800 (34%)117 | 2642 (4.6%)30838 (54%)96 |
| Epsilon-Proteobacteria | *Campylobacter* | - | - | 66 | 50 |
| Gamma-Proteobacteria | *Pseudomonadaceae\***Stenotrophomonas* | 1233 (1.6%)71 | 1082 (3.2%)- | 1333 (2.2%)56 | 1618 (2.8%)70 |
| Bacteroidia  | *Bacteroides**Dysgonomonas**S24-7\** | 148125538 | 22978365 (1.1%) | 1465 (2.4%)155226 | 1448 (2.5%)1711301 (2.3%) |
| Cloacamonae | *Cloacamonaceae\** | 336 | 179 | 107 | 114 |
| Clostridia | *Acidaminococcus**Anaerovorax**Christensenellaceae\***Oscillospira**Ruminococcus**vadinHB04* | 324 1696238019887 | 396 (1.2%)16053255234- | 1034 (1.7%)299137757526470 | 647 (1.1%)18768681230169 |
| Elusimicrobia | *Elusimicrobium* | 752 (1.0%) | 396 (1.2%) | 813 (1.3%) | 441 (0.8%) |
| Erysipelotrichi | *RFN20* | 388 | 281 | 156 | 183 |
| Spirochaetes | *Treponema* | 5352 (6.7%) | 6258 (19%) | 17785 (29%) | 9445 (17%) |
| Synergistia | *Aminiphilus**Dethiosulfovibrionaceae\** | 840 (1.1%)132 | 746 (2.2%)78 | 2435 (4.0%)393 | 1378 (2.4%)177 |

- (not found or less than 50 sequences)

\* Family

TABLE 2. Distribution of bacterial 16S rRNA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Genus | Sample1(n=112689) | Sample2(n=46565) | Sample3(n=49883) | Sample4(n=45876) |
| Alpha-Proteobacteria  | *Telmatospirillum**Xanthobacter* | -- | -- | 50- | 6860 |
| Beta-Proteobacteria | *Achromobacter**Comamonadaceae\***Rhodocyclaceae\** | --97 | --93 | --58 | 50164436 (1.0%) |
| Delta-Proteobacteria | *Desulfovibrio**Geobacter**Pelobacter* | -108437 (96%)174 | -40931 (88%)128 | 6540622 (81%)203 | 5520336 (44%)191 |
| Epsilon-Proteobacteria | *Campylobacter* | - | - | - | 90 |
| Gamma-Proteobacteria | *Pseudomonadaceae\** | 1498 (1.3%) | 790 (1.7%) | 490 (1.0%) | 2451 (5.3%) |
| Cloacamonae | *Cloacamonaceae\** | 50 | 57 | - | - |
| Clostridia | *Anaerovorax**Oscillospira* | -- | -- | 63- | 7954 |
| Elusimicrobia | *Elusimicrobium* | - | 63 | - | 228 |
| Spirochaetes | *Treponema* | 1100 (1.0%) | 3271 (7.0%) | 6303 (13%) | 18107 (40%) |
| Synergistia | *Aminiphilus**Dethiosulfovibrionaceae*  | 6756 | 146- | 385188 | 1507 (3.3%)159 |

- (not found or less than 50 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.