**SUPPLEMENTARY INFORMATION:**

# Automated Sampling Procedures Supported by High Persistence of Bacterial Fecal Indicators and *Bacteroidetes* Genetic Microbial Source Tracking Markers in Municipal Wastewater during Short-Term Storage at 5°C

Mayer RE[1,2], Vierheilig J[1,2,5,§], Egle L[3,5], Reischer GH [1,2], Saracevic E[3], Mach RL[1,2], Kirschner AKT[2,4], Zessner M[3,5], Sommer R[2,4*], and Farnleitner AH[1,2]

[1] Institute of Chemical Engineering, Research Division Biotechnology and Microbiology, Research Group Environmental Microbiology and Molecular Ecology, Vienna University of Technology, Gumpendorferstraße 1a/166-5-2, A-1060 Vienna, Austria

[2] InterUniversity Cooperation Centre for Water and Health, www.waterandhealth.at

[3] Institute for Water Quality Resources and Waste Management, Vienna University of Technology, Karlsplatz 13/226, 1040 Vienna, Austria

[4] Medical University Vienna, Institute for Hygiene and Applied Immunology, Water Hygiene, Kinderspitalgasse 15, A-1090 Vienna, Austria

[5] Centre for Water Resource Systems, Vienna University of Technology, Karlsplatz 13/222, 1040 Vienna, Austria

[§] Present address: Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, University of Vienna, Althanstraße 14, A-1090 Vienna, Austria

[#] Corresponding author: regina.sommer@meduniwien.ac.at, Tel.: +43(0)1-40160-33051 (secr.-33050)

**Table S1. Investigated wastewater treatment plants.**

| WWTP | design capacity [PE] | actual average loading [PE]*** | inhabitants connected | sludge age (average) [d] | wastewater treatment | COD** In/Ef [mg $L^{-1}$] | TN** In/Ef [mg $L^{-1}$] | TP** In/Ef [mg $L^{-1}$] |
|---|---|---|---|---|---|---|---|---|
| 2 | 40,000 | 48,700* | 23,500 | 8-10 | M, C, N, P | 500/45 | 54/26 | 10/1.0 |
| 3 | 23,000 | 13,600 | 10,800 | 22-57 | M, C, N, D, P | 560/20 | 54/12 | 9/0.4 |
| 4 | 140,000 | 44,000 | 30,800 | 13 | M, C, N, D, P | 460/14 | 45/9 | 4/0.3 |

Abbreviations: *overloaded WWTP, **Annual mean values, *** Annual mean chemical oxygen demand (COD) load (kg/a) divided by a COD load per person of 110 g COD/d; TN: Total Nitrogen; TP: Total Phosphorus; In, influent; Ef, effluent; M, primary treatment: mechanical treatment step; C, secondary treatment: biological carbon removal; tertiary treatment: nutrient removal including nitrification (N), denitrification (D) and phosphorus removal (P)

**Analysis of chemophysical parameters.** The chemical oxygen demand was measured as described in DIN 38409-43 (1). Total phosphorus and total nitrogen were investigated with a Merck 500 microwave + SKALAR segment flow analyzer (Skalar, Netherlands) according to the ISO standards (2, 3).

**Next Generation Sequencing (NGS).** The DNA extracts (n = 16) of the one representative microcosm series from the WWTP2 effluent, which was chosen for additional 16S rDNA 454 pyrosequencing analysis, were used as templates in PCR to amplify the variable regions V1-V2 of the 16S rRNA gene for 25 cycles. All reactions were run in triplicate with the bacterial specific primers S-D-Bact-0008-a-S-20 (5′-AGAGTTTGATCCTGGCTCAG-3′, as described by Edwards et al. (4), and S-D-Bact-0338-a-A-19 (5′-TGCTGCCTCCCGTAGGAGT-3′, as described by Etchebehere and Tiedje (5), the latter equipped with a distinct 12-nucleotide error-correcting Golay barcode for each extract as a multiplex tag (6-8). The nomenclature for the PCR primers was standardized according to Alm et al. (9). The sample amplicons (n = 16) were purified, pooled in equimolar amounts and sent to Selah Clinical Genomic Center, formerly EnGenCore (Columbia, SC, USA) for 454 pyrosequencing (titanium chemistry).

**Bioinformatics** Sequence analysis was done using the software package Quantitative Insights Into Microbial Ecology, QIIME (10). Raw sequences (n = 214,978) were quality filtered and assigned to the samples according to their barcodes. The flowgrams were denoised to reduce sequencing noise (11). After removing the primers, chimeric sequences identified by de novo (abundance based) and reference based chimera detection with UCHIME were filtered out (12, 13). Remaining sequences (n = 185,374) were binned into Operational Taxonomic Units (OTUs) using USEARCH, with a minimum pairwise identity of 97 % (13). Greengenes OTUs (97 %; version May 2013) were specified as a reference database at the previous two steps (14). Rare OTUs represented by less than four sequences were filtered out, leading to 182,914 remaining sequences for further analysis.. The most abundant sequence in each OTU was chosen as a representative and aligned using PyNAST (15) and the Greengenes reference alignment (14) trimmed to the V1-V2 region of the 16S rRNA gene (16) with a minimum percent identity of 75 %. The hypervariable regions were filtered out with the V1-V2 trimmed version of the lanemask and a phylogenetic tree was constructed using FastTree (17). Taxonomy was assigned with the Ribosomal Database Project (RDP) classifier (18) with a minimum confidence of 80 % and the greengenes taxonomy of May 2013 (19). The sequences assigned to the phylum *Bacteroidetes* were filtered out. Subsequently, 515 *Bacteroidetes* sequences (i.e. smallest number of taxon-specific sequences per sample) were

randomly selected from each sample for further analyses (rarefaction). To compare the diversity within this taxon between the samples, we calculated the unweighted UniFrac distance metric (20) for the phylum *Bacteroidetes* and clustered the resulting metric using principle coordinate analysis to visualise the phylogenetic relatedness of these communities.

Sequence data from this project is available in the Sequence Read Archive of the National Center for Biotechnology Information under the study accession number SRP059025.

**Literature:**


1. **DIN.** 1981. 38409-43; German standard methods for the analysis of water, waste water and sludge; Summary action and material characteristic parameters (group H); Determination for the chemical oxygen demand (COD); short duration method (H 43).
2. **ISO.** 2004. Water quality - Determination of phosphorus - Ammonium molybdate spectrometric method (ISO 6878:2004). International Organisation of Standardisation, Geneva, Switzerland.
3. **ISO.** 1997. Water quality - Determination of nitrogen - Part 1: Method using oxidative digestion with peroxodisulfate (ISO 11905-1:1997); International Organisation of Standardisation, Geneva, Switzerland.
4. **Edwards U, Rogall T, Blocker H, Emde M, Bottger EC.** 1989. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Res. **17:**7843-7853.
5. **Etchebehere C, Tiedje J.** 2005. Presence of Two Different Active nirS Nitrite Reductase Genes in a Denitrifying Thauera sp. from a High-Nitrate-Removal-Rate Reactor. Appl. Environ. Microbiol. **71:**5642-5645.
6. **Fierer N, Hamady M, Lauber CL, Knight R.** 2008. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. P Natl Acad Sci USA **105:**17994-17999.
7. **Golay MJE.** 1949. Notes on Digital Coding. Proc. IRE **37:**657.
8. **Hamady M. WJJ, Harris J. K., Gold N. J. and Knight R. .** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods **5:**235-237.
9. **Alm EW, Oerther DB, Larsen N, Stahl DA, Raskin L.** 1996. The oligonucleotide probe database. Appl. Environ. Microbiol. **62:**3557-3559.
10. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7:**335-336.
11. **Reeder J, Knight R.** 2009. The 'rare biosphere': a reality check. Nat. Methods **6:**636-637.

12. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:**2194-2200.

13. **Edgar R.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460 - 2461.

14. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72:**5069-5072.

15. **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.** 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics **26:**266-267.

16. **Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R, Ley RE.** 2012. Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. ISME J **6:**94-103.

17. **Price MN, Dehal PS, Arkin AP.** 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol. Biol. Evol. **26:**1641-1650.

18. **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.

19. **McDonald D, Price M, Goodrich J, Nawrocki E, DeSantis T, Probst A, Andersen G, Knight R, Hugenholtz P.** 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J **6:**610-618.

20. **Lozupone C, Knight R.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. **71:**8228-8235.