SUPPLEMENTARY MATERIAL

The activated sludge ecosystem contains a core community of abundant organisms

Aaron M. Saunders[1], Mads Albertsen[1], Jes Vollertsen[2], Per H. Nielsen[1]*

1: Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, Aalborg, 9220, Denmark.

2: Department of Civil Engineering, Aalborg University, Sohngaardsholmsvej 57, Aalborg, 9000, Denmark.

Contents

1. Supplementary Methods

1.1 Calculation of net growth rate using mass balance

The net growth rate of an organism can be described as its growth rate minus its decay rate. If one assumes that growth and decay both can be described as a first order process, it follows that the net growth rate also can be described as such process. For an organism *x*, the mass balance equation for a treatment plant can hence be stated as:

$$\frac{dN_{x,AS}}{dt} = kN_{x,AS} + n_{x,ww} - n_{x,SP} \qquad \text{(eq. 1)}$$

where

$N_{x,AS}$	number of organism *x* in the activated sludge of the plant [-]
k	net growth rate constant (positive for net growth, negative for net decay) [d^{-1}]
$n_{x,ww}$	number or organism *x* entering the plant with the wastewater per day [d^{-1}]
$n_{x,SP}$	number or organism *x* exiting the plant with the surplus sludge per day [d^{-1}]

At steady state there is no net change in the number of cells in the activated sludge biomass $N_{x,AS}$:

$$\frac{dN_{x,AS}}{dt} = kN_{x,AS} + n_{x,ww} - n_{x,SP} = 0 \qquad \text{(eq. 2)}$$

Rearranging equation 2 leads to the net rate constant being defined as:

$$k = \frac{n_{x,SP} - n_{x,ww}}{N_{x,AS}} \qquad \text{(eq. 3)}$$

Defining the relative abundance of a certain organism in the wastewater inflow as p_{ww}, the relative abundance of the same organism in the activated sludge as p_{AS}, and noticing that the relative abundance of a certain organism in the surplus sludge is the same as in the activated sludge, equation 3 can be rewritten as:

$$k = \frac{p_{AS}n_{SP} - p_{ww}n_{ww}}{p_{AS}N_{AS}} \qquad \text{(eq. 4)}$$

where

N_{AS}	total number of organism*s* in the activated sludge of the plant [-]
n_{ww}	total number of organism*s* entering the plant with the wastewater per day [d^{-1}]
n_{SP}	total number of organism*s* exiting the plant with the surplus sludge per day [d^{-1}]

The actual values of N_{AS}, n_{ww} and n_{SP} in equation 4 depend on the treatment plant in question:

- The total number of organisms entering the plant with the wastewater per day is governed by the wastewater load and the abundance of cells herein.
- The total number of organisms in the activated sludge of the plant is mainly governed by the wastewater load, wastewater organic matter content, the sludge retention time (sludge age) and the overall yield of the treatment process.
- The total number or organisms exiting the plant with the surplus sludge per day is mainly governed by the same parameters.

The effluent population was not included in the mass balance when calculating the growth rate as the amount of solids leaving a plant with well-functioning solids separation is only a fraction of the amount of solids leaving with the surplus sludge. The relative abundance of the populations leaving with the effluent is also essentially the same as the sludge (Morgan-Sagastume et al., 2008). The plants have a hydraulic residence time of approx. 24 h in most of which (> 15-18 h) the water is mixed with the sludge flocs. Microscopy of influent wastewater, demonstrates that the biomass is mostly present as in forms of pieces of biofilms (from sewer biofilm) or other aggregates. These cannot resist being captured into the other flocs

For an actual treatment plant all parameter values of equation 4 can be determined directly by measurement. For the general example in this study, the parameters are determined for a 'typical Danish treatment plant' with biological nutrient removal defined by the following characteristics:

- The treatment plant receives 10^5 m^3 d^{-1} of wastewater (V_w)
- The wastewater contains 0.582 kg-COD m^{-3}, corresponding to 0.410 kg-OM m^{-3} ($C_{org,ww}$) (Vollertsen *et al*. 2001)
- The observed sludge yield (Y_{obs}) is 0.28 [kg-OM (kg-OM)$^{-1}$] (Henze *et al*. 2002)
- The total sludge retention time (θ) is 35 days

Determination of n_{ww}

For the present calculations, a total abundance of organisms in the wastewater was assumed equal to previous reports for wastewater from the same region (Vollertsen *et al*. 2001): average dry weather sewage 1.4 x 10^{14} cells m^{-3} (DAPI). n_{ww} hence becomes 1.4 x 10^{14} x 10^5 m^3 d^{-1} = 1.4 x 10^{19} cells d^{-1}.

Determination of N_{AS}

For activated sludge the total abundance of organisms was assumed to 2 x 10^{15} cells kg-OM^{-1} (Morgan-Sagastume *et al*. 2008). The concentration of each OTU is its relative abundance (determined from amplicon sequencing) multiplied by the total cell concentration.
The mass of sludge in this 'typical' treatment plant can be found as θ x Y_{obs} x $C_{org,ww}$ x V_w = 35 x 0.28 x 0.410 x 10^5 = 401,800 kg-OM. Multiplying with the total abundance of organisms, the total number of cells in the plant becomes 401,800 x to 2 x 10^{15} = 8.04 10^{20} cells.

Determination of n_{SP}

The number of cells taken out with the surplus sludge is determined as N_{AS} / θ, i.e. 8.04 10^{20} / 35 = 2.30 10^{19} cells d^{-1}.

The net growth rate constant can hence for the 'typical Danish treatment plant' of this example be found from the following equation:

$$k = \frac{p_{AS}\,2.30\ 10^{19} - p_w 1.4\ 10^{19}}{p_{AS}\,8.04\ 10^{20}} \qquad \text{(eq. 5)}$$

1.2 Assumptions when using amplicon data quantitatively

Equating observed read abundance with the actual abundance of specific populations has a number of potential technical biases, most importantly the variation in the 16S rRNA gene copy number, differences in the efficiency of DNA extraction and differences in the specificity of the primer pairs used and the and phylogenetic information in the amplicon. These biases are universal for amplicon sequencing. The correction of the different copy numbers is not possible due to a lack of reference genomes for the organisms in the activated sludge system (Albertsen *et al*. 2013), but should be made possible as more genomes become available (Kembel *et al*. 2012).

The further assumption that the abundance of an organism is proportional to its carbon turnover assumes a constant biomass yield (per mass of carbon consumed). While this may be a reasonable assumption for aerobic heterotrophs, a considerable fraction of the carbon is removed by denitrifying heterotrophs that have a slightly lower yield - 0.5 compared with 0.6 and aerobic yield (Henze *et al*. 2002). Some carbon is also metabolized by fermentation under anaerobic conditions where the average yield is only 0.2. However, the fraction of the anaerobic residence time is limited and most of the characterized organisms that employ fermentation under *in situ* conditions are facultative anaerobes (e.g. *Tetrasphaera* (Kristiansen *et al*. 2013) and *Trichoccocus* (Nielsen *et al*. 2012)), which are also active denitrifiers and/or aerobic heterotrophs. The set of OTUs observed within the Aalborg West plant was highly-stable over a 6 year period, which suggests that the two sample points for the other plants were likely to have captured a representative snapshot of the diversity.

2. Supplementary Figures

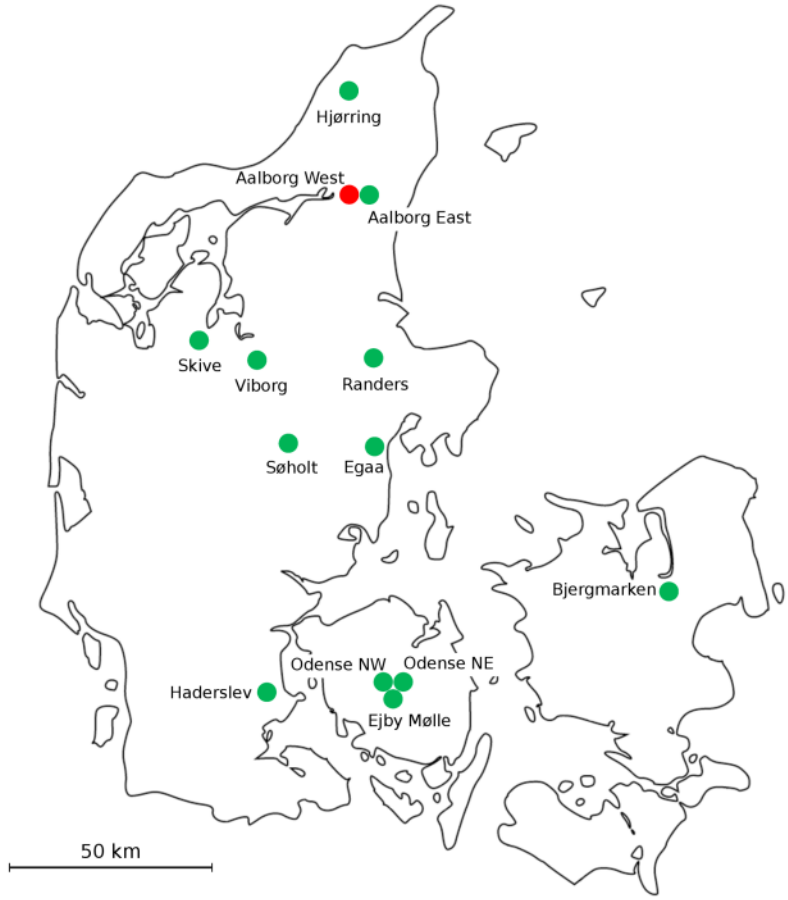

Figure S1: Location of the sampled activated sludge plants in Denmark. Plants were sampled 2 times in successive years (green) except Aalborg West (red), which was sampled 14 times over 5 successive years.

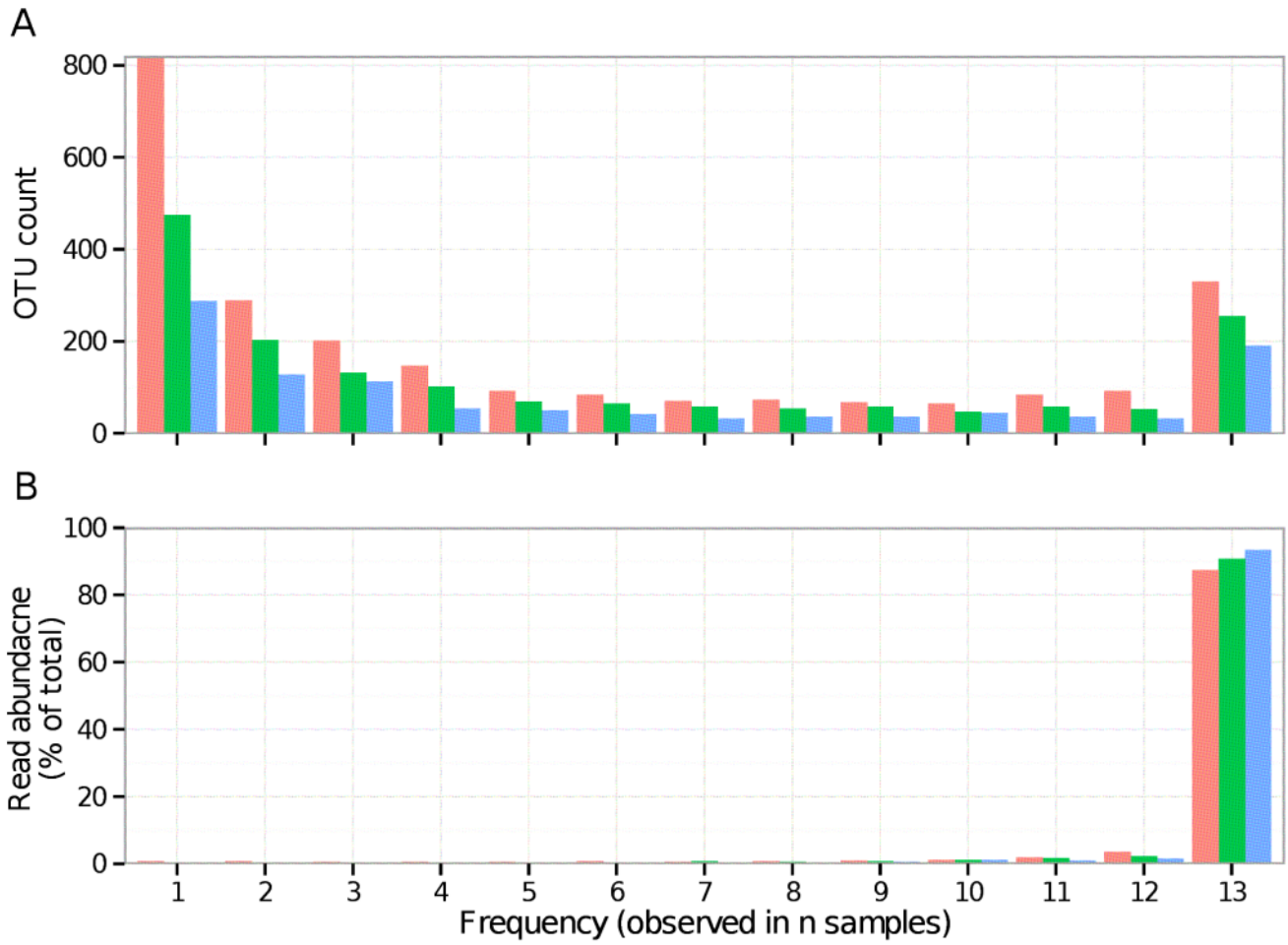

Figure S2: Frequency distribution of OTUs among samples of activated sludge (n = 13) taken at Aalborg West in a time series over 5 years. The x axis shows the number of samples in which each OTU is observed, and colors denote OTUs clustered at subspecies-level (red), species-level (green), genus-level (blue). The upper graph (A) shows the number of OTUs observed at each frequency, which demonstrates a core community of OTUs observed in every sample. The lower graph (B) show the %reads for OTUs observed at that frequency (bars) and demonstrates that core OTUs made up a large fraction of the total reads.

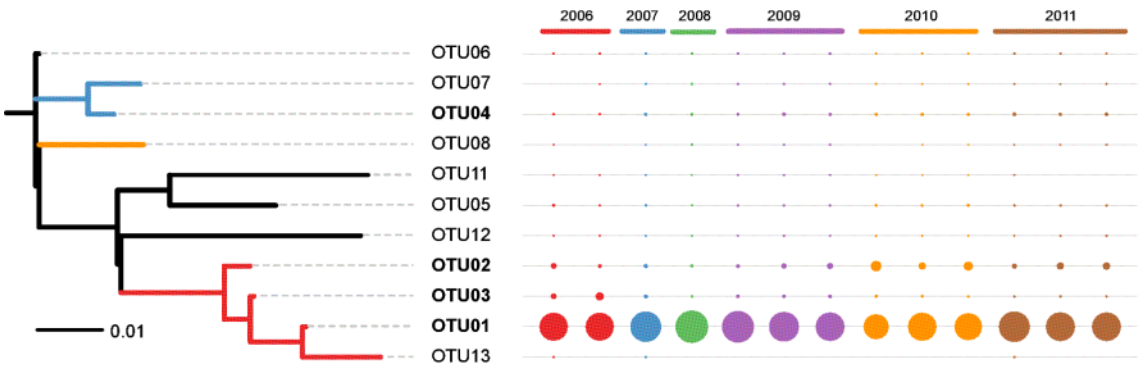

Figure S3: Relative abundance of 99% OTUs *Tetrasphaera* across 6 years in Aalborg West. The relationship between the OTUs is presented as a maximum likelihood phylogenetic tree; the colored branches denote OTUs from clade 1 (blue), clade 2 (yellow) and clade 3 (red). The size of the circles denotes the relative abundance of each OTU. The OTUs are numbered by decreasing average abundance and those in bold were the most abundant.

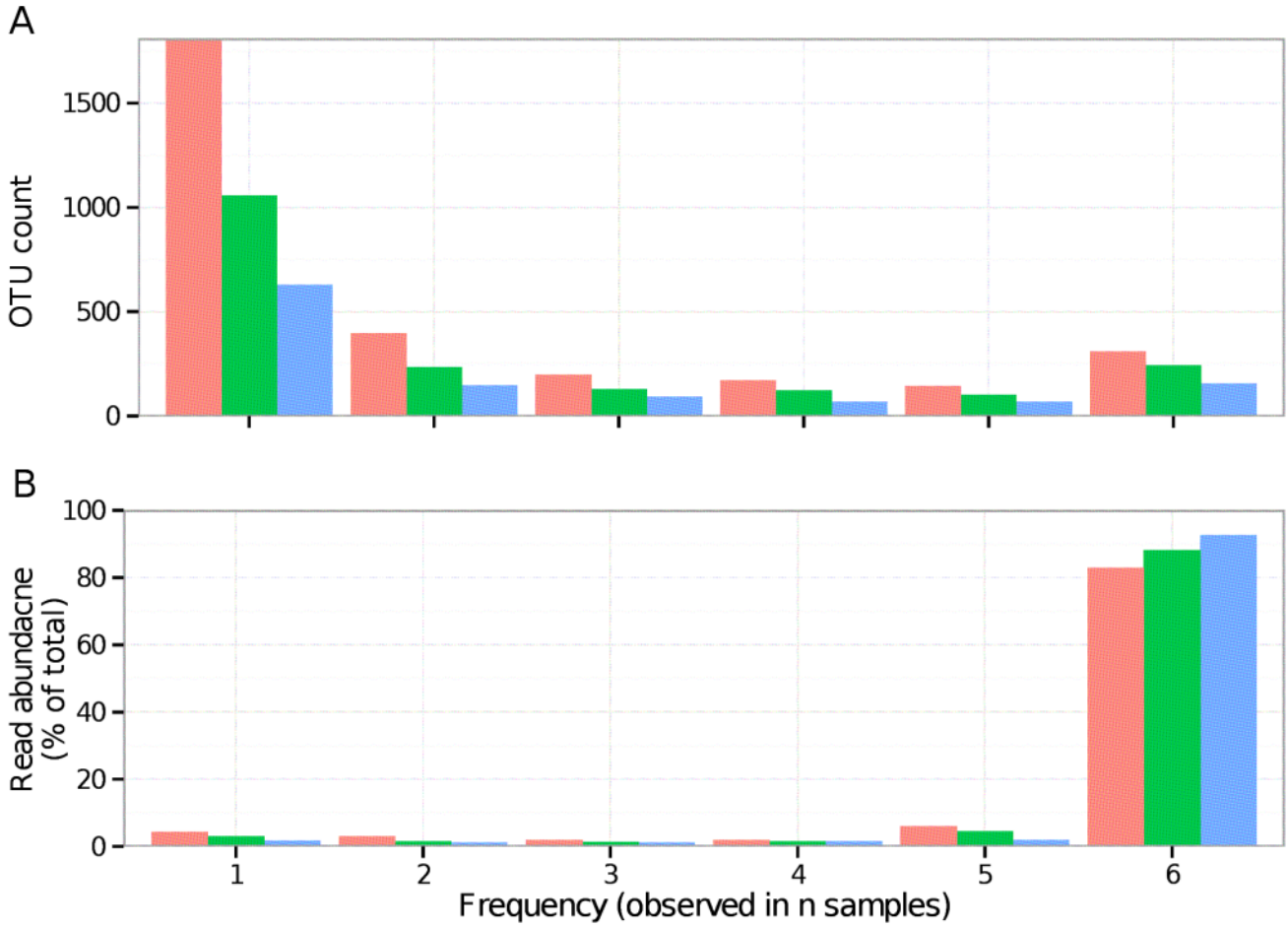

Figure S4: Frequency distribution of OTUs among 2 samples of wastewater influent to Aalborg East, Aalborg West, and Hjørring wastewater treatment plants. The x-axis shows the number of samples in which each OTU is observed, and colors denote OTUs clustered at subspecies-level (red), species-level (green), genus-level (blue). The upper graph (A) shows the number of OTUs observed at each frequency, which demonstrates a core community of abundant OTUs observed in every sample. The lower graph (B) show the % reads for OTUs observed at that frequency (bars) and demonstrates that core OTUs made up a large fraction of the total reads.

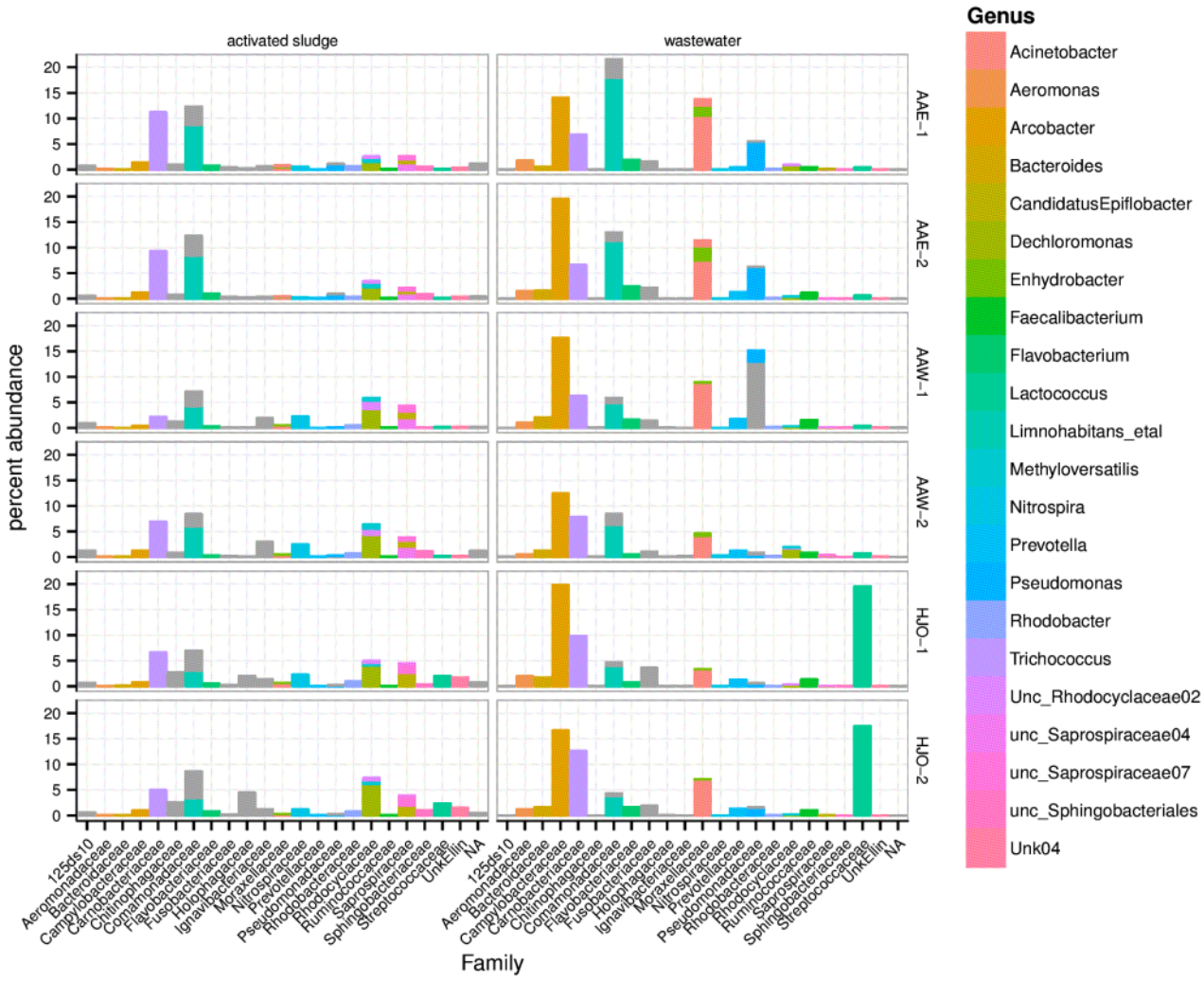

Figure S5: Comparison of the abundance of the top 20 OTUs in the activated sludge and the top 20 OTUs the influent wastewater.

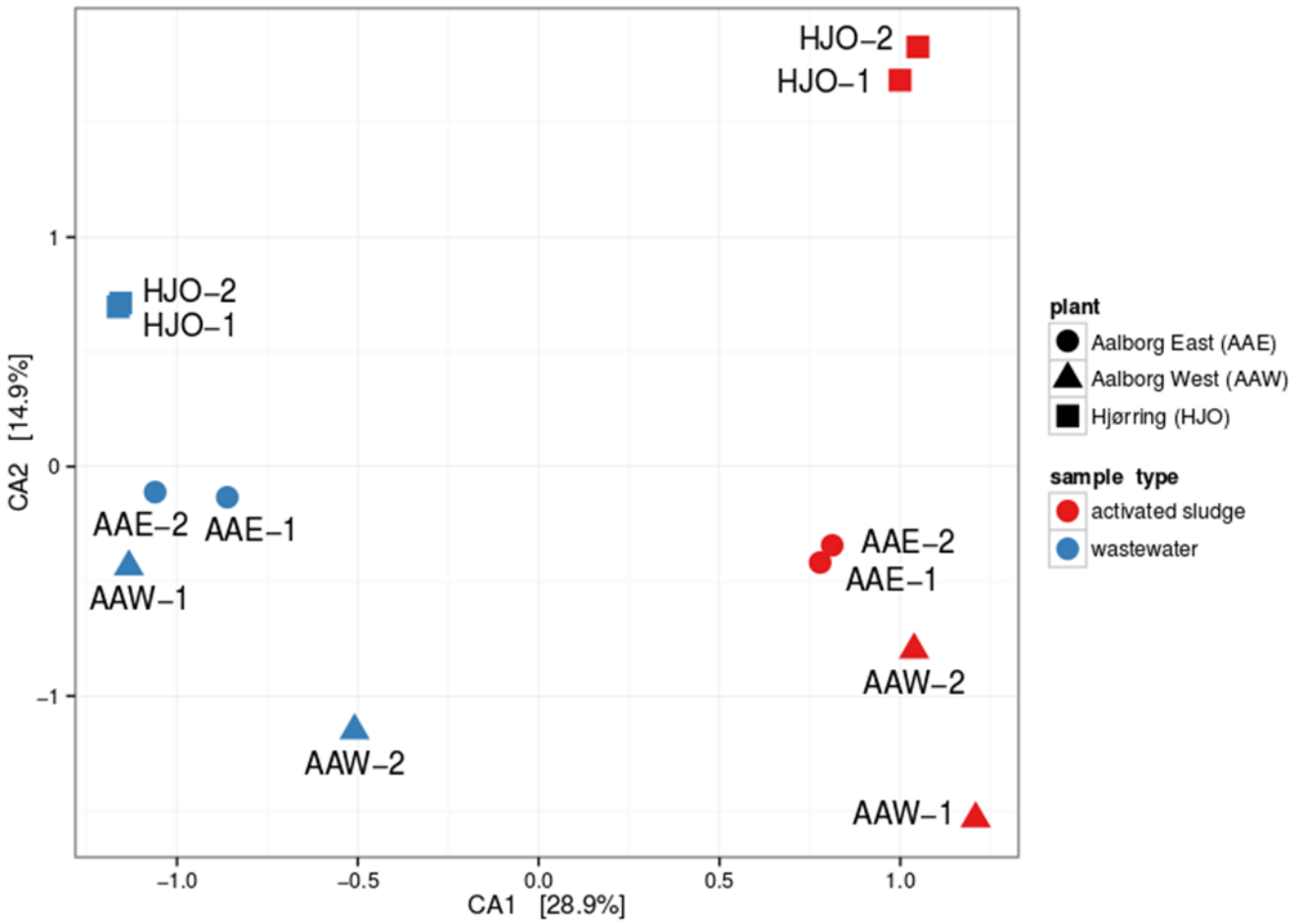

Figure S6: Canonical correspondence analysis of the relative abundance of the genus-level OTUs in the wastewater and activated sludge in three plants at two time points. CA1 separated the wastewater (blue) from the activated sludge (red). CA2 separated the two Aalborg plants (circles and triagles) from the Hjørring plant (squares).

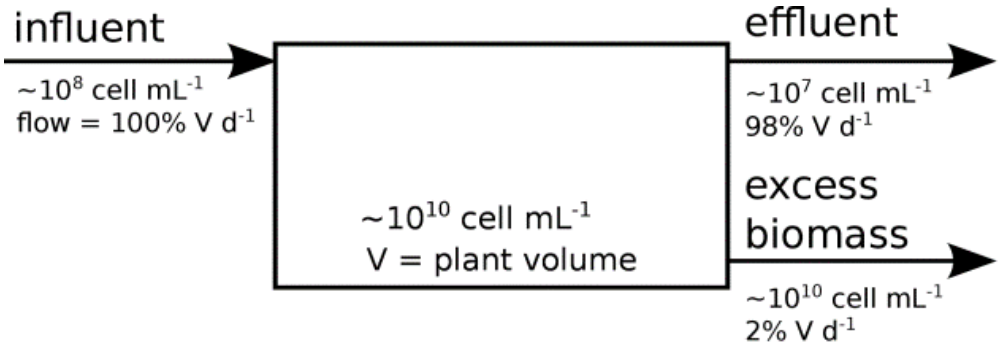

Figure S7: Balance of inflow and outflow for a typical activated sludge, wastewater treatment plant.

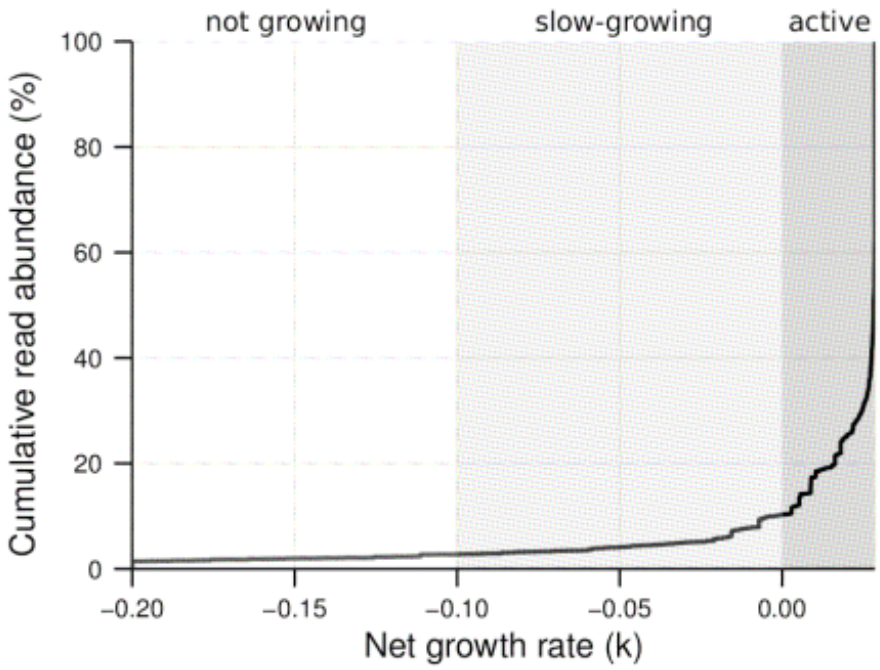

Figure S8: Cumulative distribution of relative read abundance across the net growth rates in the activated sludge. OTUs not detected in the wastewater had the maximum net growth rate (0.029 d^{-1}). Most reads (89%) had a net growth rate greater than zero (dark grey) indicating that their apparent abundance was due to growth within the plant, not from immigration with the wastewater. OTUs that had a net growth rate less than zero (light grey) and less than -0.1 (white) were abundant in the wastewater and their presence was likely due to immigration. OTUs less than -0.20 are presented as -0.20 d^{-1}.

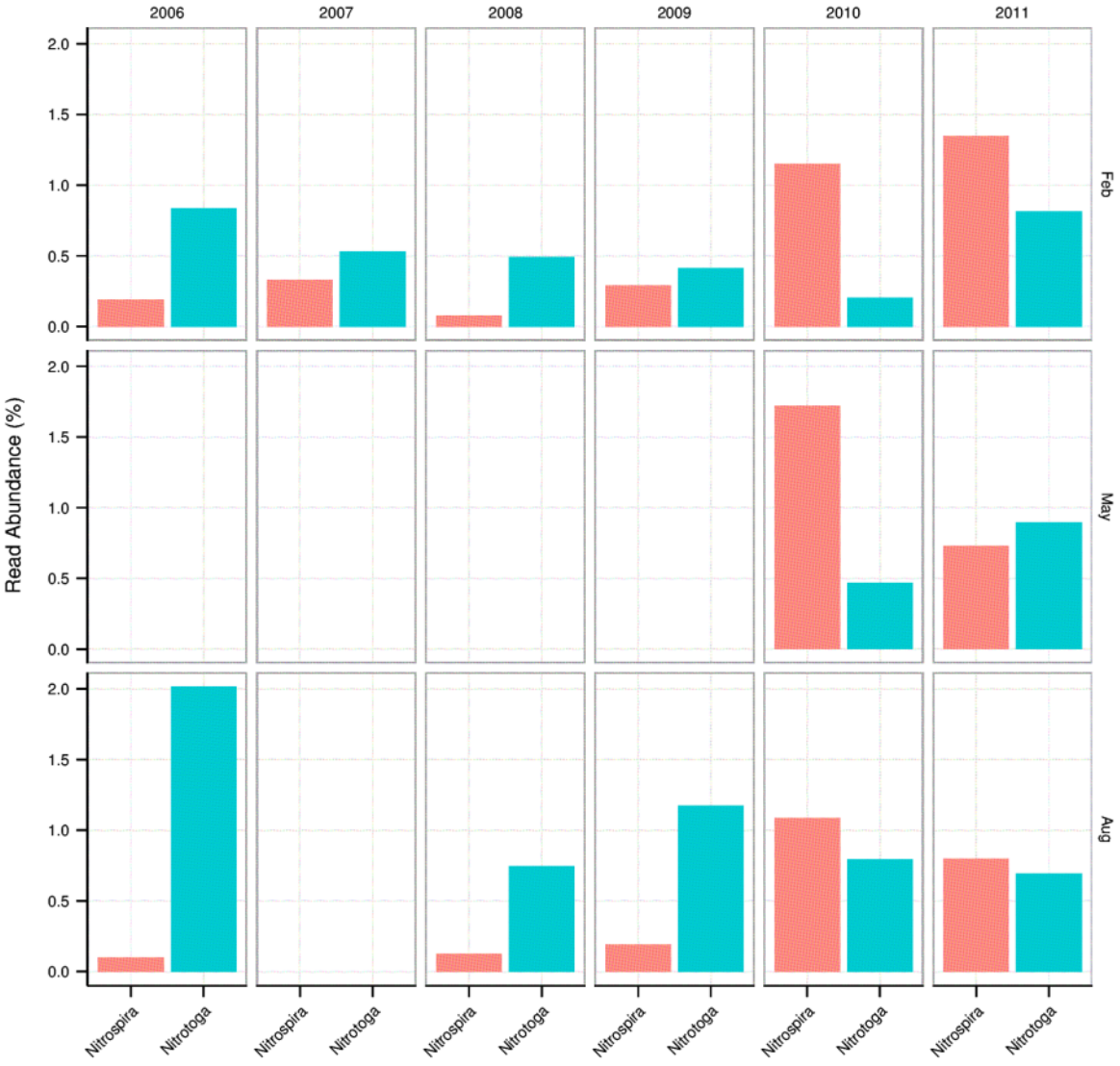

Figure S9: Abundance of nitrite oxidizing bacteria, *Nitrospira* (red) and *Nitrotoga* (blue) in activated sludge from Aalborg West. The samples (n = 13) were taken from 2006-2010 (left to right) in three different months (top to bottom).

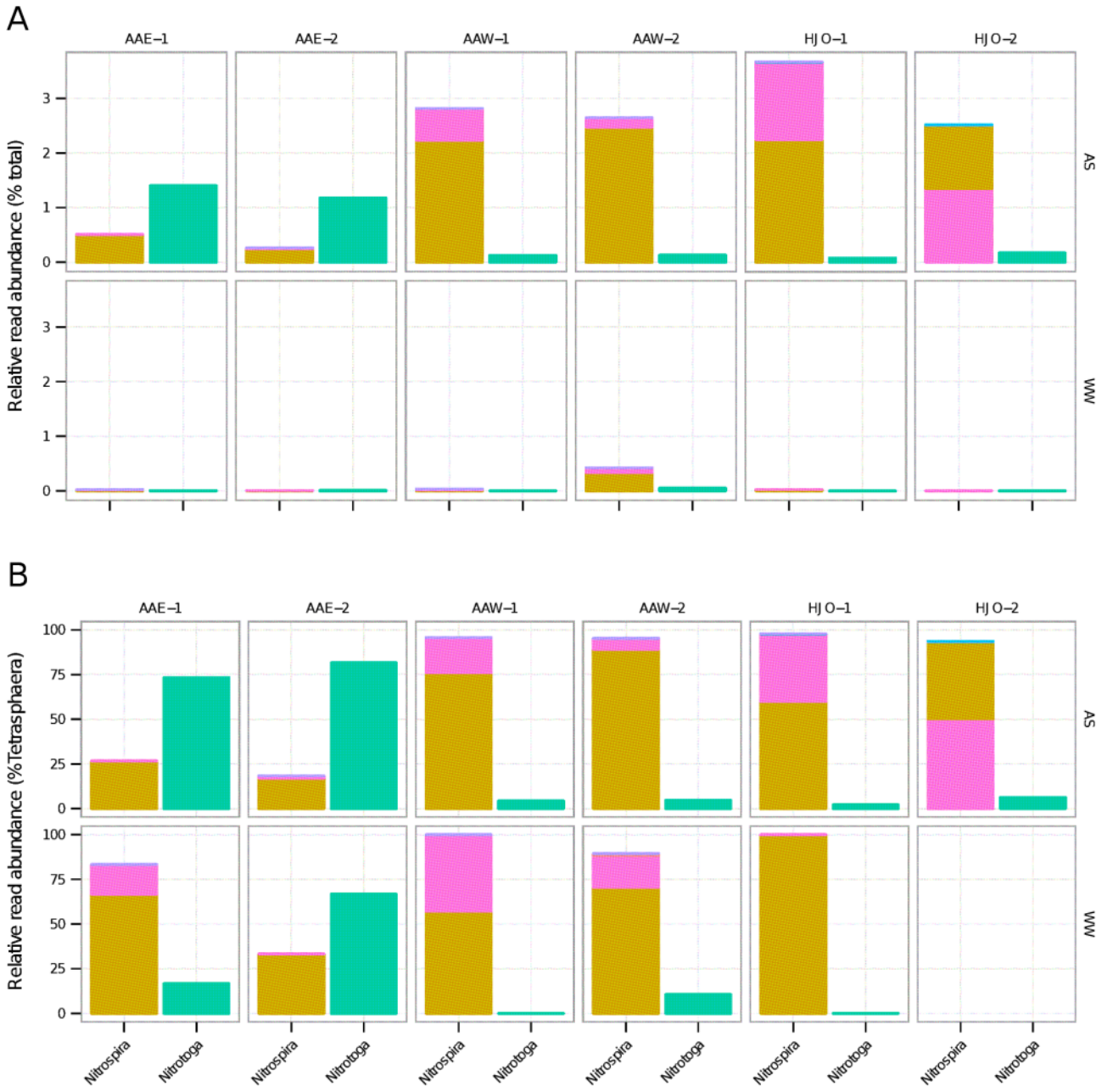

Figure S10: Comparison of the abundance of *Nitrospira* and *Nitrotoga* in the influent wastewater (WW) and in the activated sludge (AS). Two samples were taken from Aalborg East (AAE), Aalborg West (AAW) and Hjørring (HJO) The upper graph (A) shows abundance as a fraction of the total reads and the lower graph (B) shows abundance as a fraction of the total NOB (*Nitrospira* + *Nitrotoga*). Colors on the bars denote distinct sub-species level OTUs.

4. References

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533–538.

Henze M, Harremoës P, la Cour Jansen J, Arvin E (2002) *Wastewater Treatment, 3rd Edition* (Springer, New York).

Kembel SW, Wu M, Eisen J a., Green JL. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput. Biol* **8**: e1002743.

Kristiansen R, Nguyen HTT, Saunders AM, Nielsen JL, Wimmer R, Le VQ, *et al*. (2013). A metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological phosphorus removal. *ISME J* **7**: 543–554.

McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. (2010). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ Microbiol* **12**: 378–392.

Morgan-Sagastume F, Larsen P, Nielsen JL, Nielsen PH (2008) Characterization of the loosely attached fraction of activated sludge bacteria. *Water Res* 42:843–54.

Nielsen JL, Nguyen H, Meyer RL, Nielsen PH. (2012). Identification of glucose-fermenting bacteria in a full-scale enhanced biological phosphorus removal plant by stable isotope probing. *Microbiol* **158**: 1818–1825.

Vollertsen J, Jahn A, Nielsen JL, Hvitved-Jacobsen T, Nielsen PH (2001) Comparison of methods for determination of microbial biomass in wastewater. *Water Res* 35:1649–58.