MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge

Nierychlo, Marta Anna; Andersen, Kasper Skytte; Xu, Yijuan; David Green, Nick; Jiang, Chenjing; Albertsen, Mads; Dueholm, Morten Simonsen; Nielsen, Per Halkjær

Published in:
Water Research

DOI (link to publication from Publisher):
10.1016/j.watres.2020.115955

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Water Research 182 (2020) 115955

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge

Marta Nierychlo, Kasper Skytte Andersen, Yijuan Xu, Nicholas Green, Chenjing Jiang, Mads Albertsen, Morten Simonsen Dueholm, Per Halkjær Nielsen*

Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Article history:
Received 4 February 2020
Received in revised form
13 May 2020
Accepted 16 May 2020
Available online 28 May 2020

Keywords:
Activated sludge
MiDAS
Species composition
Database
Taxonomy

ABSTRACT

The function of the microbiomes in wastewater treatment systems and anaerobic digesters is dictated by the physiological activity of their members and complex interactions between them. Since functional traits are often conserved at low taxonomic ranks (genus, species, strain), high resolution taxonomic classification is crucial to understand the role of microbes in any ecosystem. Here we present MiDAS 3, a comprehensive 16S rRNA gene reference database based on full-length 16S rRNA gene amplicon sequence variants (FL-ASVs) derived from activated sludge and anaerobic digester systems in Denmark. The new database proposes unique provisional names for all unclassified microorganisms down to species level, providing a new and much-needed tool for microbiome research. The MiDAS 3 database was used to analyze the microbiome in 20 Danish wastewater treatment plants with nutrient removal, sampled over 13 years. The 50 most abundant species belonged to 42 genera, including 14 genera with provisional 'midas' name. Of those, 20 have no known function in the system, which highlights the need for more efforts towards elucidating the role of important members of wastewater treatment ecosystems. The new MiDAS 3 database also forms the backbone of the MiDAS Field Guide – an online resource linking the identity of microorganisms in wastewater treatment systems to available data related to their functional importance. The new field guide contains a complete list of genera (>1800) and species (>4200) found in activated sludge and anaerobic digesters in Denmark, but is also relevant to wastewater systems across the world. The identity of the microbes is linked to functional information, where available, and the website provides the possibility to BLAST new sequences against the MiDAS 3 database. The MiDAS Field Guide is a collaborative platform acting as an online knowledge repository, facilitating understanding of wastewater treatment ecosystem function.

1. Introduction

Biological nutrient removal plants are the principal wastewater treatment systems across the urbanized world. Besides their primary role in removing pollutants and pathogens, these systems are increasingly seen as resource recovery facilities, contributing to sustainable resource management (Nielsen, 2017). For example, at-plant anaerobic digestion exploits the value of sludge as a source of energy, reducing plant carbon footprint, and placing wastewater treatment plants (WWTPs) on the road to achieving a circular economy. Complex microbial communities, the microbiome, define the function of biological wastewater treatment systems, and understanding the role of individual microbes requires knowledge of their identity, to which known and putative functions can be assigned. The gold standard for identification and characterization of the diversity, composition, and dynamics of the microbiome is presently 16S rRNA gene amplicon sequencing (Boughner and Singh, 2016). A crucial step is a reliable taxonomic classification, but this is strongly hampered by the lack of sufficient reference sequences in public databases for many important microbes present in wastewater treatment systems (McIlroy et al., 2015). In addition, high taxonomic resolution is missing from the large-scale public reference databases such as SILVA (Quast et al., 2013), Greengenes (DeSantis et al., 2006), and RDP (Cole et al., 2014), often resulting in poor classification.

* Corresponding author. Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark.
E-mail address: phn@bio.aau.dk (P.H. Nielsen).

https://doi.org/10.1016/j.watres.2020.115955

MiDAS (Microbial Database for Activated Sludge) was established in 2015 as an ecosystem-specific database for wastewater treatment systems based on the SILVA database which provided manually curated taxonomic assignment (MiDAS taxonomy 1.0) and associated physiological information profiles (midasfieldguide.org) for all abundant and process-critical genera in activated sludge (AS) (McIlroy et al., 2015). Both taxonomy and database were later updated (MiDAS 2.0) to cover abundant microorganisms found in anaerobic digesters (AD) and influent wastewater (McIlroy et al., 2017). MiDAS 2.0 can only provide taxonomic classification down to genus level due to the resolution of the SILVA database, and classification to higher taxonomic ranks (family, order) is often observed because proper reference sequences are lacking. This limits elucidation of the diversity in these ecosystems. Consequently, physiological differences between the members of the same genus that coexist in activated sludge cannot be reliably linked to the individual phylotypes due to insufficient phylogenetic resolution. Furthermore, the growing number of sequences and frequent taxonomy updates of large-scale databases, such as SILVA (Glöckner et al., 2017), make manual curation difficult to maintain.

Recent developments of sequencing technology and bioinformatic tools have enabled generation of millions of high-quality full-length 16S rRNA gene reference sequences from any environmental ecosystem (Callahan et al., 2019; Karst et al., 2019, 2018). By obtaining full-length 16S rRNA gene sequences directly from environmental samples, we have established a near-complete reference database for high-taxonomic resolution studies of the wastewater treatment ecosystem. A Approx. one million 16S rRNA gene sequences were retrieved from the wastewater treatment plants, providing more than 9500 full-length 16S rRNA gene amplicon sequence variants (FL-ASVs) after dereplication and error correction (Dueholm et al., 2019). Moreover, we created the AutoTax pipeline for generating a comprehensive ecosystem-specific taxonomic database with *de novo* names for novel taxa. The names are first inherited from sequences present in large-scale SILVA taxonomy, and *de novo* names are provided for all remaining unclassified taxa, based on reproducible clustering with rank-specific identity thresholds (Yarza et al., 2014). These *de novo* names provide placeholder names for all novel phylotypes and act as fixed unique identifiers, making the taxonomic assignment independent of the analyzed data set and enabling cross-study comparisons.

Here we present the MiDAS 3 reference database, which is based on the FL-ASVs previously obtained from activated sludge and anaerobic digester systems (Dueholm et al., 2019), amended with additional sequences of potential importance in the field. These sequences represent bacteria from influent wastewater, potential pathogens, and genome-derived sequences of microbes found in wastewater treatment systems. The MiDAS 3 taxonomy is built using AutoTax (Dueholm et al., 2019) and proposes unique provisional names for all microorganisms important in wastewater treatment ecosystems at species-level resolution. Compared to MiDAS 2.0, which was a manually curated version of SILVA database providing classifications down to genus-level, MiDAS 3 provides a more sustainable and streamlined workflow for taxonomy assignment. The database contains only high-quality sequences coming from the ecosystem, where the existing taxonomic classification, down to species-level, is adapted from SILVA or automatically assigned using the novel AutoTax concept. In order to automate the procedure, the majority of manual curations present in MiDAS 2.0 are not maintained in MiDAS 3, thus, a minor number of changes in taxonomic classification of the abundant microorganisms present in the ecosystem are observed upon the comparison of the two MiDAS versions. We have furthermore curated some of the species-level names, and here we demonstrate the resolution power of MiDAS 3 by analyzing species-level microbiome in 20 Danish full-scale WWTPs carrying out nutrient removal with chemical phosphorus removal (BNR) or enhanced biological phosphorus removal (EBPR) over 13 years. We provide information about diversity, core communities, and the most abundant species in these ecosystems. Although the plants investigated here are Danish, the results and approach are relevant for studies of biological nutrient removal WWTPs worldwide, as previously demonstrated for Dutch and Canadian WWTPs (Dueholm et al., 2019).

We also present the updated MiDAS Field Guide, an online platform that links the MiDAS 3 taxonomy-derived identity of all microbes from wastewater treatment systems to abundance information, based on our long-term survey, and functional information (where available). MiDAS 3 alleviates important problems related to the taxonomic classification of microbes in environmental ecosystems, such as missing reference sequences and low taxonomic resolution, while our MiDAS Field Guide acts as an online knowledge repository facilitating understanding the function of microbes present in wastewater treatment ecosystems.

2. Materials and methods

2.1. MiDAS 3 reference database and taxonomy

The MiDAS 3 reference database was built using 16S rRNA gene FL-ASVs from 21 activated sludge WWTPs and 16 ADs systems, and taxonomic classification was assigned as described by Dueholm et al. (2019) and shown in Fig. 1. The database was amended with 95 additional sequences from published genomes for 60 bacteria and 35 archaea that are potentially important in wastewater treatment systems, but not present in the reference FL-ASVs obtained (Table S2). Taxonomy was assigned to all sequences using the workflow presented in Fig. 1. Gaps in the SILVA-derived taxonomy were filled with a *de novo* taxonomy which provides 'midas_x_y' placeholder names, where x represents the first letter of taxonomic level, for which *de novo* assignment was created, and y is a number derived from the FL-ASV that forms the cluster centroid for a given taxon. Where applicable, placeholder names were replaced with taxon names for FL-ASVs that represent genomes published in peer-reviewed literature, which were not included in the SILVA taxonomy. Additionally, several manual curations of the SILVA-derived taxonomy were made, based on published literature (full list of amendments is presented in Table S3).

2.2. Activated sludge samples

Sampling of the activated sludge biomass from 20 Danish full-scale municipal WWTPs was carried out within the MiDAS project (McIlroy et al., 2015). The plants were sampled up to four times a year (February, May, August, and October) from 2006 to 2018 with a total of 712 samples. All samples were taken from the aeration tank and sent overnight to Aalborg University for processing. Basic information about the design and operation of the WWTPs as well as number of samples collected are given in Table S1. WWTPs included in this study comprised 16 EBPR plants, 3 BNR plants, and 1 plant that changed from BNR to EBPR during the survey. All EBPR plants received occasional dosage of iron or aluminum salts to improve phosphorus (P) precipitation and enhance flocculation. All WWTPs had stable operation with effluent total P concentration below the required limit (1.0 mgP/L) with average effluent P concentration across all EBPR plants of 0.39 ± 0.37 mgP/L. Effluent total N and total COD concentrations were 4.59 ± 2.44 and 27.4 ± 12.3, respectively.

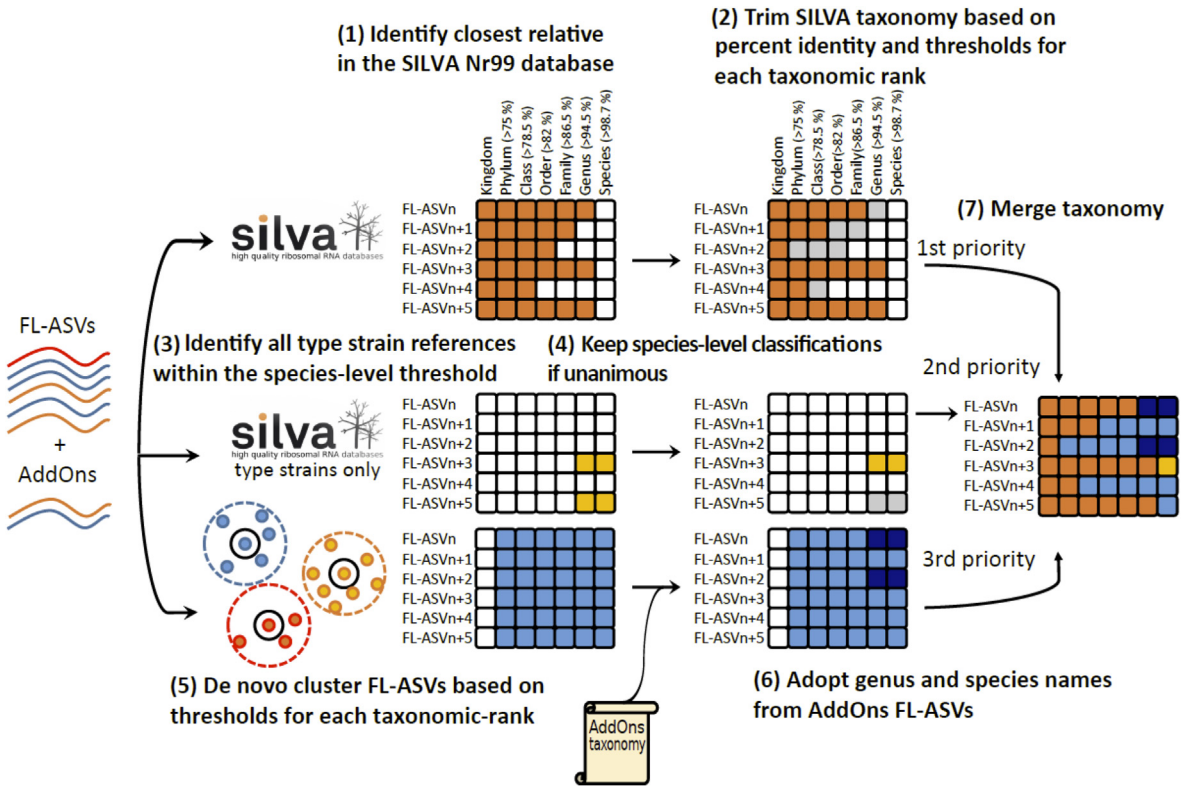

Fig. 1. MiDAS 3 taxonomic assignment workflow based on the AutoTax framework. (1) FL-ASVs from the study of Danish plants (Dueholm et al., 2019), amended with additional FL-ASVs (AddOns) coming from, e.g., published genomes (listed in Table S2), are first mapped to the SILVA_132_SSURef_Nr99 database to identify their closest relative. (2) Taxonomy is adopted from the closest relative down to the taxonomic rank that is supported by the sequence identity thresholds proposed by Yarza et al. (2014), shown in brackets. To provide species-level taxonomy, (3) FL-ASVs are mapped to sequences from type strains extracted from the SILVA database, and (4) species names are adopted if they share >98.7% identity with a single type strain. (5) FL-ASVs are also clustered by greedy clustering at different identities, corresponding to the thresholds proposed by Yarza et al. (2014) to generate a stable *de novo* taxonomy. (6) The *de novo* taxonomy is curated with taxon names for AddOns FL-ASVs that represent, e.g., genomes from strains published in peer-reviewed literature (Table S2). (7) Finally, a comprehensive taxonomy is obtained by filling gaps in the SILVA-based taxonomy with the *de novo* taxonomy with provisional "midas" placeholder names. Colored squares represent sources of taxonomic classifications of FL-ASVs present in MiDAS 3 database: orange: SILVA Nr99, yellow: SILVA type strains, light blue: de novo names, dark blue: additional sequences (AddOns) introduced to MiDAS 3 database, gray: classifications rejected during the Auto-Tax workflow (adapted from Dueholm et al., 2019).

2.3. Amplicon sequencing and bioinformatic analysis

DNA extraction, sample preparation, including amplification of V1-3 region of 16S rRNA gene using the 27F (AGAGTTTGATCCTGGCTCAG) (Lane, 1991) and 534R (ATTACCGCGGCTGCTGG) (Muyzer et al., 1993) primers, and amplicon sequencing were conducted as described by Stokholm-Bjerregaard et al. (2017). The V1-3 region was chosen for activated sludge community analysis, based on the studies by Albertsen et al. (2015) and Dueholm et al. (2019), showing this primer to give the most representative community structure and the highest taxonomic resolution. Forward reads were processed using usearch v.11.0.667 (Edgar, 2010). Raw fastq files were filtered for phiX sequences using usearch -filter_phix, trimmed to 250 bp using usearch -fastx_truncate -trunclen 250, and quality filtered using usearch -fastq_filter with -fastq_maxee 1.0. The sequences were dereplicated using usearch -fastx_uniques with -sizeout. Exact amplicon sequence variants (ASVs) were generated using -unoise3 (Edgar, 2016) with standard settings, and taxonomy was assigned using the MiDAS 3 reference database (available at https://www.midasfieldguide.org/guide/downloads) and the SINTAX classifier with a confidence threshold of 0.8 (Edgar, 2018).

2.4. Data analysis and visualization

R v.3.5.1 (R Core Team, 2018), RStudio (RStudio Team, 2015), and QIIME v.1.9.0 (Caporaso et al., 2010) were used for downstream data processing. 712 samples with minimum 13,500 reads (the cutoff value was chosen based on read histogram analysis) were normalized and analyzed using ggplot2 v.3.2.1 (Wickham, 2009) and ampvis2 v.2.4.9 (Andersen et al., 2018a). For alpha- and beta-diversity analysis samples were rarefied to 13,500 reads. Alpha-diversity was calculated in ampvis2. Beta-diversity analysis was performed using QIIME's *beta_diversity.py* script for calculating UniFrac distance (Lozupone and Knight, 2005) and visualized by principal coordinate analysis (PCoA) using ampvis2. Based on UniFrac distance matrix, statistically significant differences among WWTPs were calculated using the ANOSIM (QIIME) and Dunn's post hoc test with Bonferroni correction (dunn.test v.1.3.5, Dinno, 2017). The core communities were defined as ASV-level taxa that are 1) observed in all the plants and 2) abundant (belonging to top 80% of the reads) in all the plants. The average abundance of each ASV in all samples from given plant was summed and divided by total abundance of all ASVs in that plant. Cumulative ASV read abundance was calculated for each plant, and the ASVs comprising top 80% reads

were considered abundant. ASVs were grouped according to the number of plants in which they were observed, as well as number of plants in which they were abundant. Sequencing data is available at the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under the project number PRJNA622675. R codes and metadata files are available at https://github.com/martanierychlo/MiDAS3.

3. Results and discussion

3.1. The MiDAS reference database provides species-level classification

In the creation of the MiDAS 3 database, we have applied a novel approach, described in Fig. 1, where high-throughput sequencing of high-quality full-length 16S rRNA genes (Karst et al., 2018) was combined with a new automatic taxonomy assignment concept, AutoTax (Dueholm et al., 2019). The database contains full-length sequences from 21 Danish WWTPs carrying out nutrient removal and 16 mesophilic and thermophilic Danish ADs located at wastewater treatment plants. The MiDAS 3 reference database was additionally supplemented with high-quality full-length 16S rRNA gene sequences (AddOns) of microbes with known importance in the field, as well as bacteria present in the influent wastewater and pathogens (Table S2).

MiDAS 3 taxonomic assignment is based on the AutoTax framework, where sequences first inherit taxonomic classification from the SILVA SSURef Nr99 database (release 132, Quast et al., 2013), including species names (based on type strains), if available. The remaining unclassified sequences are assigned unique MiDAS names down to species level. Thus, the new release of MiDAS 3 taxonomy proposes a provisional genus and species name for all microorganisms found in activated sludge and anaerobic digester microbiomes, acting as stable identifiers and placeholder names for the sequences, until the representative microorganisms are further characterized and assigned approved names. Additionally, names derived from recently published genomes and other sources not yet incorporated in SILVA taxonomy were adopted in MiDAS 3 (see Table S3 for full list of curated names) to provide a near-complete database and taxonomy for microorganisms in wastewater treatment ecosystems. For a few species, there were inconsistencies between the genus and type strain name assigned by SILVA (see Table S4); these issues relate to the fact that the SILVA taxonomy is updated based on phylogenetic analyses of the available 16S rRNA gene sequences, whereas type strain names have to be approved by the International Taxonomy Committee. The possibility of including additional sequences ensures that the database can easily be updated based on the current state of the ecosystem-specific field, and will drive the future releases of the MiDAS database.

Two examples of improved taxonomy are shown in Table 1. *Acidovorax* is one of the most abundant genera in Danish influent wastewater (McIlroy et al., 2017), with one abundant species identified using the MiDAS 3 reference database. Two ASVs belonging to the species midas_s_2077 are classified down to genus level in all the reference databases tested, except Greengenes, which failed to provide genus classification for both ASVs or even family classification for ASV17. Genus *Ca.* Amarolinea contains abundant filamentous bacteria in Danish WWTPs (Nierychlo et al., 2019) and can be associated with serious bulking incidents (Andersen et al., 2018b; Nierychlo and Nielsen, 2017). All three abundant species belonging to the genus lack taxonomic classification below order level in public large-scale databases, preventing identification and analysis of this important member of the activated sludge system. The overview of the top 50 most abundant ASVs in Danish WWTPs with their taxonomic classification based on MiDAS 3 compared to SILVA is shown in Fig. S1.

3.2. Microbiome composition of Danish activated sludge plants

In this study, we performed a 13-year survey of the microbiome in 20 Danish full-scale WWTPs with nutrient removal that primarily treated municipal wastewater. All plants had biological nitrogen removal, and most also had EBPR (16 plants), 3 plants had chemical P-removal (BNR plants), while 1 changed from BNR to EBPR (Table S1). The survey was carried out using 16S rRNA gene sequencing of the V1-3 region on the Illumina MiSeq platform. Exact amplicon sequence variants (ASVs) were generated and analyzed here instead of operational taxonomic units (OTUs) in order to avoid similarity-based clustering and to reflect the true biological diversity in the samples (Callahan et al., 2017). ASVs provide consistent labels used for identification of microbes, however, they can only be compared across studies when generated in the same way, since their numbering is dependent on the data set and processing steps. MiDAS 3 taxonomy providing species names ensures deeper resolution and robust classification that allows inter-study comparisons. Using MiDAS 3 taxonomy, the microbiome was characterized at genus and species level for all plants, based on 712 samples collected 2 to 4 times per year for each plant from 2006 to 2018. The data set consisted of 75,017 ASVs that were represented by 1693 genera and 3787 species. Of the total ASV number, 49% possessed genus-level classification and 30% species-level classification. The presence of unclassified ASVs was primarily due to missing FL-ASV reference sequences for rare ASVs or the presence of highly similar 16S rRNA genes in different species of the same genus. This is especially the case for medically relevant species, which are often defined based on physiological traits rather than molecular phylogeny. An example of the latter is the genus *Trichococcus*, where the similarity of the 16S rRNA genes between species is between 99 and 100% (Strepis et al., 2016). A blast search at the MiDAS field guide can be used to evaluate whether high similarity of 16S rRNA gene within the same genus causes lack of classification for a specific ASV because several species will be hit by the ASV.

The overall microbial diversity within samples was estimated for each plant at ASV level using the Simpson index of diversity, and compared for all plants analyzed (Fig. 2). The Simpson index places greater weight on species' evenness than the richness (Kim et al., 2017), thus allowing the assessment of how evenly distributed relative taxa contributions are across the samples. Since the value of the index represents the probability that two randomly selected individuals belong to the same taxa, the low values of Simpson index of diversity (Fig. 2) indicate that the microbial communities were highly diverse in all the Danish plants. Additionally, the narrow spread of the index in each plant indicates that the microbial communities across the plants were consistently stable across the years. Similar alpha diversity indices were observed in WWTPs across the world by Wu et al. (2019), calculated as Inverse Simpson (Fig. S2).

The differences in overall activated sludge microbiome in 20 Danish WWTPs were assessed using the phylogeny-based metric UniFrac and visualized using PCoA plot (Fig. 3a). Small UniFrac distance indicates similarity between the communities composed of taxa that share a common evolutionary history. Samples belonging to different WWTPs were not completely separated from each other, suggesting high similarity between the microbial communities in those WWTPs. This is also highlighted by the fact that the two principal coordinates only explain a small fraction of variation in the data (5.6% and 5.4%). Fredericia appeared to have a more distinct microbiome composition, presumably due to a high fraction of industrial waste in the influent (Table S1). Noticeable clustering of the samples coming from individual WWTPs (Fig. 3a), as well as analysis of unweighed UniFrac distances within and

Table 1
Classification of abundant species belonging to genera *Acidovorax* and *Ca.* Amarolinea with different taxonomies.

Taxonomy	Phylum	Class	Order	Family	Genus	Species	ASV
MiDAS 3	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Acidovorax*	midas_s_2077	ASV17
MiDAS 2.1	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	-	–	ASV17
SILVA	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Acidovorax*	–	ASV17
Greengenes	Proteobacteria	Betaproteobacteria	Burkholderiales	–	–	–	ASV17
RDP	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax*	–	ASV17
MiDAS 3	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Acidovorax*	midas_s_2077	ASV1167
MiDAS 2.1	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax*	–	ASV1167
SILVA	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Acidovorax*	–	ASV1167
Greengenes	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	–	–	ASV1167
RDP	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax*	–	ASV1167
MiDAS 3	Chloroflexi	Anaerolineae	Caldilineales	Amarolineaceae	*Ca.* Amarolinea	midas_s_1	ASV3
MiDAS 2.1	Chloroflexi	SJA-15	C10_SB1A	C10_SB1A	*Ca.* Amarilinum	–	ASV3
SILVA	Chloroflexi	Anaerolineae	C10_SB1A	–	–	–	ASV3
Greengenes	Chloroflexi	–	–	–	–	–	ASV3
RDP	-	–	–	–	–	–	ASV3
MiDAS 3	Chloroflexi	Anaerolineae	Caldilineales	Amarolineaceae	*Ca.* Amarolinea	midas_s_553	ASV324
MiDAS 2.1	Chloroflexi	SJA-15	C10_SB1A	C10_SB1A	*Ca.* Amarilinum	–	ASV324
SILVA	–	–	–	–	–	–	ASV324
Greengenes	Chloroflexi	Anaerolineae	SHA-20	–	–	–	ASV324
RDP	-	–	–	–	–	–	ASV324
MiDAS 3	Chloroflexi	Anaerolineae	Caldilineales	Amarolineaceae	*Ca.* Amarolinea	*Ca.* A. aalborgensis	ASV131
MiDAS 2.1	Chloroflexi	SJA-15	C10_SB1A	C10_SB1A	*Ca.* Amarilinum	–	ASV131
SILVA	–	–	–	–	–	–	ASV131
Greengenes	Chloroflexi	Anaerolineae	SHA-20	–	–	–	ASV131
RDP	–	–	–	–	–	–	ASV131

MiDAS 2.1 (Mcllroy et al., 2017); SILVA: Release 132_SSURef_NR99 (Quast et al., 2013); Greengenes: Release 16s_13.5 (DeSantis et al., 2006); Ribosomal Database Project: Release 16S_v16 training set (Cole et al., 2014).

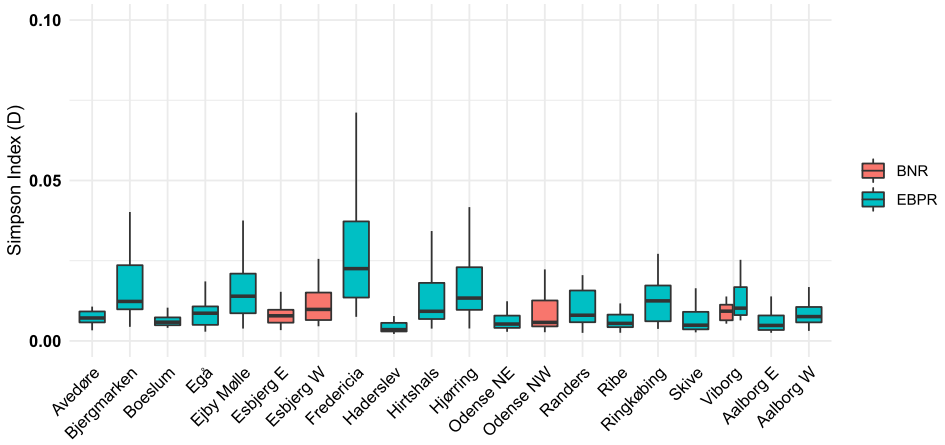

Fig. 2. Alpha diversity estimates of activated sludge microbiome using ASV-level taxa in individual plants, calculated using the Simpson index of diversity (D). Data represents 712 samples from 20 Danish full-scale WWTPs collected from 2006 to 2018 (with 17-51 samples per plant).

between WWTPs (Fig. 3b), showed that variation between the plants was higher than within individual plants over time, which suggests long-term stability of the microbial communities in Danish plants with nutrient removal. Visualization of time series data (see example Fig. S3) confirmed the stability of microbiome in individual plants with gradual fluctuation of abundance for only a few species. Fredericia WWTP demonstrated the greatest range of Unifrac distances, indicating the microbial community was less stable than in the remaining plants. Ribe had the lowest Unifrac distances within the plant, and second highest (after Fredericia) distances among the plants, suggesting highly stable and distinct microbial community present. The separation of the microbial community observed in that plant (Fig. 3a and b) most probably results from a unique composition of the nitrifying community (results not shown). Differences in community composition were analyzed statistically using Dunn's post hoc test (Fig. 3c), Fredericia and Ribe were the only plants with microbial community significantly different from all plants, whereas the community composition in all other plants was significantly different from more than half of the remaining plants. Analysis of weighed UniFrac distances (Fig. S4) showed the same trend with variation between the plants higher than within individual plants, however, those differences were less pronounced, indicating that although the total microbial community composition is different in many plants, the abundant members of these communities make them more similar to each other.

The microbiomes in WWTPs were composed of 1694 genera and 3787 species, including 1191 *de novo* genera and 3603 *de novo* species possessing midas_x_y names. A relatively small fraction of taxa comprised a large proportion of the reads, thus representing the abundant taxa that are assumed to carry out the main functions in the system (Fig. 4a and b) (Saunders et al., 2016). The abundant taxa were defined by ASVs that constitute the top 80% of reads in individual WWTPs obtained by amplicon sequencing. These

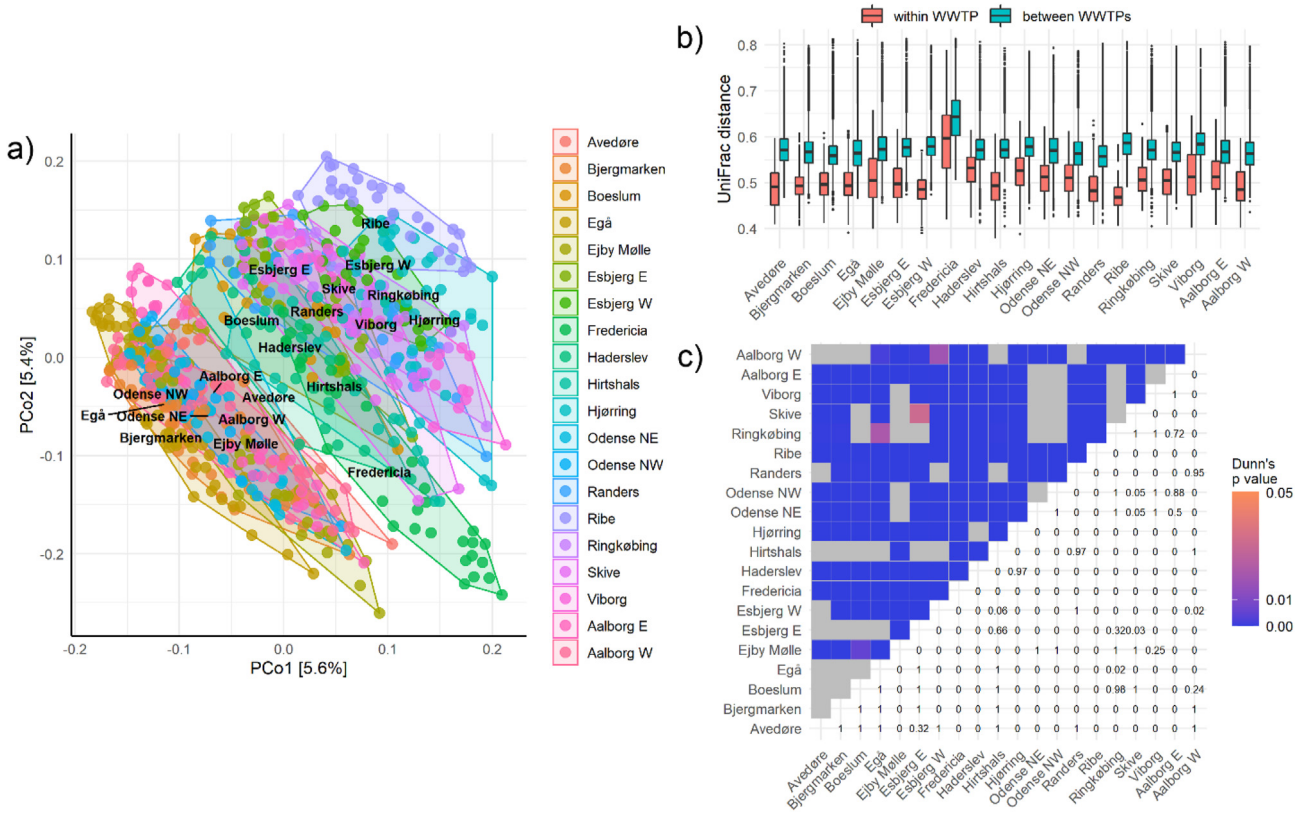

Fig. 3. a) Principal Coordinate Analysis (PCoA) showing clustering of activated sludge microbiomes in WWTPs based on unweighted UniFrac distance matrix. Axes show the percent of total variation described by each component. Statistical analysis revealed significant separation between the plants (ANOSIM, P = 0.001). Data represents 712 samples from 20 Danish full-scale WWTPs collected from 2006 to 2018, b) unweighed UniFrac distance values within each WWTP and between WWTPs, c) p-value matrix showing statistical significance of the differences among the UniFrac distances for individual WWTPs using Dunn's test with a Bonferroni correction for multiple comparisons. Gray values represent p-value > 0.05.

abundant ASVs belonged to 491 and 410 genera, and 945 and 720 species in EBPR and BNR plants, respectively (Fig. 4a and b). Of the top 80% reads in EBPR plants, only 2.9% had no genus-level classification, while 13.2% had no species-level classification. Similar numbers were observed in BNR plants.

To our knowledge, differences in the microbial community structure between EBPR and BNR plants have not been investigated. It is generally assumed that the introduction of an anaerobic stage, required for the EBPR process, will select for polyphosphate-accumulating organisms (PAO), but whether other microorganisms are also affected remains unknown. In this study, we have compared microbiomes of those two systems, however, the low number of BNR plants included may underestimate the diversity observed there. The overlap of taxa between the BNR and EBPR abundant (top 80% of the reads) communities was high (Fig. 4c). The core taxa (93 ASV-level taxa in EBPR plants and 149 ASV-level taxa in BNR plants), defined here as being present and abundant in all the plants, constituted 25% of the total reads in EBPR plants (Fig. 5) and more than 35% of total reads in BNR plants (Fig. S5). These core taxa, which belonged to 65 genera and 63 species in EBPR plants (see the full list of genus-level core taxa in Fig. S6) and to 95 genera and 99 species in BNR plants, constituted only a small fraction of the total number of genera (3.8% and 5.6% for EBPR and BNR plants, respectively) and species (1.7% and 2.6% for EBPR and BNR plants, respectively) present. The number of genera in EBPR plants is higher than the number of species because a number of ASVs possessed genus- but not species-level classification. The number of genera in BNR plants is lower than the number of species because of a number of core ASVs belonging to the same genus, but different species. The core communities in EBPR and BNR plants included taxa known to perform important processes in WWTP: nitrifying *Nitrospira* as well as a number of known filamentous genera: *Ca.* Microthrix, *Ca.* Sarcinithrix, and *Ca.* Villigracilis. Interestingly, ammonium oxidizing *Nitrosomonas* was found in the core community of BNR, but not EBPR plants. This may be due to the fact that several of the EBPR plants included had a high proportion of comammox *Nitrospira*, which is responsible for most of the nitrogen transformation, with simultaneous very low abundance of *Nitrosomonas* (results not shown). Among known and putative PAO, the genera *Tetrasphaera* and *Dechloromonas* were present in core communities in both types of plant, suggesting their versatile role in the ecosystem, while *Ca.* Accumulibacter was only found among core taxa in EBPR plants. Taxa belonging to genera *Nitrospira, Dechloromonas, Sulfuritalea, Rhodoferax,* and *Zoogloea*, known to be involved in nitrogen transformations in WWTPs, were also found to belong to the core community in WWTPs worldwide (Wu et al., 2019). Both plant types contained a number of unique genus-level core taxa (Fig. S6); however, those differences may result from the low number of BNR plants included in the analysis, so we cannot conclude that there are major differences in the microbiome of the two process types. Interestingly, 25 and 41 *de novo* genera with 'midas' names were observed in the core communities for the EBPR and BNR plants, respectively. Some of these novel genera may have important functions, e.g., novel PAOs, thus the definition of the core community helps to guide the priority list for further functional studies.

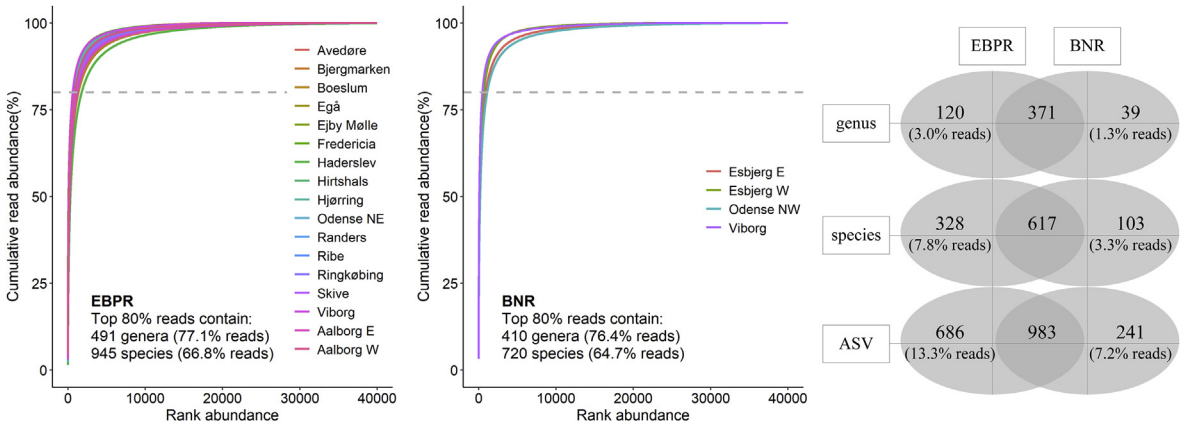

Fig. 4. Cumulative read abundance of ASV-level taxa found in a) EBPR (n = 17) and b) BNR (n = 4) WWTPs, plotted in rank order. Dashed line denotes 80% of the reads, which is assumed to cover the abundant taxa; the numbers of classified genera and species among the top 80% reads and their corresponding cumulative read abundance are given (calculated as average for all EBPR and BNR plants), c) Venn-diagram showing abundant (top 80% of the reads) taxa belonging to number of genera, species, and ASVs shared between BNR and EBPR plants.

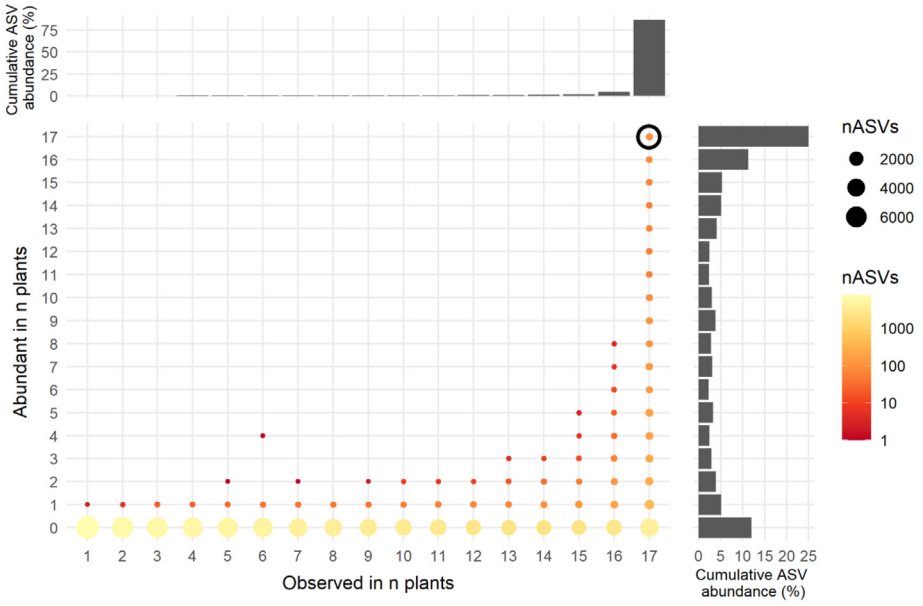

Fig. 5. Distribution of ASV-level taxa across Danish EBPR plants. ASVs are grouped based on the number of plants in which they are observed (x axis) and number of plants in which they are abundant (y axis). Number of ASVs belonging to each group is represented by color and circle size. The core community is defined as present and abundant in all EBPR plants analyzed and is marked with black, empty circle. Data represents 601 samples from 17 EBPR plants. Cumulative abundance of ASVs observed in n plants and abundant in n plants are given in upper and right margin plot, respectively.

3.3. *Species-level community composition*

The most abundant species present in the EBPR and BNR plants are shown in Fig. 6. Among the 10 most abundant species, two belong to genera with a provisional 'midas_g' name, while of the top 50 and top 100 species, 14 and 39 species, respectively, belong to genera not possessing genus-level classification in SILVA (top 100 species are presented in Fig. S7). These would be impossible to identify if MiDAS 3 was not used as the reference database, demonstrating the advantage of using the ecosystem-specific taxonomy.

A large fraction of the most abundant species is poorly described in the literature. The top 50 species belong to 42 different genera, of which almost 50% (20 genera) have no known function in wastewater treatment systems (McIlroy et al., 2017, 2015), demonstrating the need for more studies that link identity and function. A few species, however, match SILVA type strains, such as *Faecalibacterium prausnitzii* and *Intestinibacter bartletti*, so these may be applied for relevant pure-culture studies of their physiology.

The most abundant species in both BNR and EBPR plants belong to the genus *Tetrasphaera,* which represents PAO with versatile metabolism, including the potential for denitrification (Herbst et al., 2019; Kristiansen et al., 2013). Other abundant species belong to the filamentous genera *Ca.* Microthrix (McIlroy et al., 2013) and *Ca.* Villigracilis (Nierychlo et al., 2019); the former known to cause serious bulking problems (Rossetti et al., 2005). *Trichococcus* may also possess filamentous morphology in activated sludge and may be implicated in bulking (Liu and Seviour, 2001). *Rhodoferax* is a denitrifier (McIlroy et al., 2014), while the function in activated sludge is unknown for species in the *Romboutsia* and

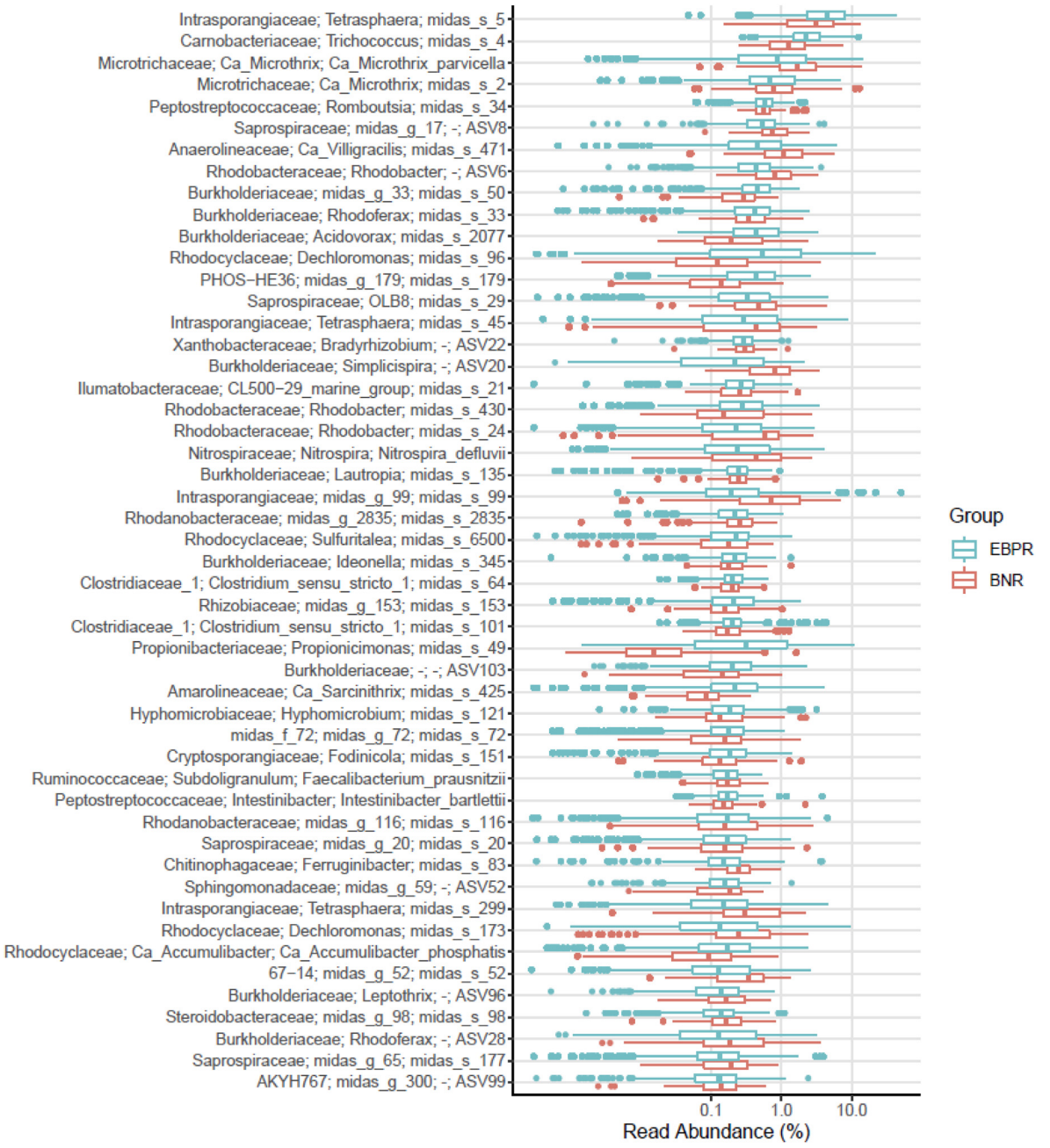

Fig. 6. Boxplot showing the occurrence of the top 50 most abundant species in Danish BNR and EBPR WWTPs. Family, genus, and species names are shown. Taxa including ASV numbers represent taxa that could not be confidently classified at higher taxonomic level. Missing taxonomic classifications are indicated by "-".

Rhodobacter genera. The novel genus midas_g_17, belonging to the family Saprospiraceae, is one of the most abundant bacteria in Danish plants, however, no functional information is available. Midas_g_17 has been incorrectly classified as *Ca.* Epiflobacter in the previous versions of MiDAS taxonomy. However, based on *in silico* analysis of the coverage of the FISH probes available, and mapping of the partial 16S rRNA gene sequences (Xia et al., 2008) to MiDAS 3 database, we have re-assigned *Ca.* Epiflobacter in MiDAS 3 to a correct genus, which is present at much lower abundance in the ecosystem (see *Ca.* Epiflobacter in Fig. S7). MiDAS 3 provides an opportunity to reclassify sequences from previous studies, for which information about distribution and/or function is available. For example, a set of partial 16S rRNA gene sequences originally classified to genus *Aquaspirillum* (Thomsen et al., 2007) could be reclassified as midas_g_33. Thomsen et al. (2007) demonstrated by FISH-microautoradiography that these sequences belong to bacteria possessing denitrifying capability, suggesting genus midas_g_33 as a potential important denitrifier in the ecosystem.

Interestingly, sub-genus level diversity among the ASVs belonging to the top 80% of the reads was generally low in the plants, with the average of 1.8 (and median 1.0) abundant species in each genus. However, the common PAO, *Tetrasphaera* and the less abundant *Ca.* Accumulibacter, exhibited much higher diversity, with 12 and 7 abundant species, respectively (five most abundant species for each PAO genus are shown in Fig. 7). *Tetrasphaera* had one dominant species (midas_s_5) present in most plants, while the other species appeared more randomly distributed. No clear dominance of any of *Ca.* Accumulibacter species could be observed

as their abundance was very different across the plants. Two of these, *Ca.* A. phosphatis and *Ca.* A. aalborgensis are described at species level, based on the genomes available (Albertsen et al., 2016; Martín et al., 2006). Interestingly, the genus *Dechloromonas*, which is also suggested to be a putative PAO (Stokholm-Bjerregaard et al., 2017), only had two dominant species present in almost all plants (Fig. 7). In general, high species diversity of the PAOs may suggest a high degree of specialization allowing the exploitation of different niches in the activated sludge microbiome. *Ca.* Accumulimonas, another putative PAO, which was described in the previous version of MiDAS, is not included in the present taxonomy version, as sequences representing that genus were reclassified to the genus *Halomonas* according to the updated SILVA taxonomy.

This study provides an insight for the first time, into species-level composition of the activated sludge microbiome. Physiological characterization is not available for the great majority of the presented species, since they represent hitherto unidentified microbes. Only a few of them are characterized based on available genomes and pure cultures. The importance of sub-genus level diversity in activated sludge ecosystem is largely unknown, however, since functional traits are often conserved at low taxonomic ranks, knowledge about identity and role of individual species in the ecosystem is of crucial importance to understanding the ecosystem function, its resilience to disturbances, and biological nutrient transformation routes. The significance of species-level microbiome composition is emphasized by recent studies of, e.g., comammox (Yang et al., 2020) or *Tetrasphaera* (Dueholm et al., 2019), where certain traits of crucial importance to the ecosystem function (ability to perform full nitrification and filamentous morphology, respectively) are only exhibited by some members of the genus.

3.4. The MiDAS reference database provides a common language for the field

Only a few studies have performed surveys of the microbiome in activated sludge systems, and in all cases, it has only been with genus-level and not species-level classification (Saunders et al., 2016; Wu et al., 2019). Since the different studies have applied different DNA extraction procedures, different primers (typically V1-3 or V4), and different databases for taxonomic assignment (MiDAS 2, SILVA, RDP, or Greengenes), it can be difficult to compare results across studies at genus level, even impossible at lower taxonomic levels. Therefore, the use of the ecosystem-specific MiDAS database provides a common language for all work in the field. Another key advantage of the MiDAS 3 taxonomy is that it includes taxa abundant in similar systems across the world (Dueholm et al., 2019), which are missing in large-scale public databases. By using this approach and similar protocols for extraction and primer selection (Albertsen et al., 2015; Dueholm et al., 2019), it will be possible to compare the results of WWTP microbiome studies in the future. Because MiDAS 3 uses the SILVA taxonomy as its backbone, it is also possible to merge classifications obtained using MiDAS 3 with those obtained from the corresponding SILVA database. This ensures high taxonomic resolution for all ASVs with a good match in the MiDAS 3 database, and allows, if reference sequences exist in the broader SILVA database, classifications of the remaining ASVs. This combined approach may be beneficial for wastewater treatment systems without nutrient removal, where other genera than those included in MiDAS 3 may dominate. In the future, however, the MiDAS reference database will continuously be updated with FL-ASVs from wastewater treatment systems from all parts of the world, and we are currently working on samples from ~1000 WWTPs from the whole world, covering all types of processes as well as wastewater and digester types.

3.5. The MiDAS reference database for anaerobic digesters

Many WWTPs have anaerobic digestion (AD) reactors associated to reduce waste sludge, generate energy, recover nutrients, and minimize the carbon footprint of the plant (Nielsen, 2017). In order to provide a comprehensive reference database for the microbes present in the whole wastewater treatment system, MiDAS 3 also includes FL-ASVs from 16 full-scale Danish ADs treating primary- and waste-activated sludge. Bacterial and archaeal microbiomes at species level, and the main factors driving their assembly based on a long-term survey of 46 Danish ADs have been presented in detail by (Jiang et al., 2020).

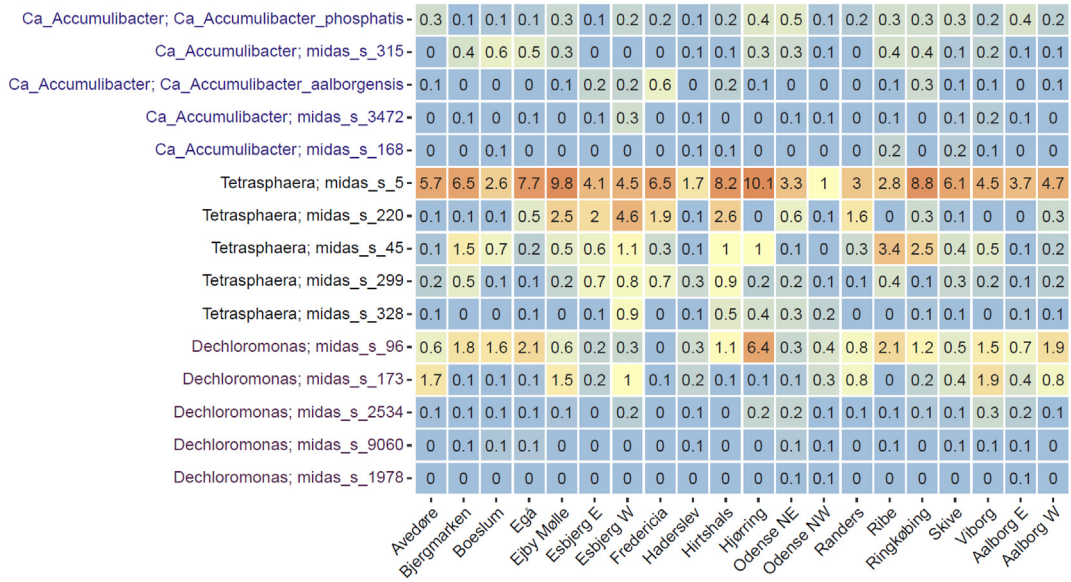

Fig. 7. Abundance of the top five species belonging to common PAO genera in 20 full-scale activated sludge WWTPs. Data represents average read abundance values for each plant, based on samples collected 2–4 times per year from 2006 to 2018. Individual PAO genera are indicated by different colors.

3.6. MiDAS field guide online platform (midasfieldguide.org)

The new MiDAS Field Guide (www.midasfieldguide.org) is a comprehensive database of microbes in wastewater treatment systems, which was updated together with MiDAS 3 release. All microbes found in activated sludge and anaerobic digesters are now included in the field guide and, for the first time, species are listed for each genus. Our database now covers more than 1800 genera and 4200 species found in biological nutrient removal treatment plants and anaerobic digesters. Detailed abundance information is now provided both at genus and species level, and functional information is described where available.

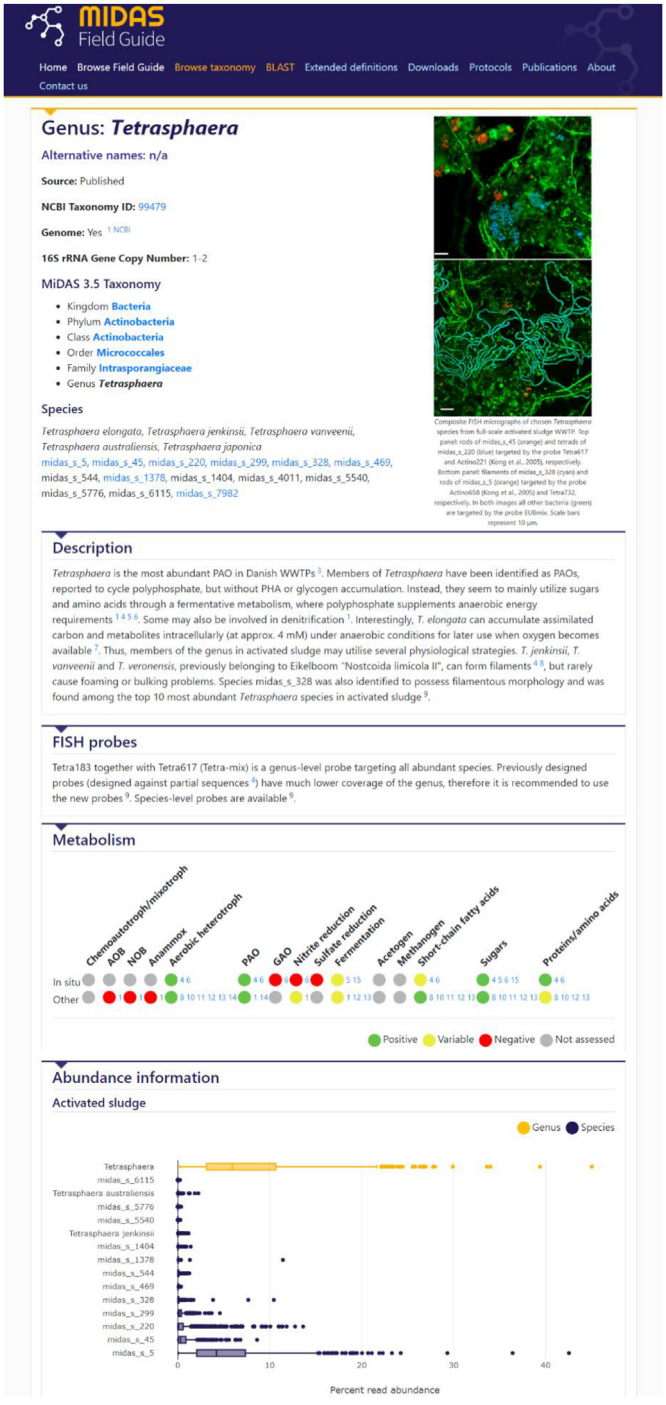

Fig. 8. MiDAS Field Guide example of *Tetrasphaera* genus entry.

The MiDAS Field Guide integrates the identity of the microbes with the available functional information (Fig. 8) into a community knowledge platform in the form of a searchable database of referenced information, thus acting as a central, on-line repository for current knowledge, where all are welcome to contribute. Due to the rapidly growing amount of relevant literature, and high number of taxa included, the MiDAS Field Guide is a continuous work in progress, with our efforts prioritized towards the microorganisms that comprise the core microbiome as well as other abundant or process-critical species.

New features launched in the updated MiDAS Field Guide include: 1) phylogenetic *Search by taxonomy* function, where all microbes found in the ecosystem are displayed in their respective taxonomic groups, 2) BLAST search function with complete MiDAS 3 taxonomy directly online. Updated protocols for DNA extraction and amplicon library preparation are available online to facilitate establishment of a common framework for studies of microbial ecology of wastewater treatment systems.

4. Conclusions

Here we present the new ecosystem-specific reference database, MiDAS 3, based on a comprehensive set of full-length 16S rRNA gene sequences derived from activated sludge and related anaerobic digester systems. It proposes unique provisional names for unclassified microorganisms in wastewater treatment microbiomes. MiDAS 3 was used to perform detailed analysis of the microbiome in 20 Danish WWTPs with nutrient removal, sampled over 13 years. This enabled unprecedented resolution, revealing for the first time abundant species present in the plants, variation of the communities across Danish plants, many abundant core taxa, and the diversity of species in important functional guilds, exemplified by the PAOs. We found many genera without known function, emphasizing the need for more efforts towards elucidating the role of important members of wastewater treatment ecosystems. We also present a new version of the MiDAS Field Guide, an online resource with a near-complete list of genera and species found in activated sludge and anaerobic digesters. Through the field guide, microbe identity is linked to the knowledge available of their function, and a blast function allows user 16S rRNA gene sequences to be blasted online against the MiDAS 3 taxonomy. The central purpose of the MiDAS Field Guide is to facilitate collaborative research into wastewater treatment ecosystem functions, which will support efforts towards the sustainable production of clean water and bioenergy, and the development of resource recovery, contributing ultimately to the circular economy of wastewater treatment systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The project has been funded by the Danish Research Council (grant 6111-00617A), Innovation Fund Denmark (OnlineDNA, grant 6155-00003A), the Villum Foundation (Dark Matter, grant 13351), and 20 Danish wastewater treatment plants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watres.2020.115955.

References

Albertsen, M., Karst, S.M., Ziegler, A.S., Kirkegaard, R.H., Nielsen, P.H., 2015. Back to basics – the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. PloS One 10, e0132783. https://doi.org/10.1371/journal.pone.0132783.

Albertsen, M., McIlroy, S.J., Stokholm-Bjerregaard, M., Karst, S.M., Nielsen, P.H., 2016. "Candidatus propionivibrio aalborgensis": a novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. Front. Microbiol. 7 https://doi.org/10.3389/fmicb.2016.01033.

Andersen, K.S., Kirkegaard, R.H., Karst, S.M., Albertsen, M., 2018a. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv 299537. https://doi.org/10.1101/299537.

Andersen, M.H., McIlroy, S.J., Nierychlo, M., Nielsen, P.H., Albertsen, M., 2018b. Genomic insights into Candidatus Amarolinea aalborgensis gen. nov., sp. nov., associated with settleability problems in wastewater treatment plants. Syst. Appl. Microbiol. https://doi.org/10.1016/j.syapm.2018.08.001.

Boughner, L.A., Singh, P., 2016. Microbial Ecology: where are we now? Postdoc J. 4, 3–17. https://doi.org/10.14304/SURYA.JPR.V4N11.2.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119.

Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 47 https://doi.org/10.1093/nar/gkz569 e103–e103.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642. https://doi.org/10.1093/nar/gkt1244.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05.

Dinno, A., 2017. Dunn's Test of Multiple Comparisons Using Rank Sums.

Dueholm, M.S., Andersen, K.S., Petriglieri, F., McIlroy, S.J., Nierychlo, M., Petersen, J.F., Kristensen, J.M., Yashiro, E., Karst, S.M., Albertsen, M., Nielsen, P.H., 2019. Comprehensive ecosystem-specific 16S rRNA gene databases with automated taxonomy assignment (AutoTax) provide species-level resolution in microbial ecology. bioRxiv.

Edgar, R.C., 2018. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ 6, e4652. https://doi.org/10.7717/peerj.4652.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv, 081257. https://doi.org/10.1101/081257.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

Glöckner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., Ludwig, W., 2017. 25 years of serving the community with ribosomal RNA gene reference databases and tools. In: Journal of Biotechnology, Bioinformatics Solutions for Big Data Analysis in Life Sciences presented by the German Network for Bioinformatics Infrastructure, 261, pp. 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198.

Herbst, F.-A., Dueholm, M.S., Wimmer, R., Nielsen, P.H., 2019. The proteome of Tetrasphaera elongata is adapted to changing conditions in wastewater treatment plants. Proteomes 7. https://doi.org/10.3390/proteomes7020016.

Jiang, C., Peces, M., Andersen, M.H., Kucheryavskiy, S., Nierychlo, M., Yashiro, E., Andersen, K.S., Kirkegaard, R.H., Hao, L., Høgh, J., Hansen, A.A., Dueholm, M.S., Nielsen, P.H., 2020. Characterizing the growing microorganisms at species-level in 46 anaerobic digesters at Danish wastewater treatment plants: a six-year survey on microbiome structure and key drivers. bioRxiv.

Karst, S.M., Dueholm, M.S., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H., Albertsen, M., 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat. Biotechnol. https://doi.org/10.1038/nbt.4045.

Karst, S.M., Ziels, R.M., Kirkegaard, R.H., Albertsen, M., 2019. Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers and Nanopore sequencing. bioRxiv 645903. https://doi.org/10.1101/645903.

Kim, B.-R., Shin, J., Guevarra, R.B., Lee, J.H., Kim, D.W., Seol, K.-H., Lee, J.-H., Kim, H.B., Isaacson, R.E., 2017. Deciphering diversity indices for a better understanding of microbial communities. J. Microbiol. Biotechnol. 27, 2089–2093. https://doi.org/10.4014/jmb.1709.09027.

Kristiansen, R., Nguyen, H.T.T., Saunders, A.M., Nielsen, J.L., Wimmer, R., Le, V.Q., McIlroy, S.J., Petrovski, S., Seviour, R.J., Calteau, A., Nielsen, K.L., Nielsen, P.H., 2013. A metabolic model for members of the genus Tetrasphaera involved in enhanced biological phosphorus removal. ISME J. 7, 543–554. https://doi.org/10.1038/ismej.2012.136.

Lane, D.J., 1991. 16S/23S rRNA sequencing. In: Nucleic acid techniques in bacterial systematics, pp. 115–175.

Liu, J.R., Seviour, R.J., 2001. Design and application of oligonucleotide probes for fluorescent *in situ* identification of the filamentous bacterial morphotype Nostocoida limicola in activated sludge. Environ. Microbiol. 3, 551–560. https://doi.org/10.1046/j.1462-2920.2001.00229.x.

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

Martín, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A.A., Szeto, E., Dalin, E., Putnam, N.H., Shapiro, H.J., Pangilinan, J.L., Rigoutsos, I., Kyrpides, N.C., Blackall, L.L., McMahon, K.D., Hugenholtz, P., 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat. Biotechnol. 24, 1263–1269. https://doi.org/10.1038/nbt1247.

McIlroy, S.J., Kirkegaard, R.H., McIlroy, B., Nierychlo, M., Kristensen, J.M., Karst, S.M., Albertsen, M., Nielsen, P.H., 2017. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database 2017. https://doi.org/10.1093/database/bax016.

McIlroy, S.J., Kristiansen, R., Albertsen, M., Michael Karst, S., Rossetti, S., Lund Nielsen, J., Tandoi, V., James Seviour, R., Nielsen, P.H., 2013. Metabolic model for the filamentous 'Candidatus Microthrix parvicella'' based on genomic and metagenomic analyses. ISME J. 7, 1161–1172. https://doi.org/10.1038/ismej.2013.6.

McIlroy, S.J., Saunders, A.M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A.A., Karst, S.M., Nielsen, J.L., Nielsen, P.H., 2015. MiDAS: the field guide to the microbes of activated sludge. Database 2015, bav062. https://doi.org/10.1093/database/bav062.

McIlroy, S.J., Starnawska, A., Starnawski, P., Saunders, A.M., Nierychlo, M., Nielsen, P.H., Nielsen, J.L., 2014. Identification of active denitrifiers in full-scale nutrient removal wastewater treatment systems. Environ. Microbiol. https://doi.org/10.1111/1462-2920.12614 n/a-n/a.

Muyzer, G., de Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700.

Nielsen, P.H., 2017. Microbial biotechnology and circular economy in wastewater treatment. Microb. Biotechnol. 10, 1102–1105. https://doi.org/10.1111/1751-7915.12821.

Nierychlo, M., Miłobędzka, A., Petriglieri, F., McIlroy, B., Nielsen, P.H., McIlroy, S.J., 2019. The morphology and metabolic potential of the Chloroflexi in full-scale activated sludge wastewater treatment plants. FEMS Microbiol. Ecol. 95 https://doi.org/10.1093/femsec/fiy228.

Nierychlo, M., Nielsen, P.H., 2017. Experiences in various countries: Denmark. In: Tandoi, V., Rossetti, S., Wanner, J. (Eds.), Activated Sludge Separation Problems: Theory, Control Measures, Practical Experiences. IWA Publishing, pp. 198–209.

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. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at: https://www.R-project.org/.

Rossetti, S., Tomei, M.C., Nielsen, P.H., Tandoi, V., 2005. "Microthrix parvicella," a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. FEMS Microbiol. Rev. 29, 49–64. https://doi.org/10.1016/j.femsre.2004.09.005.

RStudio Team, 2015. RStudio. Integrated Development for R. RStudio, Inc., Boston, MA. URL. http://www.rstudio.com/.

Saunders, A.M., Albertsen, M., Vollertsen, J., Nielsen, P.H., 2016. The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 10, 11–20.

Stokholm-Bjerregaard, M., McIlroy, S.J., Nierychlo, M., Karst, S.M., Albertsen, M., Nielsen, P.H., 2017. A critical assessment of the microorganisms proposed to be important to Enhanced Biological Phosphorus Removal in full-scale wastewater treatment systems. Front. Microbiol. 8 https://doi.org/10.3389/fmicb.2017.00718.

Strepis, N., Sánchez-Andrea, I., van Gelder, A.H., van Kruistum, H., Shapiro, N., Kyrpides, N., Göker, M., Klenk, H.-P., Schaap, P., Stams, A.J.M., Sousa, D.Z., 2016. Description of Trichococcus ilyis sp. nov. by combined physiological and in silico genome hybridization analyses. Int. J. Syst. Evol. Microbiol. 66, 3957–3963. https://doi.org/10.1099/ijsem.0.001294.

Thomsen, T.R., Kong, Y., Nielsen, P.H., 2007. Ecophysiology of abundant denitrifying bacteria in activated sludge. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol. 60, 370–382. https://doi.org/10.1111/j.1574-6941.2007.00309.x.

Wickham, H., 2009. ggplot2 - Elegant Graphics for Data Analysis. Springer Science + Business Media.

Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M., Li, Z., Nostrand, J.D.V., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Zhang, T., He, Z., Keller, J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis, T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat. Microbiol. 1 https://doi.org/10.1038/s41564-019-0426-5.

Xia, Y., Kong, Y., Nielsen, P.H., 2008. *In situ* detection of starch-hydrolyzing microorganisms in activated sludge. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol. 66, 462–471. https://doi.org/10.1111/j.1574-6941.2008.00559.x.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645. https://doi.org/10.1038/nrmicro3330.