



Low Global Diversity of *Candidatus* *Microthrix*, a Troublesome Filamentous Organism in Full-Scale WWTPs

Marta Nierychlo, Caitlin M. Singleton, Francesca Petriglieri, Lisette Thomsen, Jette F. Petersen, Miriam Peces, Zivile Kondrotaitė, Morten S. Dueholm and Per H. Nielsen*

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

OPEN ACCESS

Edited by:

Harold J. Schreier,
University of Maryland, Baltimore
County, United States

Reviewed by:

Ran Mei,
National Institute of Advanced
Industrial Science and Technology
(AIST), Japan
Despina Tsementzi,
Emory University, United States

*Correspondence:

Per H. Nielsen
phn@bio.aau.dk

Specialty section:

This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

Received: 02 April 2021

Accepted: 25 May 2021

Published: 25 June 2021

Citation:

Nierychlo M, Singleton CM,
Petriglieri F, Thomsen L, Petersen JF,
Peces M, Kondrotaitė Z, Dueholm MS
and Nielsen PH (2021) Low Global
Diversity of *Candidatus* *Microthrix*,
a Troublesome Filamentous Organism
in Full-Scale WWTPs.
Front. Microbiol. 12:690251.
doi: 10.3389/fmicb.2021.690251

Candidatus *Microthrix* is one of the most common bulking filamentous microorganisms found in activated sludge wastewater treatment plants (WWTPs) across the globe. One species, *Ca. M. parvicella*, is frequently observed, but global genus diversity, as well as important aspects of its ecology and physiology, are still unknown. Here, we use the MiDAS ecosystem-specific 16S rRNA gene database in combination with amplicon sequencing of Danish and global WWTPs to investigate *Ca. Microthrix* spp. diversity, distribution, and factors affecting their global presence. Only two species were abundant across the world confirming low diversity of the genus: the dominant *Ca. M. parvicella* and an unknown species typically present along with *Ca. M. parvicella*, although usually in lower abundances. Both species were mostly found in Europe at low-to-moderate temperatures and their growth was favored in municipal WWTPs with advanced process designs. As no isolate is available for the novel species, we propose the name “*Candidatus* *Microthrix* subdominans.” Ten high-quality metagenome-assembled genomes recovered from Danish WWTPs, including 6 representing the novel *Ca. M. subdominans*, demonstrated high genetic similarity between the two species with a likely preference for lipids, a putative capability to reduce nitrate and nitrite, and the potential to store lipids and poly-P. *Ca. M. subdominans* had a potentially more versatile metabolism including additional sugar transporters, higher oxygen tolerance, and the potential to use carbon monoxide as energy source. Newly designed fluorescence *in situ* hybridization probes revealed similar filamentous morphology for both species. Raman microspectroscopy was used to quantify the *in situ* levels of intracellular poly-P. Despite the observed similarities in their physiology (both by genomes and *in situ*), the two species showed different seasonal dynamics in Danish WWTPs through a 13-years survey, possibly indicating occupation of slightly different niches. The genomic information provides the basis for future research into *in situ* gene expression and regulation, while the new FISH probes provide a useful tool for further characterization *in situ*. This study is an important step toward understanding the ecology of *Ca. Microthrix* in WWTPs, which may eventually lead to optimization of control strategies for its growth in this ecosystem.

Keywords: activated sludge, *Ca. Microthrix*, bulking, diversity, genome, Raman, fish, amplicon sequencing

INTRODUCTION

The activated sludge process is the most common wastewater treatment technology applied worldwide. In recent years, process design has developed rapidly across the globe to meet the demands of growing urban populations. Additionally, the role of wastewater treatment plants (WWTPs) has moved toward resource recovery as societies strive toward sustainability and a circular economy. However, the unsolved problems frequently experienced at the treatment plants are bulking and foaming resulting from overgrowth of filamentous bacteria, affecting stable operation, plant hydraulic capacity, and effluent quality.

Filamentous microorganisms are often linked to deteriorated sludge settleability and foam formation (Rossetti et al., 2005; Seviour and Nielsen, 2010; Wanner, 2017). A number of studies, mainly from Europe and northern China (Wang et al., 2016; Wanner, 2017; Fan et al., 2019), have identified species in the genus *Ca. Microthrix* as some of the most troublesome filamentous bacteria in municipal WWTPs in addition to *Gordonia*, *Ca. Amarolinea*, and other phylotypes belonging to the phylum Chloroflexi (Martins et al., 2004; Speirs et al., 2019; Nierychlo et al., 2020b). *Ca. Microthrix* has long been known by both researchers and practitioners in the field, and was identified almost 50 years ago (van Veen, 1973). It has been known as morphotype “*Microthrix parvicella*,” and can successfully be identified by light microscopy due to its distinct morphological features. These include unbranched, curled filaments with a hydrophobic surface staining Gram-positive, and often with clear polyphosphate (poly-P) granules visible after Neisser staining (van Veen, 1973; Eikelboom and van Buijsen, 1983; Nielsen et al., 2002). However, despite numerous studies on the physiology and ecology and potential control measures (Fan et al., 2020), a better understanding of its ecophysiology in activated sludge is needed to prevent *Ca. Microthrix*-related bulking and foaming events.

Two species of *Ca. Microthrix* are described in the literature: *Ca. M. parvicella*, abundant in municipal plants with nutrient removal (Erhart et al., 1997), and *Ca. M. calida* found in some industrial plants (Levantesi et al., 2006). Both species have been isolated and their physiology assessed in pure culture, although *Ca. M. calida* lacks a published reference genome. *Ca. M. parvicella* was shown to grow aerobically using volatile fatty acids (VFAs) or long chain fatty acids (LCFAs) (Slijkhuis, 1983; Tandoi et al., 1998), while *Ca. M. calida* is unable to grow on volatile fatty acids (VFAs) and requires LCFAs (Levantesi et al., 2006). However, the very slow growth of *Ca. Microthrix* isolates and difficulties with culture maintenance have hampered its extensive physiological characterization (Blackall et al., 1996; Fan et al., 2019).

Several *in situ* studies using fluorescence *in situ* hybridization (FISH) combined with microautoradiography or nano-scale secondary-ion mass spectrometry visualized important differences compared to the isolate studies. *Ca. M. parvicella* has never been observed *in situ* to consume VFAs but only LCFAs and glycerol (Andreasen and Nielsen, 1998; Kindaichi et al., 2013;

Sheik et al., 2016). Furthermore, LCFA uptake followed by storage as lipids was observed under both aerobic and anaerobic conditions (Andreasen and Nielsen, 2000; Nielsen et al., 2002), which provides a competitive advantage in nutrient removal plants. *Ca. M. parvicella* were also shown to likely grow under anoxic conditions *in situ* by reducing nitrate to nitrite (Hesselsoe et al., 2005). The *in situ* activity of extracellular lipases (Schade and Lemmer, 2005) and their hydrophobic cell surface (Nielsen et al., 2002) further support the specialization of *Ca. Microthrix* spp. toward lipidic substrates. Presence of polyphosphate (poly-P) granules inside *Ca. M. parvicella* filaments has been shown in isolates and *in situ* (Erhart et al., 1997; Tandoi et al., 1998; Wang et al., 2014b; Fei et al., 2021; Petriglieri et al., 2021), placing it on the list of putative polyphosphate accumulating organisms (PAO). However, P cycling during alternating aerobic and anaerobic conditions has not been confirmed so far (Andreasen and Nielsen, 2000; Petriglieri et al., 2021). While the genome of *Ca. M. calida* has not been sequenced, two genomes representing *Ca. M. parvicella* are published: Dutch isolate Bio-17 (Muller et al., 2012), and Italian isolate RN1 (McIlroy et al., 2013). These genomes confirmed the physiology observed by *in situ* studies, showing assimilation and storage of LCFAs, the ability to store poly-P, and the potential to reduce nitrate and nitrite.

A few recent studies have made it possible to investigate the global diversity of *Ca. Microthrix* (Karst et al., 2018; Dueholm et al., 2020; Nierychlo et al., 2020a). The comprehensive and WWTP ecosystem-specific full-length 16S rRNA gene reference database, MiDAS4 (Dueholm et al., 2021), has enabled studies of the global WWTP microbiota at species-level resolution and provided unique placeholder names for the large diversity of undescribed populations. The ecosystem-specific database can also be used for the design and re-evaluation of FISH probes for the microbes of interest. These can be used in combination with other techniques to validate genomic information and confirm microorganism function in the WWTP ecosystem. Recent studies have granted valuable access to a high number of metagenome-assembled genomes (MAGs) representing microorganisms present *in situ* in the WWTP environment (Parks et al., 2017; Singleton et al., 2021). These MAGs together with high-throughput amplicon sequencing techniques provide a valuable tool set to explore the diversity of microbes present in the WWTP ecosystem.

In this study we investigated the diversity of genus *Ca. Microthrix* in activated sludge plants worldwide and found one abundant undescribed species: *Ca. Microthrix* subdominans. Using 16S rRNA gene amplicon sequencing data we investigated factors correlating to their global presence, and their influence on sludge settling properties. Furthermore, we compared the metabolic potential of *Ca. M. parvicella* and the *Ca. M. subdominans* based on new high-quality (HQ) MAGs. Both species were visualized *in situ* using novel FISH probes, which were further applied to investigate poly-P content using FISH-Raman microspectroscopy. Our results provide important new information about this troublesome filamentous genus in the activated sludge ecosystem, which may help to develop more targeted strategies for its control.

MATERIALS AND METHODS

Biomass Sampling and Fixation and Sample Processing

Sampling of the activated sludge biomass from Danish full-scale municipal WWTPs was carried out within the MiDAS project (McIlroy et al., 2015; Nierychlo et al., 2020a). Sampling of full-scale plants across the world was coordinated by the MiDAS Global Consortium and carried out within the Global MiDAS project (Dueholm et al., 2021). Activated sludge was sampled from an aeration tank and sent directly to Aalborg University (Danish samples) or preserved in RNAlater and shipped to Aalborg University with cooling elements (Global samples). All samples were frozen upon arrival and stored until processing. For details, see Nierychlo et al. (2020a) and Dueholm et al. (2021).

Amplicon Sequencing and Bioinformatic Analysis

DNA extraction, sample preparation, including amplification and amplicon sequencing were conducted as described in Nierychlo et al. (2020a) and Dueholm et al. (2021). Briefly, V1-V3 16S rRNA gene regions were amplified using the 27F (AGAGTTTGATCCTGGCTCAG) (Lane, 1991) and 534R (ATTACCGCGGCTGCTGG) (Muyzer et al., 1993) primers, and the resulting amplicons were used in all the analyses. The V4 16S rRNA gene region was amplified using the 515F (5'-GTGYCAGCMGCCGCGGTAA-3') (Parada et al., 2016) and 806R (5'-GGACTACN VGGGTWTCTAAT-3') (Apprill et al., 2015) primers. The Danish samples (Nierychlo et al., 2020a) were only sequenced with the V1-V3 primers, while the global samples (Dueholm et al., 2021) were sequenced by both primer pairs (V1-V3 and V4) to compare *Ca. Microthrix* abundances between the two amplicon datasets. The handling of the Danish and the global dataset in regard to DNA extraction and library preparation protocol were similar, providing comparable datasets. 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 -fastq_maxee 1.0`. Dereplication was performed using `usearch -fastx_uniques` with `-sizeout`. Amplicon sequence variants (ASVs) were generated using `-unnoise3` (Edgar, 2016) with standard settings, and taxonomy was assigned using the MiDAS3 (for Danish dataset) or MiDAS4 (for Global dataset) reference database¹ and the SINTAX classifier with `-strand both` and `-sintax_cutoff 0.8` options (Edgar, 2018).

Data Analysis and Visualization

R v.3.5.1 (R Core Team, 2020) and RStudio (RStudio Team, 2020) were used for data processing. The dataset with Danish plants contained 712 samples from 20 nutrient removal WWTPs with minimum 13,500 reads, the dataset with Global samples contained 847 samples from 438 WWTPs with 4 different process designs [carbon removal plants (C), carbon removal and nitrification (C,N), carbon removal, nitrification and

denitrification (C,N,DN), and EBPR plants (C,N,DN,P)] with minimum 10,000 reads. All data was visualized using R packages `ggplot2` v.3.2.1 (Wickham, 2009) and `ampvis2` v.2.4.9 (Andersen et al., 2018). The Kruskal-Wallis test was used to determine statistically significant differences in *Ca. Microthrix* abundances across different groupings. Sequencing data for Danish and Global is available at the Sequence Read Archive² as stated in Nierychlo et al. (2020a) and Dueholm et al. (2021), respectively.

Phylogenetic Analysis of 16S rRNA Gene Sequences and Fluorescence *in situ* Hybridization (FISH) Probes Design

Phylogenetic analysis of 16S rRNA gene sequences and design of FISH probes for the novel species were performed using the MiDAS4 database containing high-quality full-length 16S rRNA gene sequences from the activated sludge ecosystem (Dueholm et al., 2021) using the ARB software v.6.0.6 (Ludwig et al., 2004). Novel probes were assessed *in silico* with the mathFISH software for hybridization efficiencies with target sequences, and potentially weak, non-target matches (Yilmaz et al., 2011). Where needed, unlabeled helper probes were designed to facilitate probe access to the target side. All probes were purchased from Biomers (Ulm, Germany), labeled with cyanine-3 (Cy3), cyanine-5 (Cy5) or 6-FAM fluorochromes.

A phylogenetic 16S rRNA gene tree was calculated based on full-length gene sequences retrieved from MiDAS4 (Dueholm et al., 2021) and SILVA 138 SSURef Nr99 (Quast et al., 2013) databases as well as gene sequences from the recovered MAGs using maximum likelihood method using the GTR model and 1000- replicates bootstrap analysis. Two of the MAGs had identical 16S rRNA gene sequences (Bjer_MB2 and AalW_MB3), so one was chosen for incorporation into the **Figure 4** tree, which is labeled to indicate this. Trees were viewed in ARB v6.0.3 (Ludwig et al., 2004), and displayed in iTOL v5.7 (Letunic and Bork, 2019) with final presentation in Inkscape v0.92.

FISH, Probe Optimization, and Quantitative FISH (qFISH)

FISH was performed as previously described (Nielsen, 2009). Validation and optimization of novel probes were based on formamide (FA) dissociation curves by carrying out hybridizations over a range of 0–70% (v/v) FA concentrations with increments of 5% (data not shown). Activated sludge samples from Danish WWTPs with high abundance of the target organism, as predicted by amplicon sequencing, were used. Microscopic analysis was performed with Axioskop epifluorescence microscope (Carl Zeiss, Germany) equipped with LEICA DFC7000 T CCD camera or with a white light laser confocal microscope (Leica TCS SP8 X). The intensity of at least 50 cells at each FA concentration was measured with the ImageJ software (National Institutes of Health, Bethesda, MD, United States). Optimal hybridization conditions and details about the coverage and specificity of all the probes used in the study are presented in **Table 1**. EUBmix (Amann et al., 1990;

¹<https://www.midasfieldguide.org/guide/downloads>

²<https://www.ncbi.nlm.nih.gov/sra>

TABLE 1 | Overview of the existing and new FISH probes targeting *Ca. Microthrix*.

Probe	<i>E. coli</i> pos.	Target group	Coverage*	Non-target hits	Sequence (5'–3')	Pre-treatment**	%FA***	References
MCX840	840–859	Genus <i>Ca. Microthrix</i>	MIDAS3: 8/14 MIDAS4: 21/28 Silva138: 24/28	0 1 0	CGG CGC GGA GAG AGT TGA GT	A	20	This study
MCX840_H1	–	Helper for MCX840			TCT CCC CAC ACC TAG TGC CCA ACG		–	
MCX840_H2	–	Helper for MCX840			GCG GGG CAC TTA ATG CGT TAG CTA		–	
Mpa177	177–199	<i>Ca. Microthrix parvicella</i>	MIDAS3: 3/3 MIDAS4: 4/6 Silva138: 11/28	0 0 1	GCG GTG AAG AGA AGG TAT GCG GT	A	45	This study
Mpa177_H1	–	Helper for Mpa177			AGA GCG ATA AAT CTT TCT TCA ACT CAC CAT		–	
Mpa177_H2	–	Helper for Mpa177			TCG AAT TTC TTC GAG TTA TTC CCC ACT CC		–	
Msu181	181–203	<i>Ca. Microthrix subdominans</i>	MIDAS3: 2/4 MIDAS4: 4/10 Silva138: 8/28	0 0 0	CCA CCA TTC GAC GGT AAG AAG GT	A	40	This study
Msu181_H1	–	Helper for Msu181			CTC CCA GAG CGA TAA ATC TTT CTT CAA A		–	
Msu181_H2	–	Helper for Msu181			ATG CGG TAT TAG CTC GAA TTT CTT CGA		–	
MPA60	60–77	<i>Ca. Microthrix parvicella</i>	MIDAS3: 3/3 MIDAS4: 6/6 Silva138: 8/28	0 0 1	GGA TGG CCG CGT TCG ACT	B/C/D	20	Erhart et al., 1997
MPA223	223–240	Genus <i>Ca. Microthrix</i>	MIDAS3: 4/14 MIDAS4: 10/28 Silva138: 9/28	8 20 28	GCC GCG AGA CCC TCC TAG	B/C/D	20	Erhart et al., 1997
MPA645	546–661	Genus <i>Ca. Microthrix</i>	MIDAS3: 14/14 MIDAS4: 27/28 Silva138: 27/28	0 1 1	CCG GAC TCT AGT CAG AGC	B/C/D	20	Erhart et al., 1997
MPA650	650–666	Genus <i>Ca. Microthrix</i>	MIDAS3: 14/14 MIDAS4: 28/28 Silva138: 27/28	41 103 151	CCC TAC CGG ACT CTA GTC	B/C/D	20	Erhart et al., 1997
Mpa_all_1410	1411–1430	<i>Ca. Microthrix calida</i>	MIDAS3: 5/6 MIDAS4: 6/7 Silva138: 1/1	1 1 0	GGT GTT GTC GAC TTT CGG CG	B/C/D	35	Levantesi et al., 2006
Mpa-T1-1260	1261–1280	<i>Ca. Microthrix calida</i>	MIDAS3: 6/6 MIDAS4: 7/7 Silva138: 1/1	0 1 0	TTC GCA TGA CCT CAC GGT TT	B/C/D	25	Levantesi et al., 2006

*Probe coverage is given for MIDAS 3.7 (Nierychlo et al., 2020a), 4.8 (Dueholm et al., 2021), and Silva 138 (Quast et al., 2013) database. Values given as group hits/group totals; **A: 30 min lysozyme followed by 30 min achromopeptidase and overnight hybridization; B: 15–25 min mutanolysin (Erhart et al., 1997); C: achromopeptidase and lysozyme (Kragelund et al., 2007); D: lysozyme and extended hybridization time 15–18 h; ***Recommended optimal FA concentration for use in FISH hybridizations.

(Daims et al., 1999) was used to target all bacteria. The NON-EUB probe (Wallner et al., 1993) was used as a negative control for sequence independent probe binding. Quantitative FISH (qFISH) biovolume fractions of individual species and genera were calculated as a percentage area of the total biovolume, hybridizing with both EUBmix and specific probes. qFISH analysis was performed using the Daime image analysis software (Daims et al., 2006), based on 30 fields of view taken at 630× magnification.

Morphological Characterization

A minimum of 60 probe-defined filaments were randomly chosen to measure cell length and width using the segmented line tool in ImageJ (National Institutes of Health, Maryland United States).

Raman Microspectroscopy

Fresh activated sludge samples were collected from full-scale Danish WWTPs and aerated for 30 min to exhaust most intracellular carbon and refill poly-P reserves. Samples were fixed in 50% ethanol (final concentration), as previously described

(Nielsen, 2009), and stored at –20°C until analysis. Raman microspectroscopy was applied in combination with FISH as previously described (Fernando et al., 2019). Briefly, FISH was conducted on optically polished CaF₂ Raman windows (Crystran, United Kingdom). The specific FISH probes were used to locate the target cells with a 50X dry objective (Olympus M Plan Achromat- Japan) of the in-built Olympus (model BX-41) fluorescence microscope. After bleaching, spectra from different filaments were obtained using a Horiba LabRam HR 800 Evolution (Jobin Yvon—France) equipped with a Torus MPC 3000 (United Kingdom) 532 nm 341 mW solid-state semiconductor laser. The specific settings for the spectrophotometer were: 5% neutral density (ND) filters, 600 mm/groove diffraction grating, 100 μm and 72 μm slit width and confocal pinhole, respectively. The Raman spectra collected spanned the wavenumber region of 200 cm⁻¹ to 3,000 cm⁻¹. The Raman spectrometer was calibrated prior to obtaining all measurements to the first-order Raman signal of Silicon, occurring at 520.7 cm⁻¹. Raman spectrometer

operation and subsequent processing of spectra were conducted using LabSpec version 6.4 software (Horiba Scientific, France). Absolute quantification of intracellular poly-P was carried out as described by Fernando et al. (2019). The method assumes that the intensity of the Raman signal is directly dependent on the amount of the analyte in a determined area. An average amount of poly-P per cell was calculated as a factor of a constant determined during calibration for poly-P (Fernando et al., 2019), the average charge-coupled device (CCD) counts determined during the experiment, and the average area of cells measured by image analysis. For filaments, it is very difficult to determine the internal area of single cells with common microscopic techniques. An arbitrary area was determined by multiplying the filament width by 1 μm , thus obtaining the amount of poly-P in an average filament segment of 1 μm length.

Identification, Annotation, and Metabolic Reconstruction of *Ca. Microthrix*

Ca. Microthrix MAGs were identified from within a set of 1083 high-quality MAGs from Danish activated sludge WWTPs (Singleton et al., 2021, NCBI BioProject PRJNA629478). Ten MAGs were identified, and followed the MIMAG standard for high-quality draft genomes, including full-length rRNA genes, and a completeness > 90%, and contamination < 5%. GTDB-Tk v1.4.1 (Chaumeil et al., 2020) with RefSeq release 95 was used for taxonomic classification, and the species representatives were selected using dRep v2.3.2 (Olm et al., 2017) at 95% average nucleotide identity and the CheckM v1.1.2 (Parks et al., 2015) completeness and contamination quality statistics. 16S rRNA genes in the MAGs were identified within the genomes using Infernal v1.1.2 (Nawrocki and Eddy, 2013), and BEDTools v2.27 (Quinlan and Hall, 2010) was used to extract the sequences for inclusion in the 16S rRNA gene phylogenetic tree. *Ca. M. parvicella* genomes and MAGs present in NCBI but not the dereplicated GTDB-Tk Refseq release 95 database were downloaded (on 26/02/2021) and incorporated with the Danish MAGs and the GTDB-Tk set for phylogenomic analysis. The genome tree was created from the concatenated, trimmed alignment of 120 single copy proteins produced by GTDB-Tk. IQ-TREE v2.0 (Nguyen et al., 2015) was used to create a maximum likelihood tree from this alignment following the WAG+G model with 100 bootstrap iterations. Trees were viewed in ARB v6.0.3 (Ludwig et al., 2004), and displayed in iTOL v5.7 (Letunic and Bork, 2019) with final presentation in Inkscape v0.92.

EnrichM v0.5.0³ “annotate” was used to annotate the MAGs against the KEGG Orthology (KO) number (Kanehisa and Goto, 2000) annotated uniref100 database (EnrichM database v10) using the default settings. EnrichM “enrichment” was used to determine the enriched presence of KOs in the *Ca. M. parvicella* or *Ca. M. subdominans* MAGs, to determine species differences. Additionally, the MAGs were uploaded to the “MicroScope Microbial Genome Annotation and Analysis Platform” (Vallenet et al., 2020) in order to examine gene synteny and cross-validate KO annotations found using EnrichM. Core and species

specific genes were also examined using the MicroScope platform (Vallenet et al., 2020). Specifically, the *Ca. M. subdominans* MAGs were compared to the *Ca. M. parvicella* MAGs and the *Ca. M. parvicella* RN1 isolate genome using the pan-genome analysis function using 50% amino acid identity and 80% alignment coverage. Core genes specific to each species were identified and examined within the surrounding genomic context (i.e., syntenic genes), and additional checks were conducted using protein BLAST against the NCBI nr database (Camacho et al., 2009).

Average nucleotide identity based on blast (ANIb) was conducted using pyani v0.2.10 (Pritchard et al., 2016) and the “average_nucleotide_identity.py -m ANIb” arguments. R v4.0.3 (R Core Team, 2020), and the libraries gplots, RColorBrewer and reshape2, were used to process the produced files, ANIb_percentage_identity.tab and ANIb_alignment_coverage.tab, into ANI and alignment heatmaps with the heatmap.2 function. Inkscape v0.92 was used for the final presentation.

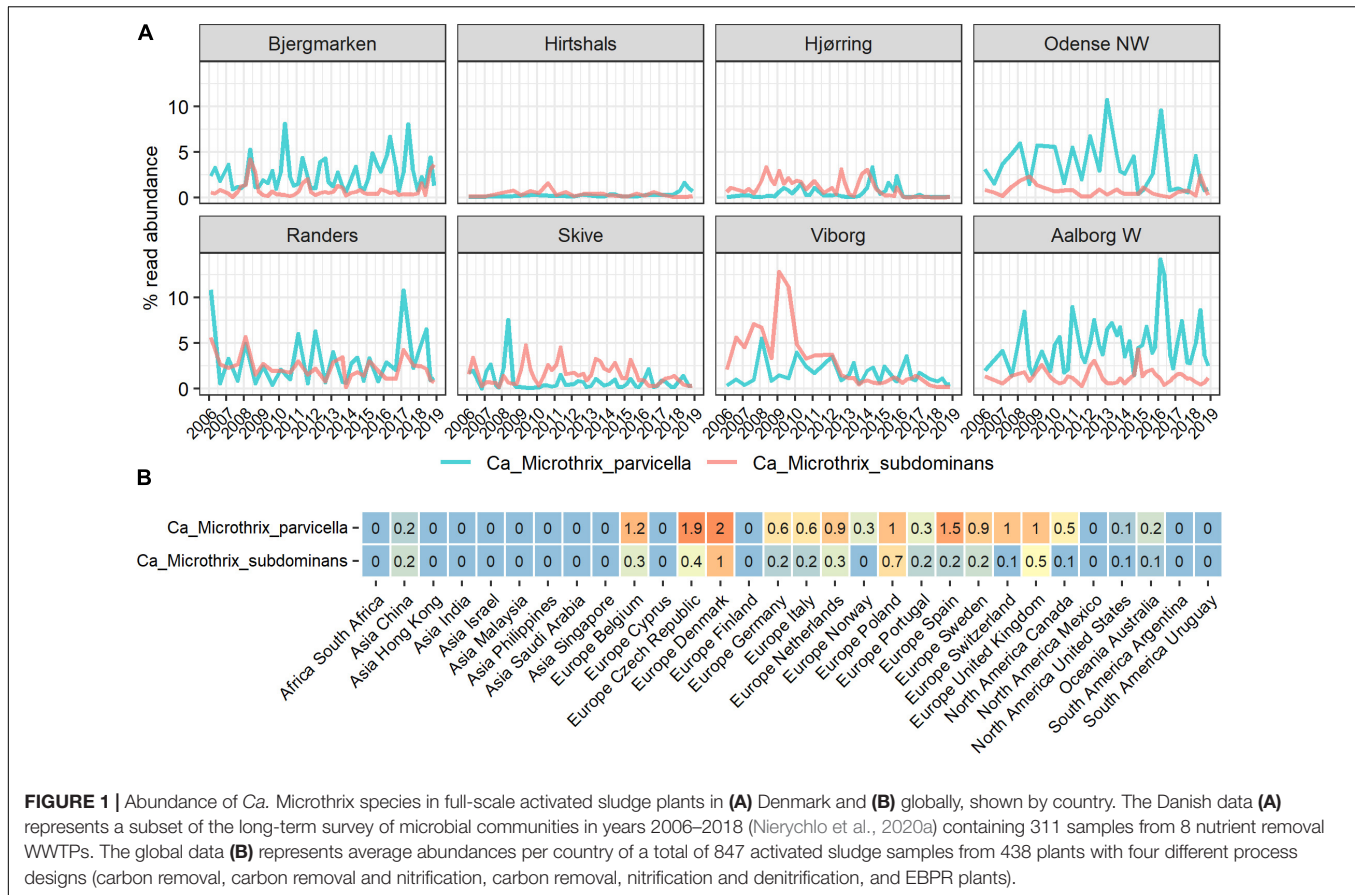
RESULTS AND DISCUSSION

Diversity and Distribution of *Ca. Microthrix* Species in Full-Scale WWTPs by Amplicon Sequencing

The diversity and occurrence of *Ca. Microthrix* species in the activated sludge ecosystem were analyzed using data from long-term surveys of Danish WWTPs (Nierychlo et al., 2020a) as well as the recent Global WWTP survey of 438 samples (Dueholm et al., 2021). We observed a widespread presence of two *Ca. Microthrix* species (**Figure 1**): the well-known and characterized *Ca. M. parvicella*, and a second, novel species with the provisional name in the MiDAS taxonomy: midas_s_2, but here given the name *Candidatus Microthrix subdominans* (see section Etymology). The two other low-abundant species were not observed as abundant in any WWTP. *Ca. M. parvicella* was dominant in the majority of the Danish and global plants, with the average read abundance of $1.7 \pm 2.0\%$ and $0.6 \pm 1.2\%$, respectively. In some plants their abundance exceeded 14% likely causing severe sludge separation problems and deteriorated plant performance. *Ca. M. subdominans* was generally observed at lower abundances, with average and maximum read abundances in Denmark of 1.1 ± 1.3 and 13%, respectively, and average and maximum abundances globally of 0.2 ± 0.5 and 4%, respectively. Both species co-existed in the majority of the plants, indicating a similar ecological niche, with the highest global abundances recorded in a number of European countries (**Figure 1B**). The only other known *Ca. Microthrix* species, *Ca. M. calida*, previously isolated from industrial WWTP (Levantesi et al., 2006), was not observed in any of the plants investigated.

In the Danish nutrient removal plants *Ca. M. parvicella* was the second most abundant genus (**Supplementary Figure 1**), underlining its importance for wastewater treatment processes and floc properties. Long-term survey of Danish plants provided insight into the abundance levels and seasonal dynamics of the two species (**Figure 1A**). In some plants *Ca. Microthrix* was

³<https://github.com/geronimp/enrichM>

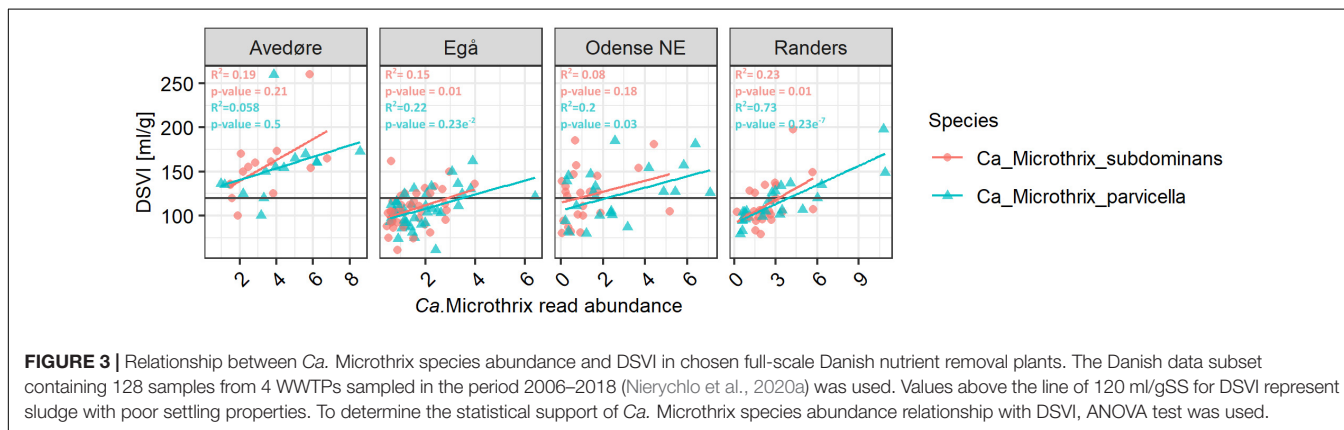
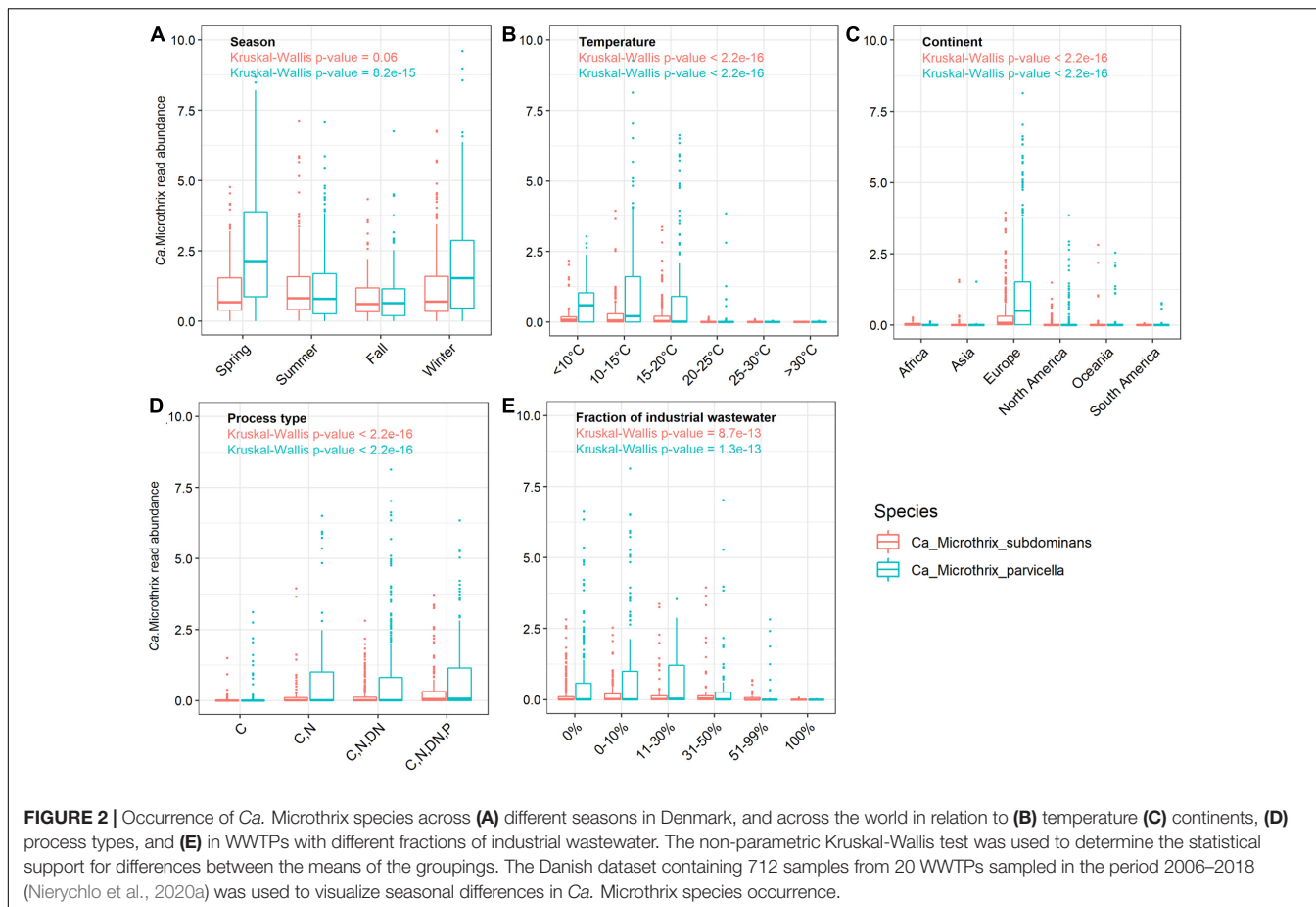


hardly present, in others consistently found over many years. *Ca. M. parvicella* was dominant in most plants. The plants have only minor differences in design and operation, so the mechanisms behind the observed differences are not known. The survey revealed clear seasonal pattern across the plants, visible in particular for *Ca. M. parvicella* (Figures 1A, 2A), with the highest abundance observed during winter (sampled in February) and spring (sampled in May). This is in accordance with earlier observations of its excessive proliferation and bulking in colder seasons (e.g., Kruit et al., 2002; Hug et al., 2006; Mielczarek et al., 2012). Interestingly, *Ca. M. subdominans* did not show a strong seasonal pattern, and had no statistical support in the Danish dataset for differential abundance across the seasons ($p = 0.06$).

The comprehensive global dataset of WWTP community composition enabled a deeper insight into factors affecting the occurrence of the two *Ca. Microthrix* species (Figures 2B–E). Temperature showed a strong effect on the occurrence of both species, with highest abundance observed in WWTPs below 15°C and hardly any *Ca. Microthrix* found in plants with temperature exceeding 20°C (Figure 2B). This observation confirms the general notion that *Ca. Microthrix* thrives at lower temperatures (Knoop and Kunst, 1998; Sheik et al., 2016), which is also reflected by its geographical distribution with the highest abundances of both species observed in Europe (Figure 2C). The process design was important for the prevalence of *Ca. Microthrix*, with highest abundances of

both species in more advanced plants with biological N and P removal (Figure 2D). This supports previous observations of *Ca. Microthrix* preferentially occurring in plants with long sludge retention time (SRT) and aerobic and anoxic stages (Noutsopoulos et al., 2007; Fan et al., 2020) reflecting its low growth rate, the ability to store lipid under anaerobic conditions, and use of nitrate as electron acceptor (Rossetti et al., 2005). The fraction of industrial wastewater, here given as the COD fraction in the influent, was also shown to be important for the occurrence of *Ca. Microthrix* (Figure 2E). Both species were primarily found in municipal plants with low to medium content of industrial wastewater. The reason may be that industrial WWTPs rarely have a high content of lipidic substrates in the influent, which is an important factor promoting growth of *Ca. Microthrix*.

Both species had substantial effects on the settling properties in Danish WWTPs. The diluted sludge volume index (DSVI), which is widely used as a measure of sludge settling properties, showed a positive relationship with the relative abundance of both *Ca. Microthrix* species (Figure 3). Such relationships have been demonstrated before (e.g., Wang et al., 2014a,b; Fan et al., 2018; Nierychlo et al., 2020b) either for the whole genus or for *Ca. M. parvicella*. With the new species-specific FISH probes (see the next section), we could observe that in plants where both species were abundant, they had a similar effect on sludge settling. This was visualized by an almost identical slope



representing the relationship between species read abundance and the corresponding DSVI values (Figure 3).

Phylogenetic Analysis and Visualization of *Ca. Microthrix* Species by FISH

Phylogenetic analysis of the genus *Ca. Microthrix* was conducted using full-length 16S rRNA gene sequences obtained from the MiDAS4 database, SILVA 138 SSURef Nr99, and from MAGs recovered from Danish WWTPs (Figure 4). *Ca.*

Microthrix is represented by 6 distinct species according to the MiDAS4 taxonomy: *parvicella*, *calida*, and several species with provisional *midas_s_x* names, including *midas_s_2* (named here *Ca. Microthrix* *subdominans*). However, only two of the species were abundant globally in WWTPs (Figure 1): *Ca. M. parvicella* and *Ca. M. subdominans*, the latter without a cultured representative or genome.

The first FISH probes targeting morphotype *Microthrix parvicella* were designed more than 20 years ago (Erhart et al., 1997), and additional probes were designed by

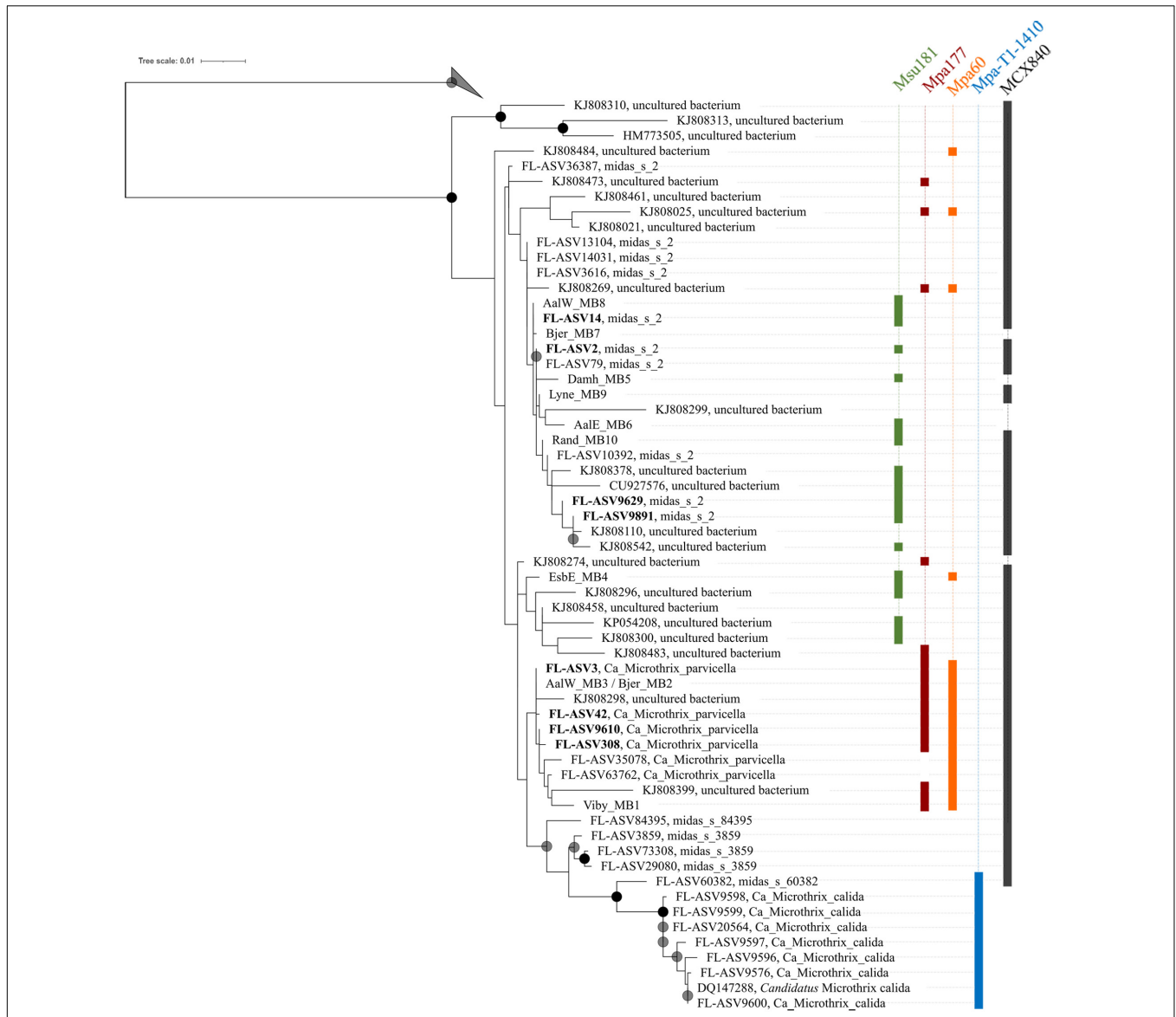


FIGURE 4 | Maximum-likelihood (PhyML) 16S rRNA gene phylogenetic tree of the genus *Ca. Microthrix* and coverage of new and existing FISH probes recommended for use. The tree includes full-length sequences from MiDAS4 database (FL-ASVxxx), MAG (xxxx_MBx_x) and chosen SILVA138 SSURef Nr99 (indicated by sequence accession numbers). FL-ASVs corresponding to the globally abundant ASVs are shown in bold. FL-ASVs representing *Ca. M. subdominans* are shown with its provisional name *midas_s_2* assigned by MiDAS4.8 taxonomy. Coverage of individual FISH probes is shown in panels to the right. A 20% conservational filter was applied to the alignment used for the tree to remove hypervariable positions, resulting in 1204 aligned positions. Bootstrap values from 1000 re-samplings are indicated for branches with > 70% (gray circle), and > 90% (black circle) support. 4 FL-ASVs representing a novel genus from the order IMCC26256 were used as the outgroup. The scale bar represents substitutions per position.

Levantesi et al. (2006) after isolation of *Ca. M. calida*. Their coverage and specificity for targeting individual *Ca. Microthrix* species were tested *in silico* using the MiDAS4 reference database (Table 1) in addition to new probes designed in this study. Probe MPA60 was originally designed to visualize the *M. parvicella* morphotype, and it targets species *Ca. M. parvicella* with perfect coverage and specificity. Probes MPA223 and MPA650 have low specificity, so their use in activated sludge is not recommended. The probe MPA645 served as a good candidate for a genus-level probe, but *in silico* testing and *in situ* hybridization directly in

activated sludge revealed sub-optimal binding resulting in weak FISH signal (Supplementary Figure 2A). The *in silico* test of probe Mpa-T1-1260 to target *Ca. M. calida* (Levantesi et al., 2006) indicated good coverage and specificity, but the use of probe Mpa_all_1410 as a general genus-level probe cannot be recommended due to insufficient coverage.

None of the existing probes targeted the novel *Ca. M. subdominans*, and the general genus-level probes were of insufficient quality, thus three new probes: two species-specific probes and one genus-level probe, were designed and optimized

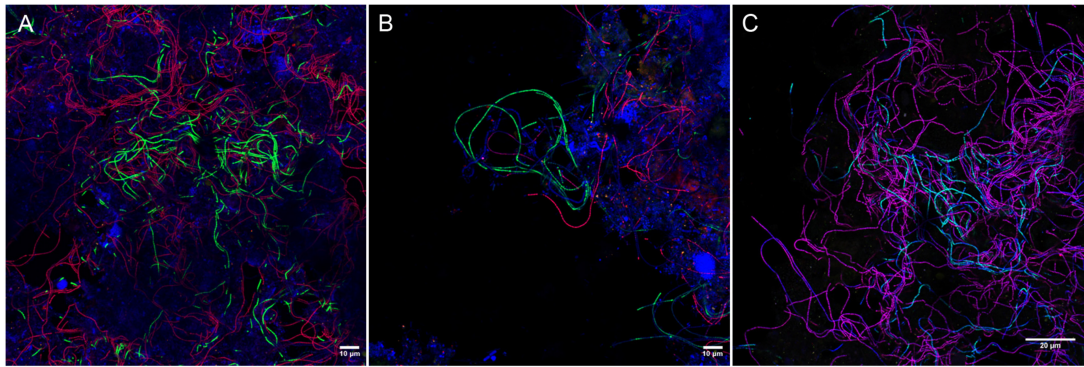


FIGURE 5 | Composite FISH micrographs of *Ca. Microthrix* species in full-scale activated sludge. **(A)** *Ca. M. parvicella* visualized with species-specific probes Mpa177 (Cy3, red), *Ca. M. subdominans* visualized with probe Msu181 (6-FAM, green), other bacteria shown with EUBmix (Cy5, blue); **(B)** *Ca. M. parvicella* visualized with species-specific probe MPA60 (Cy3, red), *Ca. M. subdominans* visualized with probe Msu181 (6-FAM, green), other bacteria shown with EUBmix (Cy5, blue); **(C)** *Ca. M. parvicella* visualized with species-specific probes Mpa177 (Cy3, red) appear magenta, *Ca. M. subdominans* visualized with probe Msu181 (6-FAM, cyan) appear white, *Ca. Microthrix* genus visualized with probe MCX840 (Cy5, blue). Activated sludge was sampled from Randers WWTP.

(Table 1). Since the existing probe MPA60 with 1 mm targets the majority of midas_s_2 sequences, an alternative species-specific probe for *Ca. M. parvicella* was designed. When applied to full-scale activated sludge biomass, all three new probes exclusively hybridized with curly filaments entangled inside, as well as protruding from, the flocs (Figure 5). *Ca. M. parvicella* filaments visualized with probe Mpa177 (Figure 5A) were $0.7 \pm 0.1 \mu\text{m}$ wide and $122 \pm 11 \mu\text{m}$ long, and *Ca. M. subdominans* visualized with probe Msu181 (Figures 5A,B) were $1.0 \pm 0.2 \mu\text{m}$ wide and $92 \pm 19 \mu\text{m}$ long. No overlap between the two species-specific probes was observed, confirming their specificity *in situ*, and both showed a good overlap with the genus-level probe MCX840 (Figure 5C). Probe Mpa177 targeting *Ca. M. parvicella* ensures high specificity as it does not target other *Ca. Microthrix* species even with one mismatch. However, *in situ* tests performed using the existing *Ca. M. parvicella* probe MPA60 also proved its high specificity, without overlap with *Ca. M. subdominans* (Figure 5B). Mpa177 and MPA60 showed very good overlap (Supplementary Figure 2B), thus both probes can be applied interchangeably, or with different fluorochromes to confirm specific coverage of *Ca. M. parvicella*, or together as a mix to give higher fluorescence signal. Helper probes are recommended with all the new probes to ensure strong and even fluorescence signal along the filaments.

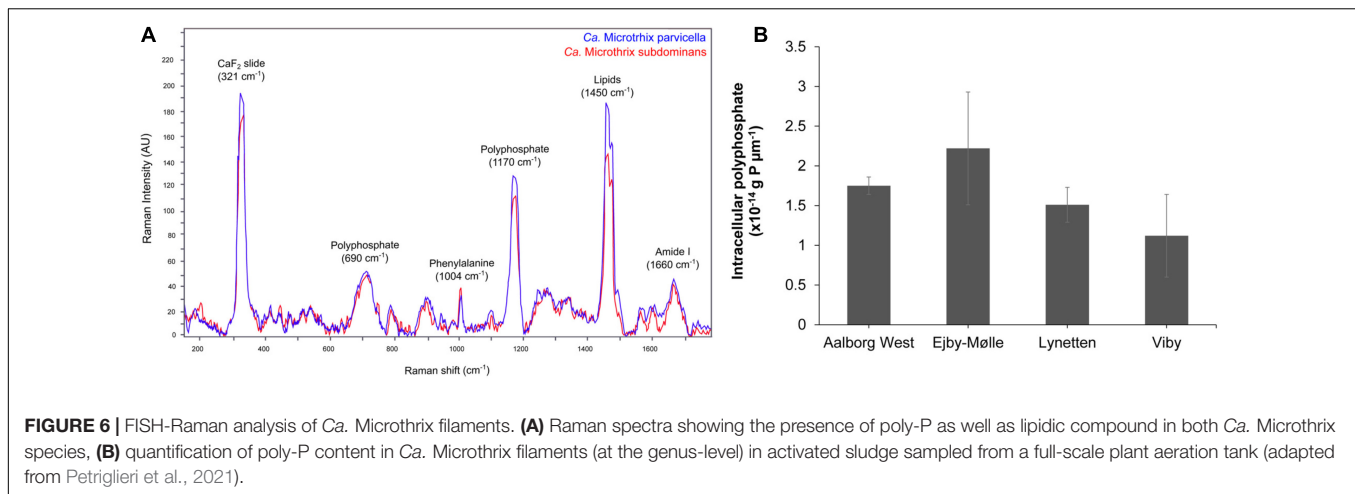
Quantitative FISH analyses with all new probes showed that the biovolume fractions of the biomass were similar or slightly higher than the read abundance by amplicon sequencing using the V1-V3 primers (Table 2). The amplicon abundance estimation for *Ca. Microthrix* depended strongly on primers used, with notable divergence between the two most commonly used primer sets in microbial community analyses. We observed that *Ca. Microthrix* abundance is underestimated by the factor of 3 by the V4 primer set, compared to V1-V3 primer set (Supplementary Figure 3). This is of importance when comparing abundance estimates by the two methods and when the potential effect on settling properties is evaluated using amplicon sequencing data.

High sequence similarity among the species in genus *Ca. Microthrix* makes it difficult to construct robust phylogenetic trees, resulting in low bootstrap values (Figure 4). Consequently, the exact placement of the individual sequences, especially those clustering with the two most abundant species, may not be definitive. Not all full-length 16S rRNA gene sequences coming from the MAGs or retrieved from the SILVA database are targeted by the new or existing FISH probes (e.g., the MAG Bjer_MB7 representing *Ca. M. subdominans* or KJ808299). However, it should be kept in mind that the MAGs recovered may not always represent the abundant members of the microbial community, and sequences retrieved from SILVA are clustered at 97% similarity, both of which may result in the *in silico* observed lack of probe coverage. All the MAG-derived sequences

TABLE 2 | *Ca. Microthrix* abundance estimation using amplicon sequencing and qFISH.

FISH Probe (target) WWTP	Date	Abundance [%]	
		Amplicon reads*	qFISH
Mpa177 (<i>Ca. M. parvicella</i>)			
Randers	February 2017	10.9	19.3 (± 5.9)
Viborg	February 2010	3.6	5.0 (± 2.0)
Kolding	February 2008	7	8.9 (± 4.8)
Msu181 (<i>Ca. M. subdominans</i>)			
Randers	February 2017	4.2	5.8 (± 2.2)
Viborg	February 2010	4.3	4.9 (± 1.3)
Kolding	February 2008	4.2	4.2 (± 2.5)
MCX840 (genus <i>Ca. Microthrix</i>)			
Randers	February 2017	15.1	22.9 (± 7.0)
Viborg	February 2010	7.9	15.5 (± 3.5)
Kolding	February 2008	11.2	16.2 (± 6.1)

*Standard deviation of amplicon sequencing is approx. 10%.



clustered either with *Ca. M. parvicella* or *Ca. M. subdominans*, except EsbE_MB4, which formed a separate cluster with several SILVA sequences. This particular sequence was targeted by *Ca. M. parvicella*-specific probe (MPA60) and the *Ca. M. subdominans*-specific probe (Msu181).

Detection and Quantification of Storage Polymers in *Ca. Microthrix*

Application of the species-specific FISH probes with Raman microspectroscopy revealed the presence of biological peaks for nucleic acids (784 cm^{-1}), phenylalanine ($1,004\text{ cm}^{-1}$), lipids ($1,450\text{ cm}^{-1}$), amide I linkages of proteins ($1,660\text{ cm}^{-1}$), as well as poly-P storage ($1,170\text{ cm}^{-1}$) (Schuster et al., 2000; Huang et al., 2004) in both *Ca. M. parvicella* and *Ca. M. subdominans* (Figure 6A). No other storage polymers pertinent to the PAO metabolism (glycogen or PHA) were detected *in situ* under aerobic or anaerobic conditions. Interestingly, the peak assigned to lipids ($1,450\text{ cm}^{-1}$) was significantly higher than similar peaks observed in other PAOs (Fernando et al., 2019), supporting the prevalence of lipid metabolism suggested by the metabolic potential.

The amount of stored poly-P in *Ca. Microthrix* filaments was quantified using FISH-Raman microspectroscopy in fresh activated sludge from four full-scale EBPR plants (Figure 6B). The content of poly-P varied slightly among the different plants, with the highest value measured in Ejby-Mølle ($2.22 \times 10^{-14}\text{ g P } \mu\text{m}^{-1}$) and the lowest in Viby ($1.12 \times 10^{-14}\text{ g P } \mu\text{m}^{-1}$). Even though the *Ca. Microthrix* trichome width was relatively small, and thus also the P-content per μm of the filament, it can substantially contribute to P removal in the plants as the filament length often exceeds $100\text{ } \mu\text{m}$, and the genus often present at high abundances (Petriglieri et al., 2021).

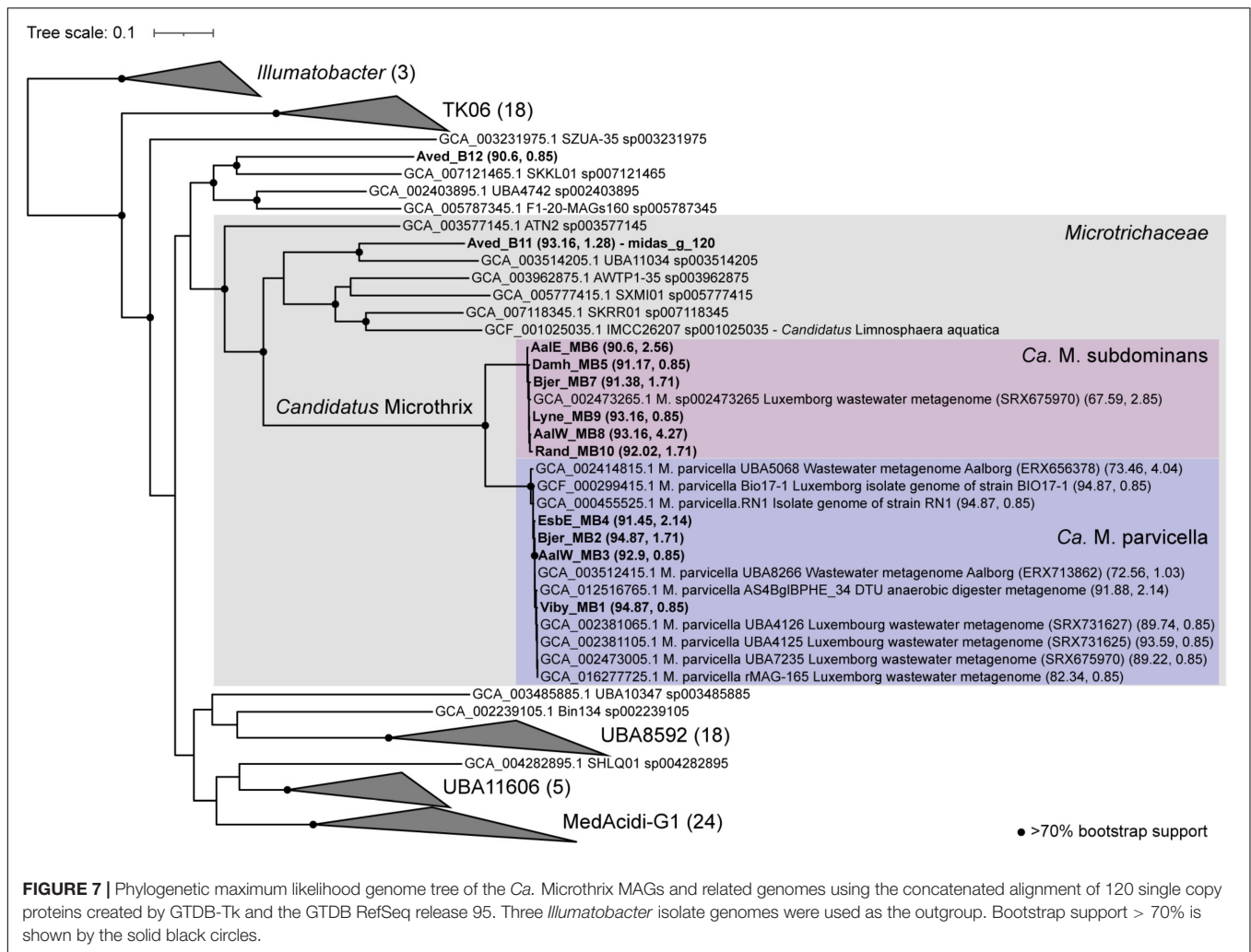
In order to investigate whether *Ca. Microthrix* had a dynamic poly-P uptake and release in anaerobic- aerobic cycles as are typical in EBPR plants, we carried out a number of experiments with fresh activated sludge with a high content of *Ca. M. parvicella* and *Ca. M. subdominans* (Supplementary Note 2 and Supplementary Table 3). None of the substrates tested (mixture acetate-glucose-casamino acids or oleic acid) induced anaerobic

depletion of intracellular poly-P after 3 h without oxygen or nitrate present. This supports previous observations (Andreasen and Nielsen, 2000) that *Ca. Microthrix* does not have a dynamic P-cycling similar to PAOs.

Ca. Microthrix Genome Recovery and Metabolic Reconstruction

Eleven high-quality MAGs from the family Microtrichaceae were recovered from across 23 Danish WWTPs (Singleton et al., 2021). Ten of these MAGs were from the genus *Ca. Microthrix*, forming two different species (95% ANI clustering), *Ca. M. parvicella* (4 MAGs) and *Ca. M. subdominans* (6 MAGs) (Figure 7). One MAG (Aved_B11) belonged to a novel and undescribed genus with the provisional name *midas_g_120*. These two genera, as well as genus IMCC26207 represented by the genome of *Candidatus Limnosphaera aquatica* strain IMCC26207 (Kim et al., 2017), were found to be abundant across global WWTPs (Supplementary Figure 4).

The number of coding sequence (CDS) regions for *Ca. Microthrix* MAGs ranged between 3915 and 4743, and completeness and contamination ranged between 90.6–94.87 and 0.85–4.27%, respectively (Supplementary Data 1). To date, only one medium quality MAG (GCA_002473265, 62% completeness), recovered from a WWTP metagenome in Luxembourg (SAMN02862044), exists for *Ca. M. subdominans* in the GTDB database (Parks et al., 2017). The four *Ca. M. parvicella* MAGs are > 99% ANI to each other (alignment coverage > 79%; Supplementary Figure 5), representing near identical lineages despite being recovered from separate WWTPs. The six *Ca. M. subdominans* MAGs had ANIs of > 98% (alignment coverage > 74%) to each other, suggesting a similar lack of diversity (Supplementary Figure 5). This supports previous findings suggesting lower genomic diversity between *Ca. Microthrix* lineages when compared to other common genera in WWTPs, such as *Ca. Accumulibacter* (McIlroy et al., 2013). Additionally, the analyses of single-nucleotide polymorphisms in *Ca. Microthrix* spp. has indicated low population-level diversity (Muller et al., 2014a).



As expected based on the high similarity of the *Ca. M. parvicella* MAGs to the previously recovered genomes, metabolic reconstruction revealed similar metabolisms to the published models. The *Ca. M. parvicella* MAGs encoded the EMP pathway, TCA cycle, pentose phosphate pathway, as well as lipid activation and oxidation capabilities, high-affinity phosphate transporters, potential triacylglycerol (TAG) storage, and the potential to reduce nitrite and nitrate (McIlroy et al., 2013). Unlike *Ca. M. parvicella*, the metabolic potential of *Ca. M. subdominans* is currently unknown. The pangenome of *Ca. Microthrix* based on the orthologous proteins shared between the two species was large (11,765 gene families, **Supplementary Table 1**). However, the core genome of 1,513 gene families showed similar metabolic capabilities to *Ca. M. parvicella*. Numerous acyl-CoA ligases (K00666, K01897, K03822, up to 8 copies per genome) and enoyl-CoA hydratases (K01692, K01715, up to 17 copies per genome) were detected in every MAG, supporting lipid use as a key metabolic strategy (**Figure 8**). Diacylglycerol O-acyltransferase / wax synthase (K00635, **Figure 8**) genes (3–5 copies per MAG) were identified in the *Ca. M. subdominans* MAGs, supporting the hypothesized use of TAG as an energy storage compound

(McIlroy et al., 2013). While PHB storage and use has been suggested in previous work (Muller et al., 2012), we could only identify copies of the *phaA* gene. No copies of *phaB* or the key gene, PHA synthase (*phaC*), were detected in the KO analysis (**Figure 8**). Absence of these genes supports the Raman microspectroscopy analyses above showing no PHA storage. The potential for nitrate (putative genes identified in McIlroy et al., 2013) and nitrite reduction (*nirK*, K00368) in anaerobic respiration was also identified in *Ca. M. subdominans* (**Figure 8**). Genes for further reduction to N₂O or N₂ were not detected. The phosphate metabolism genes were highly similar between the two species. Genes for the high affinity phosphate transporter (*pstSCAB*) and transport regulator (*phoU*) were detected in *Ca. M. subdominans* (**Figure 8**). Similar to *Ca. M. parvicella*, the low affinity Pit transport system, often used as a likely indicator gene for polyphosphate accumulating organisms such as *Tetrasphaera* and *Ca. Accumulibacter* (McIlroy et al., 2014; Oyserman et al., 2016), was not detected. No acetate transporter genes were identified in the genomes, though genes for the transport of branched chain amino acids were detected (*livGFHMK*, K01995-K01999). While the main metabolisms were

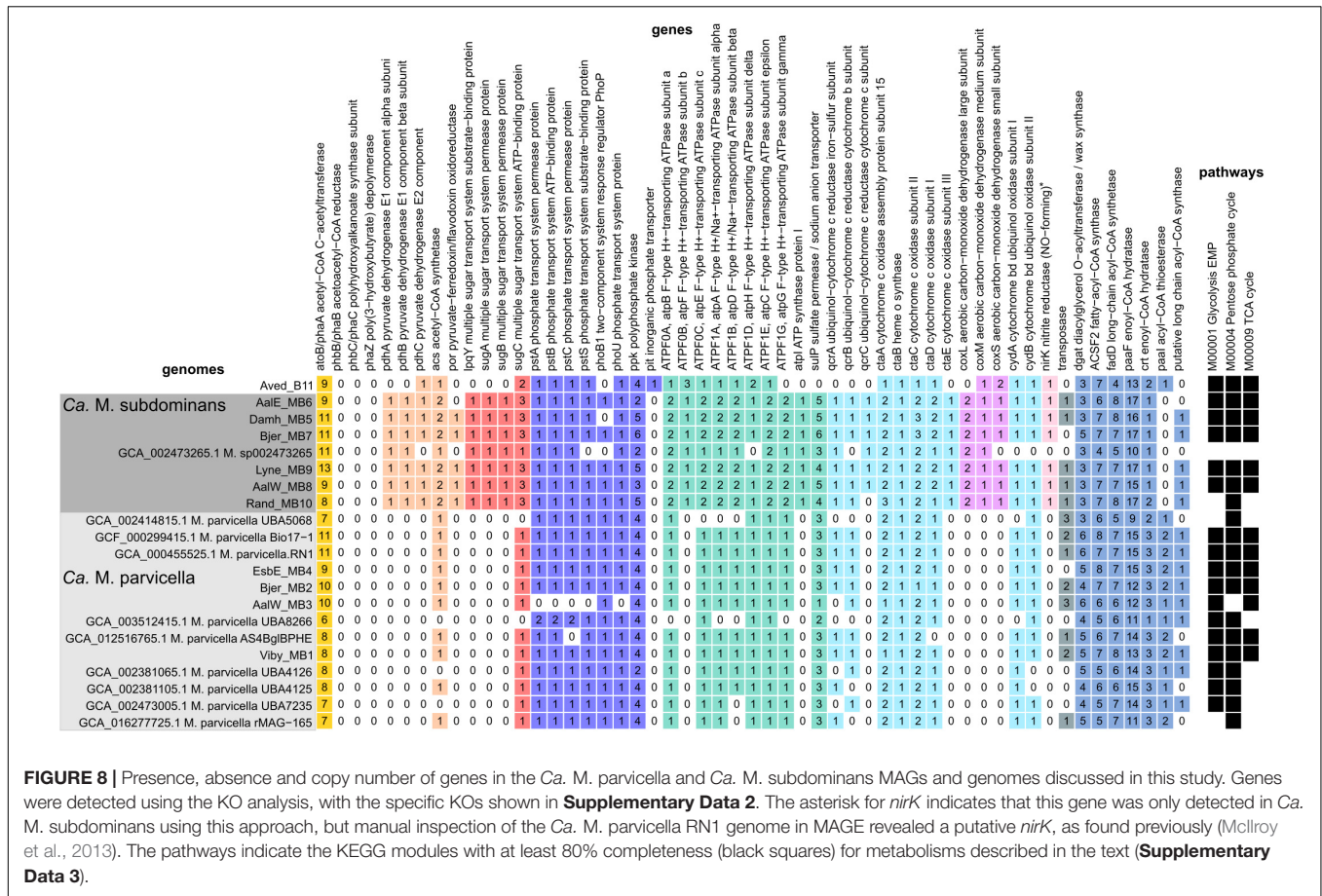


FIGURE 8 | Presence, absence and copy number of genes in the *Ca. M. parvicella* and *Ca. M. subdominans* MAGs and genomes discussed in this study. Genes were detected using the KO analysis, with the specific KOs shown in **Supplementary Data 2**. The asterisk for *nirK* indicates that this gene was only detected in *Ca. M. subdominans* using this approach, but manual inspection of the *Ca. M. parvicella* RN1 genome in MAGE revealed a putative *nirK*, as found previously (McIlroy et al., 2013). The pathways indicate the KEGG modules with at least 80% completeness (black squares) for metabolisms described in the text (**Supplementary Data 3**).

very similar between the two species, a subset of genes comprised the species-specific core metabolisms of *Ca. M. parvicella* (380) and *Ca. M. subdominans* (364) (**Supplementary Table 1**).

Genes unique to *Ca. M. subdominans* indicated a slightly different niche to *Ca. M. parvicella*. Firstly, this lineage may be able to use atmospheric levels of carbon monoxide as a supplementary energy source, as it appears to encode several genes for a putative form II aerobic carbon monoxide dehydrogenase (CODH, K03518, K03519, K03520, *coxLMS*) (**Figure 8; Supplementary Note 1**). This could be an adaptation of *Ca. M. subdominans* to its upstream habitats, most likely soil and water systems (King, 2003; King and Weber, 2007). An ABC sugar transporter unique to *Ca. M. subdominans* was also identified (**Figure 8**). The four genes comprising the ABC transporter operon (*lpqY*, *sugABC*, K02027, K02025, K02026, K10112) are most closely related to those found in the ammonia oxidizing Gammaproteobacteria, *Nitrosococcus*, and *Nitrosomonas* (**Supplementary Data 4**). These genes could facilitate the transport of trehalose, as has been shown in *Mycobacterium tuberculosis* (Kalscheuer et al., 2010), which encodes an annotated trehalose transport system permease protein of 56.2% AAI (SugA) and 65.4% AAI (SugB) to the *Ca. M. subdominans* proteins. This transporter could perhaps provide *Ca. M. subdominans* with additional substrate options compared to *Ca. M. parvicella*. Trehalose synthesis and hydrolysis genes are

present in both species, but transporter genes are absent in the *Ca. M. parvicella* MAGs, including *Ca. M. parvicella* RN1 (McIlroy et al., 2013). This compound could be used as a mechanism for managing environmental stress, including low temperatures, or serve as an energy storage compound (Elbein et al., 2003).

While *Ca. M. parvicella* prefers microaerobic conditions based on growth experiments (Rossetti et al., 2005), *Ca. M. subdominans* may be better adapted to oxic environments based on the genomic potential. Both *Ca. M. subdominans* and *Ca. M. parvicella* have the genomic potential to use nitrate or nitrite as a terminal electron acceptor under anoxic conditions as well as the microaerobic cytochrome *bd* oxidase and aerobic cytochrome *c aa3* oxidase. However, *Ca. M. subdominans* encodes an additional *bc1-aa3* type cytochrome super complex (*qcrABC*, *ctaCDEF*) in one syntenic region (**Figure 8**). Specifically, the QcrC component is a small di-heme cytochrome *c*, and has been found to be the only *c*-type cytochrome in most Actinobacteria (Kao et al., 2016), although both *Ca. M. parvicella* and *Ca. M. subdominans* also encode an additional mono-heme cytochrome *c* domain containing protein. This super complex is widespread in aerobic Actinobacteria, and is considered more bioenergetically efficient under aerobic conditions, enabling a direct link between menaquinol oxidation and dioxygen reduction (Kao et al., 2016; Davoudi et al., 2019). If these metabolisms are correct and active, *Ca. M. subdominans* may

grow better in oxic conditions compared to the microaerophile *Ca. M. parvicella*.

Ecophysiology of *Ca. Microthrix* and the Importance for Full-Scale WWTPs Performance

Ca. Microthrix has been known for decades by researchers and practitioners in the field, and *Ca. M. parvicella* has been considered the model organism for this genus. Our study has shown that there are only two abundant species present in municipal activated sludge plants across the world, and that an undescribed species, *Ca. Microthrix* subdominans, is usually present along with *Ca. M. parvicella*, but is generally less abundant. The abundance of both species may be strongly underestimated by using the commonly applied primer V4, compared to abundances obtained by the V1-V3 primers or by FISH, so the use of the V1-V3 primer set for future amplicon studies is highly recommended. Both species had very similar morphology and seem to impair wastewater settling properties, if abundant. *Ca. Microthrix* can also cause foaming due to its hydrophobic cell surface (de los Reyes, 2010). Foaming is, however, uncommon in Danish WWTPs, so we could not investigate whether the two species had different foaming properties.

The low level of variation in metabolic potential between the core genomes of *Ca. M. parvicella* and *Ca. M. subdominans* suggests that both species occupy a very similar niche in the WWTP ecosystem, yet unknown factors dictate the differences in their abundance and seasonal variation. This may result from the subtle differences in their genetic makeup or unknown conditions upstream of the WWTPs. Some detected genomic differences were related to the presence of a sugar transporter in *Ca. M. subdominans*, and the potential to use CO as an additional energy source, which may increase the chance for long-term survival during carbon starvation when organic carbon is not available (Cordero et al., 2019).

Wastewater treatment plants with nutrient removal promote the growth of *Ca. Microthrix* as indicated by our global survey. Such plants have dynamic oxic and anoxic conditions, which are selective for bacteria capable of storing carbon anaerobically, are microaerophilic, and can use nitrate and/or nitrite as electron acceptors, such as *Ca. Microthrix*. *Ca. Microthrix* is specialized in the uptake and storage of LCFAs, particularly under anoxic conditions, which allows them to successfully grow in the aerobic tanks with little external substrate present (Andreasen and Nielsen, 2000; Muller et al., 2014a; Roume et al., 2015). The storage compound present in *Ca. M. parvicella* has been suggested to be PHA in isolates and *in situ* based on Nile Blue A staining (Tandoi et al., 1998; Rossetti et al., 2002), while annotations of the two available genomes have yielded contradicting results (Muller et al., 2012; McIlroy et al., 2013). We verified that none of the *Ca. Microthrix* MAGs had the full set of genes for PHA storage, and that neither of the species possessed PHA *in situ*, as analyzed by Raman microspectroscopy. Instead we found the potential for TAG storage and observed an unspecified lipid-peak indicating storage of a lipid compound

in both species, in agreement with our previous observations using isotope-labeled substrates (Nielsen et al., 2002). The results clearly indicate that both species of *Ca. Microthrix* are lipid accumulating bacteria, perhaps making them interesting targets for lipid recovery and subsequent biofuel conversion (Muller et al., 2014b).

Ca. M. parvicella is likely microaerophilic (Andreasen and Nielsen, 2000) as supported by our MAGs, while the additional cytochrome oxidase found in *Ca. M. subdominans* may potentially make this species less sensitive to higher oxygen levels. There is experimental evidence for the use of nitrate as an electron acceptor (Hesselsoe et al., 2005), although the annotations of the *nar* genes in the MAGs are ambiguous (McIlroy et al., 2013). In contrast, the *nirK* gene was more confidently annotated in all MAGs, so the details regarding *Ca. Microthrix* activity under anoxic conditions are still uncertain.

Poly-P granules have been commonly observed in *Ca. Microthrix* (e.g., Tandoi et al., 1998; Wang et al., 2014a,b), and their presence was confirmed *in situ* in both species. *Ca. Microthrix* could be a PAO, but the lack of active cycling of poly-P during anaerobic/aerobic phases, and the absence of a pit transporter across all the MAGs, indicates a less dynamic role in P removal. Instead, the poly-P reserves may be used for maintenance purposes as the granules are slowly depleted during longer starvation periods (Rossetti et al., 2002).

The global survey clearly demonstrated higher abundances of both species in countries with colder climates. This agrees with several studies (e.g., Knoop and Kunst, 1998; Sheik et al., 2016; Fan et al., 2017) and may be due to an increased availability of lipids to *Ca. Microthrix* compared to other species at low temperatures because of their hydrophobic surface as well as surface-associated lipases (Rossetti et al., 2005). Our study did not reveal obvious new ecophysiological traits that can be applied for the design of efficient control measures, but it provides a complete set of FISH probes and genomic information for future research into *in situ* gene expression and regulation, an important step toward understanding the ecology of *Ca. Microthrix* in WWTPs.

ETYMOLOGY

Description of “*Candidatus Microthrix subdominans*” sp. nov. (midas_s_2) “*Candidatus Microthrix subdominans*,” (sub.do`mi.nans. L. prep *sub* below; L. pres. part. *dominans* dominant; N.L. part. adj. *subdominans* indicating the abundance of this organism often below the dominant species *Ca. Microthrix parvicella*). This taxon was represented by the MAG Lyne_MB9. The complete protologue can be found in **Supplementary Table 2**.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: sequencing data from Danish WWTPs is available at the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under the project number PRJNA622675.

AUTHOR CONTRIBUTIONS

MN and PN devised the study and its main conceptual ideas. MN, CS, and MD performed data analysis with contributions from PN. FP and JP performed Raman measurements and data analysis. LT and ZK performed FISH analyses. MP and JP performed P-release lab experiments. The manuscript was drafted by MN and CS with contributions from FP, and revised by PN. All authors contributed to the article and approved the submitted version.

FUNDING

This project was funded by the VILLUM Foundation (16578) and 20 Danish wastewater treatment plants.

REFERENCES

- Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., and Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* 56, 1919–1925.
- Andersen, K. S., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* [Preprint] doi: 10.1101/299537bioRxiv:299537
- Andreasen, K., and Nielsen, P. H. (1998). *In situ* characterization of substrate uptake by *Microthrix parvicella* using microautoradiography. *Water Sci. Technol.* 37, 19–26. doi: 10.2166/wst.1998.0571
- Andreasen, K., and Nielsen, P. H. (2000). Growth of *Microthrix parvicella* in nutrient removed activated sludge plants: studies of *in situ* physiology. *Water Res.* 34, 1559–1569. doi: 10.1016/S0043-1354(99)00319-X
- Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753
- Blackall, L. L., Stratton, H., Bradford, D., Del Dot, T., Sjørup, C., Seviour, E. M., et al. (1996). “*Candidatus Microthrix parvicella*,” a filamentous bacterium from activated sludge sewage treatment plants. *Int. J. Syst. Bacteriol.* 46, 344–346. doi: 10.1099/00207713-46-1-344
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Cordero, P. R. F., Bayly, K., Man Leung, P., Huang, C., Islam, Z. F., Schittenhelm, R. B., et al. (2019). Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J.* 13, 2868–2881. doi: 10.1038/s41396-019-0479-8
- Daims, H., Brühl, A., Amann, R., Schleifer, K.-H., and Wagner, M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* 22, 434–444. doi: 10.1016/S0723-2020(99)80053-8
- Daims, H., Lückner, S., and Wagner, M. (2006). *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* 8, 200–213. doi: 10.1111/j.1462-2920.2005.00880.x
- Davoudi, C.-F., Ramp, P., Baumgart, M., and Bott, M. (2019). Identification of Surf1 as an assembly factor of the cytochrome bc1-aa3 supercomplex of Actinobacteria. *Biochim. Biophys. Acta BBA Bioenerg.* 1860:148033. doi: 10.1016/j.bbabi.2019.06.005
- de los Reyes, F. L. (2010). “Foam in wastewater treatment facilities,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed. K. N. Timmis (Berlin: Springer), 2401–2411. doi: 10.1007/978-3-540-77587-4_176
- Dueholm, M. S., Andersen, K. S., McIlroy, S. J., Kristensen, J. M., Yashiro, E., Karst, S. M., et al. (2020). Generation of comprehensive ecosystems-specific reference databases with species-level resolution by high-throughput full-length 16S rRNA gene sequencing and automated taxonomy assignment (AutoTax). *mBio* 11:e01557-20. doi: 10.1128/mBio.01557-20
- Dueholm, M. S., Nierychlo, M., Andersen, K. S., Rudkjøbing, V., Knudsen, S., The MiDAS Global Consortium, et al. (2021). MiDAS 4—a global WWTP ecosystem-specific full-length 16S rRNA gene catalogue and taxonomy for studies of bacterial communities. *BioRxiv*
- Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* [Preprint] doi: 10.1101/081257bioRxiv:081257
- Edgar, R. C. (2018). Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ* 6:e4652. doi: 10.7717/peerj.4652
- Eikelboom, D. H., and van Buijsen, H. J. J. (1983). *Microscopic Sludge Investigation Manual*. Delft: TNO Research Institute for Environmental Hygiene.
- Elbein, A. D., Pan, Y. T., Pastuszak, I., and Carroll, D. (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology* 13, 17R–27R. doi: 10.1093/glycob/cwg047
- Erhart, R., Bradford, D., Seviour, R. J., Amann, R., and Blackall, L. L. (1997). Development and use of fluorescent *in situ* hybridization probes for the detection and identification of “*Microthrix parvicella*” in activated sludge. *Syst. Appl. Microbiol.* 20, 310–318.
- Fan, N., Qi, R., Rossetti, S., Tandoi, V., Gao, Y., and Yang, M. (2017). Factors affecting the growth of *Microthrix parvicella*: batch tests using bulking sludge as seed sludge. *Sci. Total Environ.* 609, 1192–1199. doi: 10.1016/j.scitotenv.2017.07.261
- Fan, N., Wang, R., Qi, R., Gao, Y., Rossetti, S., Tandoi, V., et al. (2018). Control strategy for filamentous sludge bulking: bench-scale test and full-scale application. *Chemosphere* 210, 709–716. doi: 10.1016/j.chemosphere.2018.07.028
- Fan, N., Yang, M., Rossetti, S., Levantesi, C., and Qi, R. (2019). Monitoring, isolation and characterization of *Microthrix parvicella* strains from a Chinese wastewater treatment plant. *Water Sci. Technol.* 79, 1406–1416. doi: 10.2166/wst.2019.136
- Fan, N.-S., Qi, R., Huang, B.-C., Jin, R.-C., and Yang, M. (2020). Factors influencing *Candidatus Microthrix parvicella* growth and specific filamentous bulking control: a review. *Chemosphere* 244:125371. doi: 10.1016/j.chemosphere.2019.125371
- Fei, X., Li, W., Wang, C., Jiao, X., and Zhang, X. (2021). Simulation and experimental study of fluorescence labeled polyphosphate in *Microthrix parvicella*. *J. Mol. Graph. Model.* 104:107842. doi: 10.1016/j.jmkgm.2021.10.7842
- Fernando, E. Y., McIlroy, S. J., Nierychlo, M., Herbst, F.-A., Petriglieri, F., Schmid, M. C., et al. (2019). Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman–FISH. *ISME J.* 13, 1933–1946. doi: 10.1038/s41396-019-0399-7

ACKNOWLEDGMENTS

This project was funded by the VILLUM Foundation (16578) and 20 Danish wastewater treatment plants. The LABGeM (CEA/Genoscope and CNRS UMR 8030), the France Génomique and French Bioinformatics Institute National Infrastructures (funded as part of Investissement d’Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.690251/full#supplementary-material>

- Hesselsoe, M., Nielsen, J. L., Roslev, P., and Nielsen, P. H. (2005). Isotope labeling and microautoradiography of active heterotrophic bacteria on the basis of assimilation of $^{14}\text{CO}_2$. *Appl. Environ. Microbiol.* 71, 646–655. doi: 10.1128/AEM.71.2.646-655.2005
- Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J., and Whiteley, A. S. (2004). Raman microscopic analysis of single microbial cells. *Anal. Chem.* 76, 4452–4458. doi: 10.1021/ac049753k
- Hug, T., Gujer, W., and Siegrist, H. (2006). Modelling seasonal dynamics of *Microthrix parvicella*. *Water Sci. Technol.* 54, 189–198. doi: 10.2166/wst.2006.387
- Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G. S., and Jacobs, W. R. (2010). Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21761–21766. doi: 10.1073/pnas.1014642108
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Kao, W.-C., Kleinschroth, T., Nitschke, W., Baymann, F., Neehaul, Y., Hellwig, P., et al. (2016). The obligate respiratory supercomplex from Actinobacteria. *Biochim. Biophys. Acta BBA Bioenerg.* 1857, 1705–1714. doi: 10.1016/j.bbabi.2016.07.009
- Karst, S. M., Dueholm, M. S., McIlroy, S. J., Kirkegaard, R. H., Nielsen, P. H., and Albertsen M. (2018). Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* doi: 10.1038/nbt.4045
- Kim, S., Kang, I., and Cho, J.-C. (2017). Genomic analysis of a freshwater actinobacterium, *Candidatus Limnosphaera aquatica* strain IMCC26207, isolated from Lake Soyang. *J. Microbiol. Biotechnol.* 27, 825–833. doi: 10.4014/jmb.1701.01047
- Kindaichi, T., Nierychlo, M., Kragelund, C., Nielsen, J. L., and Nielsen, P. H. (2013). High and stable substrate specificities of microorganisms in enhanced biological phosphorus removal plants. *Environ. Microbiol.* 15, 1821–1831. doi: 10.1111/1462-2920.12074
- King, G. M. (2003). Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by Mycobacteria†. *Appl. Environ. Microbiol.* 69, 7266–7272. doi: 10.1128/AEM.69.12.7266-7272.2003
- King, G. M., and Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5, 107–118. doi: 10.1038/nrmicro1595
- Knoop, S., and Kunst, S. (1998). Influence of temperature and sludge loading on activated sludge settling, especially on *Microthrix parvicella*. *Water Sci. Technol.* 37, 27–35. doi: 10.1016/S0273-1223(98)00080-8
- Kragelund, C., Remesova, Z., Nielsen, J. L., Thomsen, T. R., Eales, K., Seviour, R., et al. (2007). Ecophysiology of mycolic acid-containing Actinobacteria (Mycolata) in activated sludge foams. *FEMS Microbiol. Ecol.* 61, 174–184. doi: 10.1111/j.1574-6941.2007.00324.x
- Kruit, J., Hulsbeek, J., and Visser, A. (2002). Bulking sludge solved?! *Water Sci. Technol.* 46, 457–464. doi: 10.2166/wst.2002.0517
- Lane, D. J. (1991). “16S/23S rRNA sequencing,” in *Nucleic Acid Techniques in Bacterial Systematics*, eds E. Stackebrandt and M. Goodfellow (New York, NY: John Wiley and Sons), 115–175.
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Levantesi, C., Rossetti, S., Thelen, K., Kragelund, C., Krooneman, J., Eikelboom, D., et al. (2006). Phylogeny, physiology and distribution of “*Candidatus Microthrix calida*”, a new *Microthrix* species isolated from industrial activated sludge wastewater treatment plants. *Environ. Microbiol.* 8, 1552–1563. doi: 10.1111/j.1462-2920.2006.01046.x
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293
- Martins, A. M. P., Pagilla, K., Heijnen, J. J., and van Loosdrecht, M. C. M. (2004). Filamentous bulking sludge—a critical review. *Water Res.* 38, 793–817. doi: 10.1016/j.watres.2003.11.005
- McIlroy, S. J., Albertsen, M., Andresen, E. K., Saunders, A. M., Kristiansen, R., Stokholm-Bjerregaard, M., et al. (2014). ‘*Candidatus Competibacter*’-lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *ISME J.* 8, 613–624. doi: 10.1038/ismej.2013.162
- McIlroy, S. J., Kristiansen, R., Albertsen, M., Michael Karst, S., Rossetti, S., Lund Nielsen, J., et al. (2013). Metabolic model for the filamentous ‘*Candidatus Microthrix parvicella*’ based on genomic and metagenomic analyses. *ISME J.* 7, 1161–1172. doi: 10.1038/ismej.2013.6
- McIlroy, S. J., Saunders, A. M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A. A., et al. (2015). MiDAS: the field guide to the microbes of activated sludge. *Database* 2015:bav062. doi: 10.1093/database/bav062
- Mielczarek, A. T., Kragelund, C., Eriksen, P. S., and Nielsen, P. H. (2012). Population dynamics of filamentous bacteria in Danish wastewater treatment plants with nutrient removal. *Water Res.* 46, 3781–3795. doi: 10.1016/j.watres.2012.04.009
- Muller, E. E. L., Pinel, N., Gillece, J. D., Schupp, J. M., Price, L. B., Engelthaler, D. M., et al. (2012). Genome sequence of “*Candidatus Microthrix parvicella*” Bio17-1, a long-chain-fatty-acid-accumulating filamentous Actinobacterium from a biological wastewater treatment plant. *J. Bacteriol.* 194, 6670–6671. doi: 10.1128/JB.01765-12
- Muller, E. E. L., Pinel, N., Laczny, C. C., Hoopmann, M. R., Narayanasamy, S., Lebrun, L. A., et al. (2014a). Community-integrated omics links dominance of a microbial generalist to fine-tuned resource usage. *Nat. Commun.* 5:5603. doi: 10.1038/ncomms6603
- Muller, E. E. L., Sheik, A. R., and Wilmes, P. (2014b). Lipid-based biofuel production from wastewater. *Curr. Opin. Biotechnol.* 30, 9–16. doi: 10.1016/j.copbio.2014.03.007
- Muyzer, G., de Waal, E. C., and 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.
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nielsen, J. L. (2009). “Protocol for fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotides,” in *FISH Handbook for Biological Wastewater Treatment: Identification and Quantification of Microorganisms in Activated Sludge and Biofilms by FISH*, eds P. H. Nielsen, H. Daims, and H. Lemmer (New York, NY: IWA Publishing), 73–84.
- Nielsen, P. H., Roslev, P., Dueholm, T. E., and Nielsen, J. L. (2002). *Microthrix parvicella*, a specialized lipid consumer in anaerobic-aerobic activated sludge plants. *Water Sci. Technol.* 46, 73–80.
- Nierychlo, M., Andersen, K. S., Xu, Y., Green, N., Jiang, C., Albertsen, M., et al. (2020a). 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. *Water Res.* 182:115955. doi: 10.1016/j.watres.2020.115955
- Nierychlo, M., McIlroy, S. J., Kucheryavskiy, S., Jiang, C., Ziegler, A. S., Kondrotaitė, Z., et al. (2020b). *Candidatus Amarolinea* and *Candidatus Microthrix* are mainly responsible for filamentous bulking in Danish municipal wastewater treatment plants. *Front. Microbiol.* 11:1214. doi: 10.3389/fmicb.2020.01214
- Noutsopoulos, C., Mamais, D., and Andreadakis, A. (2007). Effect of solids retention time on *Microthrix parvicella* growth. *Water S. A.* 32, 315–321. doi: 10.4314/wsa.v32i3.5276
- Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126
- Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G., and McMahon, K. D. (2016). Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus Accumulibacter phosphatis*. *ISME J.* 10, 810–822. doi: 10.1038/ismej.2015.155
- Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from

- isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. doi: 10.1038/s41564-017-0012-7
- Petriglieri, F., Petersen, J. F., Peces, M., Nierychlo, M., Hansen, K., Baastrand, C. E., et al. (2021). Quantification of biologically and chemically bound phosphorus in activated sludge from full-scale plants with biological P-removal. *BioRxiv* [Preprint] doi: 10.1101/2021.01.04.425262
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- R Core Team (2020). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rossetti, S., Tomei, M., Levantesi, C., Ramadori, R., and Tandoi, V. (2002). “*Microthrix parvicella*”: a new approach for kinetic and physiological characterization. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* 46, 65–72. doi: 10.2166/wst.2002.0458
- Rossetti, S., Tomei, M. C., Nielsen, P. H., and 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. doi: 10.1016/j.femsre.2004.09.005
- Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., Satagopam, V. P., Laczny, C. C., et al. (2015). Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *NPJ Biofilms Microbiomes* 1:15007. doi: 10.1038/npjbiofilms.2015.7
- RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.
- Schade, M., and Lemmer, H. (2005). Lipase activities in activated sludge and scum—comparison of new and conventional techniques. *Acta Hydrochim. Hydrobiol.* 33, 210–215. doi: 10.1002/ahch.200400576
- Schuster, K. C., Reese, I., Urlaub, E., Gapes, J. R., and Lendl, B. (2000). Multidimensional information on the chemical composition of single bacterial cells by confocal Raman microspectroscopy. *Anal. Chem.* 72, 5529–5534. doi: 10.1021/ac000718x
- Seviour, R., and Nielsen, P. H. (eds) (2010). *Microbial Ecology of Activated Sludge*. London: IWA Publishing.
- Sheik, A. R., Muller, E. E., Audinot, J.-N., Lebrun, L. A., Grysan, P., Guignard, C., et al. (2016). *In situ* phenotypic heterogeneity among single cells of the filamentous bacterium *Candidatus Microthrix parvicella*. *ISME J.* 10, 1274–1279. doi: 10.1038/ismej.2015.181
- Singleton, C. M., Petriglieri, F., Kristensen, J. M., Kirkegaard, R. H., Michaelsen, T. Y., Andersen, M. H., et al. (2021). Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat. Commun.* 12:2009. doi: 10.1038/s41467-021-22203-2
- Slijkhuis, H. (1983). *Microthrix parvicella*, a filamentous bacterium isolated from activated sludge: cultivation in a chemically defined medium. *Appl. Environ. Microbiol.* 46:8.
- Speirs, L. B. M., Rice, D. T. F., Petrovski, S., and Seviour, R. J. (2019). The phylogeny, biodiversity, and ecology of the Chloroflexi in activated sludge. *Front. Microbiol.* 10:2015. doi: 10.3389/fmicb.2019.02015
- Tandoi, V., Rossetti, S., Blackall, L. L., and Majone, M. (1998). Some physiological properties of an Italian isolate of “*Microthrix parvicella*.”. *Water Sci. Technol.* 37, 1–8. doi: 10.2166/wst.1998.0567
- Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., et al. (2020). MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* 48, D579–D589.
- van Veen, W. L. (1973). Bacteriology of activated sludge, in particular the filamentous bacteria. *Antonie Van Leeuwenhoek* 39, 189–205. doi: 10.1007/BF02578852
- Wallner, G., Amann, R., and Beisker, W. (1993). Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* 14, 136–143. doi: 10.1002/cyto.990140205
- Wang, J., Li, Q., Qi, R., Tandoi, V., and Yang, M. (2014a). Sludge bulking impact on relevant bacterial populations in a full-scale municipal wastewater treatment plant. *Process Biochem.* 49, 2258–2265. doi: 10.1016/j.procbio.2014.08.005
- Wang, J., Qi, R., Liu, M., Li, Q., Bao, H., Li, Y., et al. (2014b). The potential role of ‘*Candidatus Microthrix parvicella*’ in phosphorus removal during sludge bulking in two full-scale enhanced biological phosphorus removal plants. *Water Sci. Technol.* 70, 367–375. doi: 10.2166/wst.2014.216
- Wang, P., Yu, Z., Qi, R., and Zhang, H. (2016). Detailed comparison of bacterial communities during seasonal sludge bulking in a municipal wastewater treatment plant. *Water Res.* 105, 157–166. doi: 10.1016/j.watres.2016.08.050
- Wanner, J. (2017). “Activated sludge separation problems,” in *Activated Sludge Separation Problems*, eds S. Rossetti, V. Tandoi, and J. Wanner (London: International Water Association), 53–66. doi: 10.2166/9781780408644_053
- Wickham, H. (2009). *ggplot2—Elegant Graphics for Data Analysis*. Berlin: Springer Science + Business Media.
- Yilmaz, L. S., Parmerkar, S., and Noguera, D. R. (2011). mathFISH, a web tool that uses thermodynamics-based mathematical models for *in silico* evaluation of oligonucleotide probes for fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.* 77, 1118–1122. doi: 10.1128/AEM.01733-10

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Nierychlo, Singleton, Petriglieri, Thomsen, Petersen, Peces, Kondrotaitė, Dueholm and Nielsen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.