Genomic insights into Candidatus Amarolinea aalborgensis gen. nov., sp. nov., associated with settleability problems in wastewater treatment plants

Andersen, Martin Hjorth; McIlroy, Simon Jon; Nierychlo, Marta; Nielsen, Per Halkjær; Albertsen, Mads

Published in:
Systematic and Applied Microbiology

DOI (link to publication from Publisher):
10.1016/j.syapm.2018.08.001

Publication date:
2019

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

Systematic and Applied Microbiology 42 (2019) 77–84

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Genomic insights into *Candidatus* Amarolinea aalborgensis gen. nov., sp. nov., associated with settleability problems in wastewater treatment plants

Martin H. Andersen, Simon J. McIlroy [1], Marta Nierychlo, Per H. Nielsen, Mads Albertsen *

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

A B S T R A C T

Settleability of particles in activated sludge systems can be impaired by an overgrowth of filamentous bacteria, a problem known as bulking. These filaments are often members of the phylum Chloroflexi, sometimes reaching abundances in excess of 30% of the biovolume. The uncultured Chloroflexi phylotype, *Candidatus* Amarolinea, has been observed in high abundances in Danish full-scale activated sludge systems by 16S rRNA gene amplicon surveys, where it has been associated with bulking. In this study, fluorescence *in situ* hybridization was applied to confirm their high abundance, filamentous morphology, and contribution to the interfloc bridging that characterizes filamentous bulking. Furthermore, genome-centric metagenomics using both Illumina and Oxford Nanopore sequencing was used to obtain a near complete population genome (5.7 Mbp) of the *Ca.* Amarolinea phylotype, which belongs to the proposed novel family Amarolineaceae within the order Caldilineales of Chloroflexi. Annotation of the genome indicated that the phylotype is capable of aerobic respiration, fermentation, and dissimilatory nitrate reduction to ammonia. The genome sequence also gives a better insight into the phylogenetic and evolutionary relationships of the organism. The name *Candidatus* Amarolinea aalborgensis is proposed for the species.

Introduction

Wastewater treatment, using the conventional activated sludge process, relies on a complex consortium of bacteria to remove pollutants and facilitate separation of particles and water, prior to discharging the water into the environment [17]. The microbial communities are co-localized in activated sludge flocs, where filamentous bacteria provide structural support and facilitate separation of sludge and water through gravitational sedimentation. However, overgrowth of filamentous bacteria can lead to impaired settling (bulking) or foaming, which are serious operational problems and can reduce performance or increase the risk that biomass is washed out with the effluent [38].

To control excessive growth of filamentous bacteria, a significant effort has been made to identify and characterize the causative species [38]. Historically, filamentous bacteria have been "identified" by their morphology [16,18], but this is now recognized as imprecise or unreliable [38]. Through the use of molecular techniques, it has been revealed that filamentous bacteria associated with foaming and bulking belong to several phyla, although, most are related to the phylum Chloroflexi, and that almost all are uncultured [6,21,23,50,51]. Filamentous Chloroflexi can account for more than 30% of the total biovolume in some treatment plants based on fluorescence *in situ* hybridization (FISH) analyses [7,33,34] and 16S rRNA amplicon sequencing [39]. Their high abundance also suggests that members of the Chloroflexi phylum are important in nutrient transformation in the activated sludge community, where they are believed to be involved in the hydrolysis of particulate and colloidal organic matter. *In situ* studies using microautoradiography and FISH have shown that most utilize carbohydrates, but not short-chain fatty acids [22,23,39]. However, many genera are still not well characterized, and their ecology is generally poorly understood.

The Chloroflexi phylum is highly diverse in wastewater treatment systems. A 10-year survey of 25 Danish full-scale wastewater treatment plants (WWTPs) with nutrient removal showed that more than 50 abundant OTUs (97% cutoff, >1% relative abundance in at least one sample) related to Chloroflexi could be identified. The most abundant phylotypes were associated with the classes Anaerolineae, Ardenticatenia, SJA-15, and Caldilineae [39], where only one (*Candidatus* Defluviifilum [61]) has been isolated, and one (*Ca.* Promineofilum breve, [30]) has an annotated genome obtained from a metagenome. Several abundant and novel genera have been described based on their 16S rRNA gene sequences (*Ca.* Promineofilum, *Ca.* Defluviifilum, *Ca.* Villogracilis, *Ca.* Sarcinathrix, and *Ca.* Amarolinea), FISH probes have been designed for their study, and their basic *in situ* physiology described [23,30,39]. However, *in situ* studies alone are limited in the metabolic detail they can provide and are generally low throughput and labor intensive. Therefore, a

* Corresponding author.
E-mail address: ma@bio.aau.dk (M. Albertsen).
[1] Present address: Australian Centre for Ecogenomics, University of Queensland, Brisbane, QLD, Australia.

https://doi.org/10.1016/j.syapm.2018.08.001

prior knowledge and hypothesis are often required for a thorough and effective characterization of novel microorganisms. Furthermore, both metagenomic and gene expression studies of activated sludge are hampered by the lack of reference genomes for the abundant organisms [1].

Of the abundant Chloroflexi, species related to *Ca.* Amarolinea has often been observed in high (>30%) and dynamic abundance [31,39] and might be associated with bulking problems. Hence, the aim of this study was to apply genome-centric metagenomics to obtain a representative genome of *Ca.* Amarolinea, and thus the first genome of the class SJA-15 within the phylum Chloroflexi.

Materials and methods

Sampling and fixation

Activated sludge samples were collected from an aerated tank in Aalborg West full-scale, biological nutrient removal WWTP. Biomass was either stored at −80 °C for sequencing workflows or fixed with 4% paraformaldehyde (final concentration) for 3 h at 4 °C for FISH. Fixed samples were washed 3 times with sterile-filtered tap water and resuspended in 50% ethanol in 1 × PBS solution (v/v) and stored at −20 °C [37].

DNA extraction

For Illumina sequencing

DNA was extracted from activated sludge samples using the FastDNA™ SPIN Kit for Soil (MP Biomedicals, CA, USA) as described by the manufacturer, but with bead beating at 6 m s^{-1} using a Fast Prep FP120 (MP Biomedicals) for 4 × 40 s. The input material was 500 μL activated sludge.

For Nanopore sequencing

DNA was extracted from activated sludge using the DNeasy PowerSoil kit (QIAGEN) following the standard protocol, except for the exchange of vortexing for 10 times tube inversions and a sample input of 500 μL activated sludge.

Community profiling using 16S rRNA gene amplicon sequencing

Profiling of the bacterial community was performed as recommended by Albertsen et al. [2] using the 16S rRNA gene regions V1–V3 with the primers 27F (AGAGTTTGATCCTGGCTCAG) [24] and 534R (ATTACCGCGGCTGCTGG) [36]. Amplification was carried out using 2 mU Platinum Taq DNA Polymerase High Fidelity, 1 × Platinum High fidelity buffer, 400 pM dNTP, 1.5 mM $MgSO_4$, 5 μM Illumina barcoded V1–V3 adaptor mix, and 10 ng template DNA. The thermocycler program was 95 °C for 2 min, 30 cycles of 95 °C for 20 s, 56 °C for 30 s, 72 °C for 60 s, and a final elongation at 72 °C for 5 min. PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA) in a 5/4 ratio of sample/bead solution. Barcoded amplicons were paired-end sequenced on a MiSeq (Illumina, USA) with a MiSeq Reagent kit v3 (2 × 301 bp) and a 10% PhiX control library (Illumina, USA) spike-in. The reads were trimmed using trimmomatic v. 0.32 [8] and merged using FLASH v. 1.2.7 [28]. The reads were dereplicated, and potential PhiX contamination was identified using USEARCH v. 7.0.1090 [15], which was also used to cluster reads into operational taxonomic units (OTUs) with the –cluster_otus command using default settings. OTU abundances were estimated using USEARCH v. 7.0.1090 with the –usearch_global command and –id 0.97. Taxonomy was assigned using the RDP classifier [54], implemented in the parallel_assign_taxonomy_rdp.py script in QIIME [11], with the MiDAS database v. 1.23 [31], which is a curated database based on the SILVA database, release 123 [44]. The resulting data were analyzed in the R environment v. 3.5.0 [45] using the ampvis2 package [5].

Metagenome sequencing

Illumina workflow

Three samples were selected based on the community profile in order to enable differential coverage binning. DNA was fragmented using a M220 ultrasonicator (Covaris, UK) at 75 W peak incident power, 10% duty factor, 200 cycles per burst, and 40 s treatment time, as described by the manufacturer in order to yield a 550 bp product. Sequencing libraries were prepared using the NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs, UK), as described by the manufacturer. Three libraries were prepared with unique indexes using the NEBNext Multiplex Oligos for Illumina (New England BioLabs, UK). Sequencing libraries were validated with gel electrophoresis using Tapestation 2200 and High Sensitivity D1K screentapes (Agilent Technologies, CA, USA). The libraries were paired-end sequenced on a MiSeq (Illumina, USA) with a MiSeq Reagent kit v3 (2 × 301 bp) and a 10% PhiX control library (Illumina, USA) spike-in.

Nanopore workflow

Two libraries were prepared. One library was prepared according to the standard protocol for 1D Low input genomic DNA with PCR v. LIP_9021_v108_revL_11Nov2016 (Oxford Nanopore Technologies, UK) with 10 PCR cycles and a 15 min 50 s extension time for a 19 kbp target. DNA repair and fragmentation were omitted. The other library was prepared from two different samples with the standard protocol for 1D Genomic DNA by ligation v. GDE_9002_v108_revT_18Oct2016 (Oxford Nanopore Technologies, UK), where fragmentation was omitted, and DNA repair and end-prep were performed in a single step, using 48 μL sample (2.4 μg), 3.5 μL NEBNext FFPE DNA Repair Buffer (New England BioLabs, UK), 2 μL NEBNext FFPE DNA Repair Mix (New England BioLabs, UK), 3.5 μL NEBNext Ultra II End Prep Reaction Buffer (New England BioLabs, UK), and 3 μL NEBNext Ultra II End Prep Enzyme Mix (New England BioLabs, UK) with 15 min incubation at room temperature, and 5 min at 65 °C. Sequencing libraries were validated with gel electrophoresis using Tapestation 2200 and Genomic DNA screentapes (Agilent Technologies, CA, USA). The libraries were sequenced on a MinION 106 flow cell v. 9.4 (Oxford Nanopore Technologies, UK) according to protocol and with live base-calling in MinKNOW v. 1.11.5.

Assembly and binning

Cutadapt v. 1.16 [29] was applied to trim the Illumina FASTQ data for quality (phred ≥ 20), length (≥150 bp), and removal of NEBNext adapters. Nanopore FASTQ data were filtered for length (≥4000 bp) using Filtlong v. 0.1.1 [55], and adapters were removed with Porechop v. 0.2.3 [56] with the commands –min_split_read_size 4000 and –check_reads 1000. A *de novo* metagenome assembly was performed on one Illumina sample using metaSPAdes v. 3.11.1 with the kmer lengths 33, 55, 77, 99, and 127. Contigs with a coverage of <3 were removed along with contigs shorter than 1000 bp. The trimmed Illumina reads were mapped to the assembly using minimap2 v. 2–2.5 [25] with the commands '–ax sr' to specify short paired-end reads. Samtools v. 1.3.1 [26] was used to sort, merge, and estimate read depth of the resulting files. Coverage was calculated with the 'calc.coverage.in.bam.depth.pl' script. Automatic annotation was carried out using Prokka v. 1.12 with the –metagenome command. 16S rRNA gene sequences were found using BLASTn v. 2.2.28+ [10] with the options –max_target_seqs 5 –outfmt 6 –evalue 1e-10, extracted using the 'extract.long.hits.from.blast.pl'

script with the command –m 500, and classified using the SINA Alignment service v. 1.2.11 [43]. The mmgenome workflow script 'data.generation.2.1.0.sh' was used to generate the files necessary for the binning process. Binning was performed by differential coverage binning in the R environment v. 3.4.4 [45] using the package 'mmgenome2' v. 2.0 (github.com/KasperSkytte/mmgenome2) [19]. Reassembly of extracted bins was performed with the Illumina sample used for *de novo* assembly and the trimmed Nanopore data as inputs to Unicycler v. 0.4.3 [57], running SPAdes v.3.11.1 with the commands –no_correct –min_kmer_frac 0.3 –kmer_count 3 –no_pilon. Unicycler dependencies were racon [52], makeblastdb v. 2.2.28+, and tblastn v. 2.2.28+ [10]. The reassembled bin was further manually curated using mmgenome2 v. 2.0.

Annotation and statistics of reassembled bin

Statistics of the reassembled bin were obtained using QUAST v. 2.3 with default settings and CheckM lineage_wf v. 1.0.11 with default settings. Annotation of the reassembled bin was performed using Prokka v. 1.12 with the –metagenome command. Prokka annotations were compared to the manually curated annotations of *Ca.* Promineofilum breve [30] by BLASTp [3] search with the options –evalue 0.1 and max_target_seqs 1. Unidentified genes were compared to the NCBI Non-redundant protein sequences database by standard protein BLAST (blastp) with default settings.

Phylogeny

Taxonomic classification was obtained for the 16S rRNA gene using blastn with the 16S rRNA sequence database [3]. Furthermore, genome classification was obtained using the protein phylogeny in the Genome Taxonomy Database [40] with the gtdbtk classify_wf command. The resulting Newick tree was imported into ARB v. 6.0.6 [27] for visualization, and the section of interest exported for further formatting. Average nucleotide identities (ANIb) were computed in JSpecies v. 1.2.1 [46] using BLAST with default options. Average amino acid identities (AAI) were calculated using the AAI calculator in the enveomics toolbox [47].

FISH

The FISH protocol was performed as described by Daims et al. [13]. The CFX64 (5′-TCT ACC TAA GCA GAC CGT TC-3′) FISH probe was applied to target the genus *Ca.* Amarolinea using the optimal hybridization conditions [39]. EUBmix [4,12] was applied to target all bacteria, and NON-EUB [53] was applied as a negative hybridization control. Microscopic analysis was performed using a white light laser confocal microscope (Leica TCS SP8 X).

Data availability

The raw sequencing data described in this study is available from the European Nucleotide Archive (ENA) under the accession number PRJEB27012. The RMarkdown file describing the exact manual binning of the genome is available at github along with specific data generation scripts (https://github.com/martinhjorth/Publications/tree/Andersen2018).

Results and discussion

Selection of samples with high abundance of Ca. *Amarolinea*

Full-scale WWTPs were screened to identify samples with a high abundance of *Ca.* Amarolinea for use in genome-centric metagenomics. The relative abundances were evaluated from two large 16S rRNA gene amplicon surveys [31,39] and confirmed by FISH analyses for selected samples. The *Ca.* Amarolinea phylotype was observed in high abundance in several full-scale WWTPs with nutrient removal, while in most samples it was present in low abundance (663 samples with a mean relative read abundance of 0.32% and a maximum of 36.3%). In WWTPs with a high abundance of *Ca.* Amarolinea, bulking problems were often observed. In Aalborg West WWTP, a drastic increase in abundance of *Ca.* Amarolinea over a period of 1.5 years was correlated with strongly impaired settling properties. These results were confirmed by Nierychlo et al. (2018) through FISH analyses using probe CFX64 specific for *Ca.* Amarolinea, which were observed in up to of 15.8% of the total biovolume in Aalborg West WWTP [39]. In this study, the probe hybridized solely to filamentous bacteria with an Eikelboom 0092 morphotype, in agreement with previous *in situ* characterization of the *Ca.* Amarolinea [39], and was confirmed to contribute to interfloc bridging that is characteristic of filamentous bulking episodes (Fig. 1).

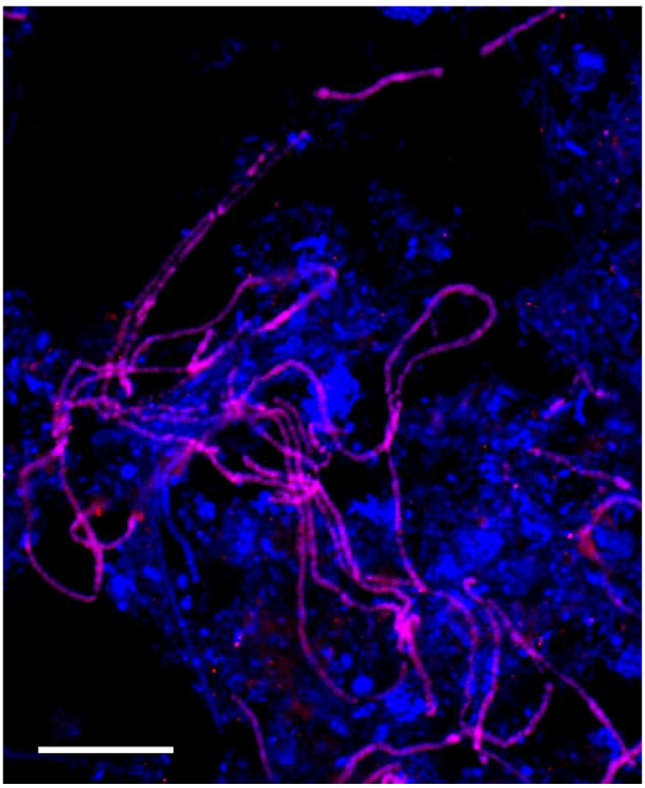

Fig. 1. Composite FISH Micrograph of the Aalborg WWTP biomass. Ca. Amarolinea filaments appear magenta (CFX64 (red) + EUBmix (blue)) and all other bacteria appear blue (EUBmix only). The scale bar represents 20 μm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Genome recovery of the Ca. *Amarolinea*

Five samples from Aalborg West WWTP with the target organism in high abundance were used for genome-centric metagenomics. All sampled were taken from the aeration tank in the period 2013–2015 (Fig. 2). Three samples were sequenced using short-read Illumina technology to enable differential coverage binning from several metagenomes, while two samples were sequenced using long-read technology through the Oxford Nanopore Technologies MinION platform to enable scaffolding.

An initial Chloroflexi bin, comprising 467 scaffolds, was extracted from the Illumina based metagenome assembly using differential coverage binning (Fig. 3). The extracted bin was

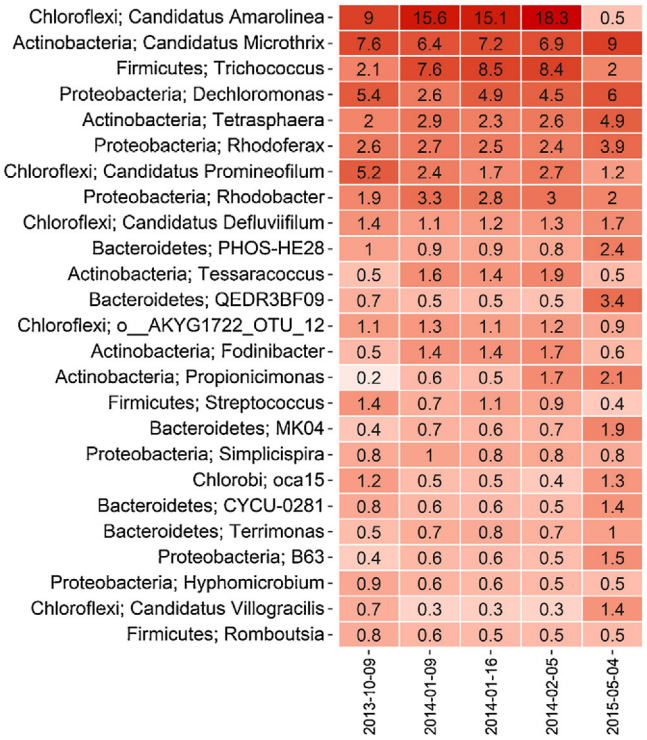

Fig. 2. Heatmap of the 25 most abundant bacterial genera and their phyla in the selected activated sludge samples from Aalborg Vest WWTP. The values indicate relative abundance in percentage, based on amplicon sequencing of the 16S rRNA gene regions V1–V3. The 2014-01-09 and 2014-01-16 samples were sequenced on Nanopore MinION, while the others were sequenced on Illumina MiSeq.

Table 1
Genome statistics for the assembled Chloroflexi genome. [a] indicates values estimated by the checkM software.

Collection site	Aalborg West WWTP, Denmark
Number of contigs	80
Total length (bp)	5,723,216
Longest contig (bp)	652,831
N50 (contigs, bp)	252,621
L50 (# of contigs)	7
GC (%)	62.0
Genome completeness (%)[a]	95.5
Genome contamination (%)[a]	1.1
Number of unique tRNA genes	46
rRNA operons found in genome	1
Number of CDS	4701
Coding density (%)	88.5

re-assembled through Unicycler [57] using both Illumina and Nanopore data, which resulted in a high-quality draft genome [9] of 5.7 Mbp in 80 contigs, with a GC content of 62.0% (Table 1). The completeness of the genome was estimated by CheckM [41] to be 95.5% with contamination of 1.1%. A single complete rRNA operon was present in the genome, and the 16S rRNA gene shared 99.8% similarity with the abundant OTU from the amplicon-surveys [31,39], confirming that the genome represented the abundant species in the full-scale plant.

Phylogeny

Based on its 16S rRNA gene, the *Ca.* Amarolinea sequences classify to the narrowly defined C10-SB1A class in the SILVA taxonomy, which is made up of only four sequences in the non-redundant (99% similarity cut-off) database base tree. The *Ca.* Amarolinea genus is defined in the MiDAS taxonomy (v. 2.1, as *Ca.* Amarilinum), which is a version of the SILVA taxonomy (v. 1.23) curated for activated sludge sequences [31], and the class C10-SB1A is relegated to an order within the class SJA-15. The closest named relative is *Caldilinea aerophila* DSM 14535 (87% identity, NR_074397.1) and *Caldilinea aerophila* STL-6-O1 (86% identity, NR_040878.1) of the Caldilineae class. *Caldilinea aerophila* STL-6-O1 is described as thermophilic and most likely not a phylotype present in activated sludge [49]. In accordance with the 16S rRNA gene phylogenetic tree by Nierychlo et al. [39], this places *Ca.* Amarolinea in a novel genus within a cluster of known activated sludge phylotypes (Supplementary Fig. S1), which is distantly related to the genus *Calidilinea*.

The genome of *Ca.* Amarolinea enables further investigation of the phylogenetic and evolutionary relationships using the recently available Genome Taxonomic Database (GTDB), which seeks to provide standardized bacterial taxonomy inferred from a concatenated alignment of 120 single copy marker genes, normalizing ranks based on relative sequence divergence [40]. In the GTDB taxonomy, the SILVA/MiDAS classes Caldilineae, Ardenticatenia, and Thermoflexia are relegated to the orders Caldilineales, Ardenticatenales, and Thermoflexales, respectively. In this taxonomic scheme, the *Ca.* Amarolinea sp. genome is assigned to a long branch clustered with genomes of the order Caldilineales, likely representing a novel family (Fig. 4). The MiDAS defined Anaerolineae, Caldilineae, Ardenticatenia, and SJA-15 represent the most abundant classes within the phylum in full-scale activated sludge WWTPs [39]. Thus, the Chloroflexi of these systems are dominated by members of the GTDB defined Anaerolineae. The ANIb and AAI values were compared for the phylotypes, which, according to Fig. 4, were closely related to *Ca.* Amarolinea aalborgensis. Comparing the six most closely related genomes (accession numbers GCF_000281175.1, GCA_002148365.1, GCA_001871755.1, GCA_002366755.1, GCA_002415895.1, and GCA_900066015.1) to *Ca.* Amarolinea aalborgensis by ANIb yielded values in the range 65.0–67.6%, indicating that ANIb comparison was not feasible due to the divergence of the genomes [48]. Instead, a two-way AAI comparison was used and resulted in a range of 47.0–50.9% AAI, with a MAG from ground water (accession no. GCA_001871755.1) being the most similar (based on alignments of 1646 proteins). These values are below the 60% AAI genus boundary suggested by Rodriguez-R and Konstantinidis [48], supporting the novelty of the draft genome.

Metabolism

Annotation of the *Ca.* Amarolinea genome indicated the potential for aerobic respiration based on the location of a putative cytochrome c oxidase (Supplementary Table S1), respiratory complexes, and genes for a complete TCA cycle (Fig. 5). The *Ca.* Amarolinea phylotype may perform dissimilatory nitrate reduction to ammonia (DNRA) as genes encoding both dissimilatory nitrate and nitrite reductases (nasB/C and nrfA respectively) were identified. A nitrous oxide reductase (nosZ) was annotated in the genome, but no nitrite reductase (nirK) or nitric oxide reductase were found, noting that both nitrite and nitric oxide may be reduced to ammonia through the action of the annotated nitrite reductase (nrfA) [42]. The same complement of genes involved in nitrogen metabolism were also located in the *Ca.* Promineofilum breve genome [30]. Genes coding for a glucose/mannose:H+ symporter and the complete glycolysis/gluconeogenesis pathway were found, although the gene coding for glucose-6-phosphate isomerase was not identified in the bin. Genes coding for a complete pathway for synthesis and degradation of glycogen were also identified. The *Ca.* Amarolinea genome contains the genes required for fermentation of carbohydrates, which confirms previous *in situ* studies for the genus and members of the phylum abundant in WWTPs [23,30,39]. The phenotype of

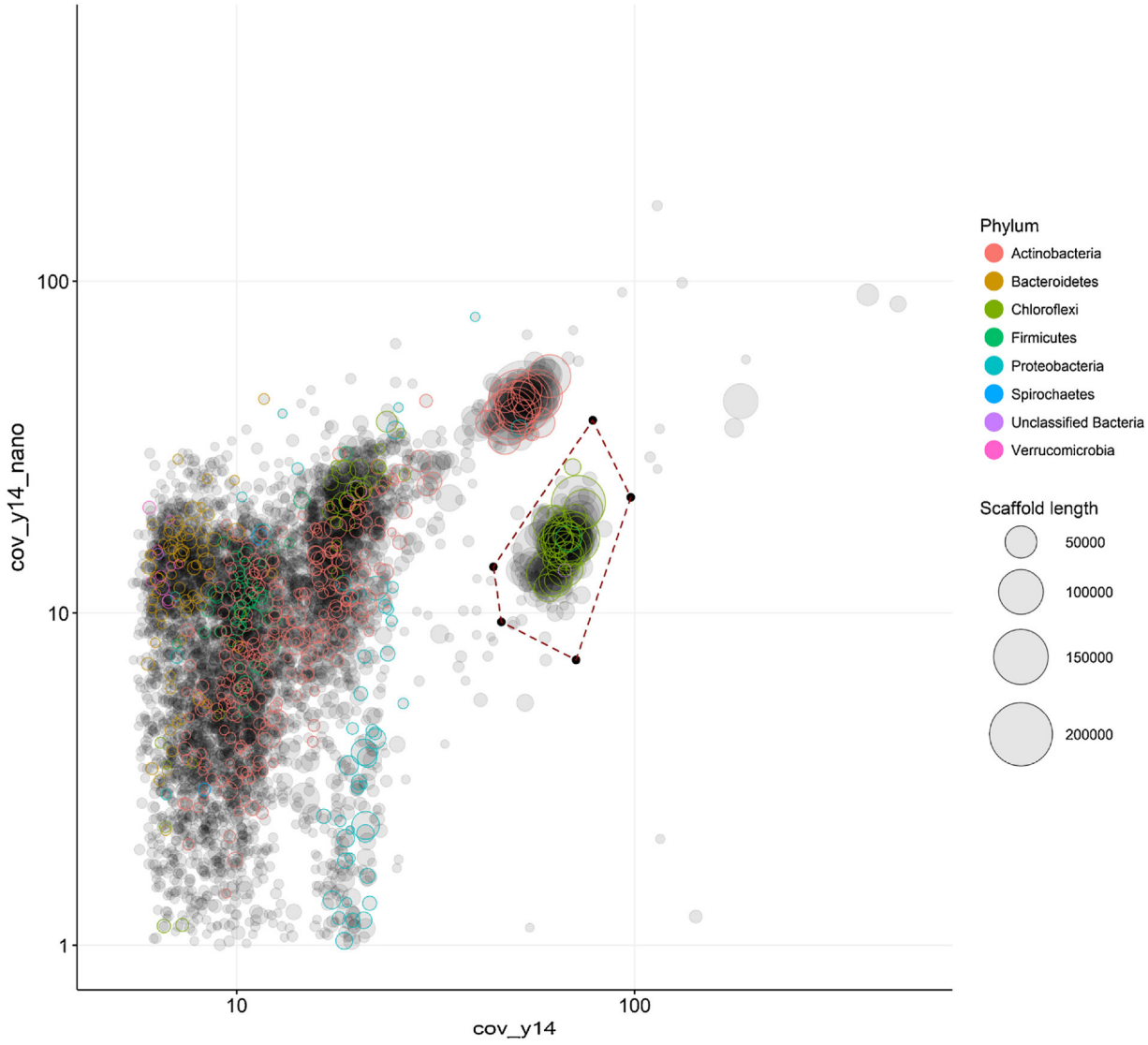

Fig. 3. Extraction of the bin of interest from the metagenome assembly. The read coverage of the Illumina data is plotted on the x-axis and the read coverage of the Nanopore MinION data is plotted on the y-axis. Both axes are log-scaled. Taxonomic classification at phylum level is used as color overlay. The polygon marks the extracted genome bin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the *Ca.* Amarolinea is also consistent with other members of the class Anaerolineae (GTDB defined), with all currently characterized members possessing a chemoheterotrophic metabolism, and most being able to grow anaerobically by the fermentation of carbohydrates and proteinaceous material [14,20,49,58,59]. The pathways described are visualized in a metabolic model (Fig. 5), with the curated gene annotations and the associated locus tags available in Supplementary Table S1. BLASTp results of the annotations are available in Supplementary Table S2.

In wastewater treatment plants configured for nutrient removal, where carbon is predominantly available under anaerobic conditions, *Ca.* Amarolinea and the other abundant Chloroflexi are likely to exhibit a fermentative lifestyle, utilizing carbohydrates. Under aerobic conditions, they may respire residual carbon or anaerobically stored glycogen (Supplementary Table S3). As such, their excessive overgrowth, as well as causing bulking problems, may be detrimental to the enhanced biological phosphate removal (EBPR) activated sludge systems through competition with the key polyphosphate-accumulating organisms (PAO). They would compete directly with the *Tetrasphaera* PAO for simple carbohydrates, and any carbon stored as glycogen anaerobically, and not fermented to volatile fatty acids, would not be available for the *Ca.* Accumulibacter PAO [32].

Taxonomic proposals for *Candidatus* Amarolinea aalborgensis

As no pure cultures are available, based on the recommendations by Murray and Stackebrandt [35], we propose the name *Candidatus* Amarolinea aalborgensis for the species. The formal proposal of the new species and genus *Candidatus* Amarolinea aalborgensis gen. nov., sp. nov. is given in Table 2 with the Taxonumber CA00038. Based on the phylogenetic position of the novel genus, we propose a new family in the Caldilineales order, the Amarolineaceae fam. nov., currently containing the type genus *Ca.* Amarolinea gen. nov. as the only genus. This genus is represented by the single type species *Ca.* Amarolinea aalborgensis gen. nov., sp. nov.

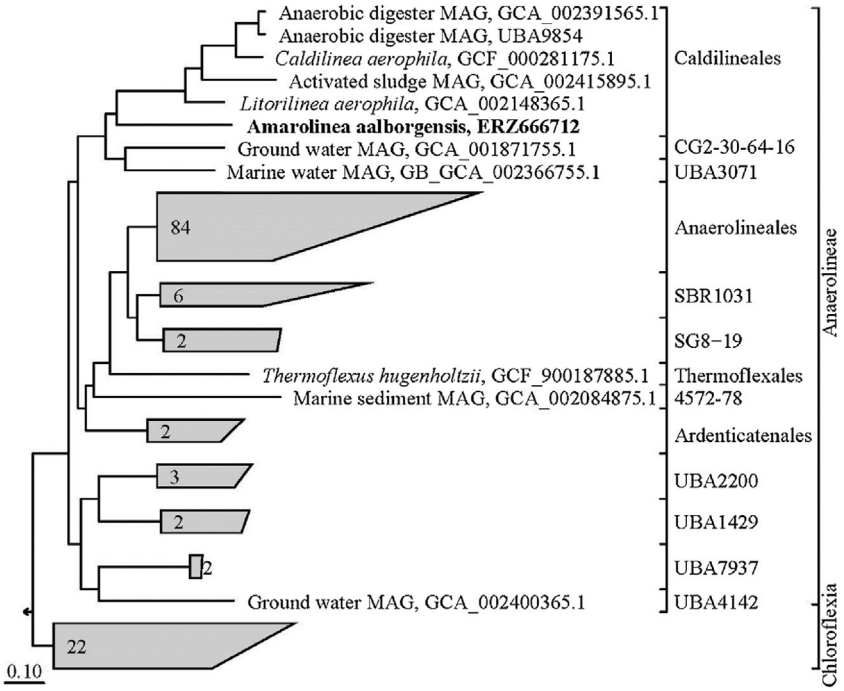

Fig. 4. Phylogenetic tree based on the Genome Tree Database, created with the gtdbtk workflow and visualized in ARB. The right, outermost brackets contain class-level classifications, while the inner brackets contain order-level classifications. 'MAG' stands for 'metagenome assembled genome'. The scale bar represents amino acid substitutions per site.

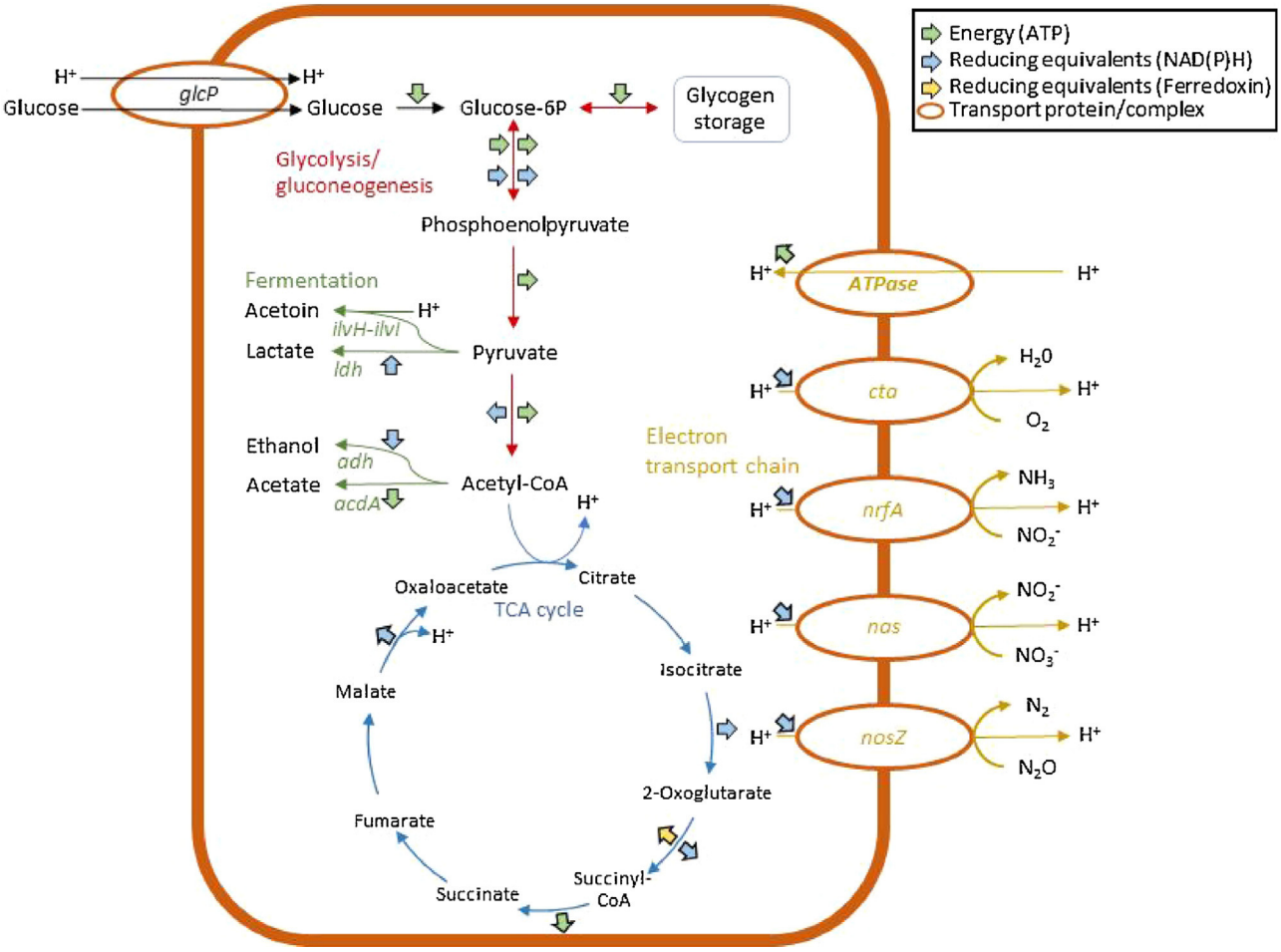

Fig. 5. Annotated metabolic pathways for the *Ca*. Amarolinea phylotype. Annotated pathways are color coded. Additional information is available in Supplementary Table S1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Digital Protologue table for *Ca.* Amarolinea aalborgensis gen. nov., sp. nov.

Taxonumber	CA00038
Species name	Amarolinea aalborgensis
Genus name	Amarolinea
Specific epithet	aalborgensis
Genus etymology	Amarolinea (A.ma.ro.li'ne.a. Gr. fem. n. amara conduit, channel, sewer; L. fem. n. linea a thread, a line; N.L. fem. n. Amarolinea a thread from a sewer).
Type species of the genus	Amarolinea aalborgensis
Genus status	gen. nov.
Species etymology	aalborgensis (aal.borg.en'sis. N.L. fem. adj. aalborgensis pertaining to the city of Aalborg, where the sample containing the organism was obtained)
Species status	sp. nov.
Authors	Andersen M. H., Nierychlo M., McIlroy S. J., Nielsen P. H., Albertsen M.
Title	Genomic insights into Candidatus Amarolinea aalborgensis gen. nov., sp. nov., associated with settleability problems in wastewater treatment plants
Journal	Systematic and Applied Microbiology
Corresponding author	Mads Albertsen
E-mail of the corresponding author	ma@bio.aau.dk
Submitter	Martin Hjorth Andersen
E-mail of the submitter	mha@bio.aau.dk
16S rRNA gene accession number	MH537630
MAG/SAG accession number [EMBL]	ERZ666712
Genome status	draft
Genome size	5723
GC mol %	62.0
Country of origin	Denmark
Region of origin	Aalborg, North Jutland
Source of sample	Activated sludge
Sampling date	2014-02-15
Geographic location	Aalborg West wastewater treatment plant
Latitude	57°04′80.3″N
Longitude	9°86′32.2″E
Relationship to O2	Facultative aerobe
Energy metabolism	Chemoorganotroph
Assembly	Replicate different samples
Sequencing technology	Illumina MiSeq and Oxford Nanopore MinION
Binning software used	mmgenome2 through the R environment
Assembly software used	Unicycler using SPAdes

Description of Amarolineaceae *fam. nov.*

Amarolineaceae (A.ma.ro.li'ne.a'ce.ae. Gr. fem. n. *Amarolinea* type genus of the family; *-aceae* ending to denote a family; N.L. fem. n. Amarolineaceae family of the genus *Amarolinea*).

The description is as for the genus *Amarolinea*. The type and only genus is *Amarolinea.*

Acknowledgements

This work was supported by the Innovation Fund Denmark (6155-00003A). S.J.M. was supported by a Danish Council for Independent Research grant (no. 4093-00127A). M.A. and P.H.N. was supported by research grants (15510 and 16578) from VILLUM FONDEN. The authors wish to thank Aharon Oren for help with naming and etymology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.08.001.

References

[1] Albertsen, M., Hansen, L.B.S., Saunders, A.M., Nielsen, P.H., Nielsen, K.L. (2012) A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. ISME J. 6 (6), 1094–1106, http://dx.doi.org/10.1038/ismej.2011.176.
[2] 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 (7), e0132783.
[3] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
[4] Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56 (6), 1919–1925, http://dx.doi.org/10.1111/j.1469-8137.2004.01066.x.
[5] Andersen, K.S.S., Kirkegaard, R.H., Karst, S.M., Albertsen, M. (2018) ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. BioRxiv, http://dx.doi.org/10.1101/299537.
[6] Beer, M., Seviour, E.M., Kong, Y., Cunningham, M., Blackall, L.L., Seviour, R.J. (2002) Phylogeny of the filamentous bacterium Eikelboom Type 1851, and design and application of a 16S rRNA targeted oligonucleotide probe for its fluorescence *in situ* identification in activated sludge. FEMS Microbiol. Lett. 207, 179–183.
[7] Beer, M., Stratton, H.M., Griffiths, P.C., Seviour, R.J. (2006) Which are the polyphosphate accumulating organisms in full-scale activated sludge enhanced biological phosphate removal systems in Australia? J. Appl. Microbiol. 100 (2), 233–243, http://dx.doi.org/10.1111/j.1365-2672.2005.02784.x.
[8] Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120, http://dx.doi.org/10.1093/bioinformatics/btu170.
[9] Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35 (8), 725–731, http://dx.doi.org/10.1038/nbt.3893.
[10] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009) BLAST+: architecture and applications. BMC Bioinf. 10, 1–9, http://dx.doi.org/10.1186/1471-2105-10-421.
[11] 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. Publ. Gr. 7 (5), 335–336, http://dx.doi.org/10.1038/nmeth0510-335.
[12] Daims, H., Brühl, A., Amann, R., Schleifer, K.H., 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 (3), 434–444, http://dx.doi.org/10.1016/S0723-2020(99)80053-8.
[13] Daims, H., Stoecker, K., Wagner, M. (2005) Fluorescence *in situ* hybridization for the detection of prokaryotes. In: Osborn, A.M., Smith, C.J. (Eds.), Molecular Microbial Ecology, Taylor & Francis, New York, NY, pp. 213–239.
[14] Dodsworth, J.A., Gevorkian, J., Despujos, F., Cole, J.K., Murugapiran, S.K., Ming, H., Li, W.J., Zhang, G., Dohnalkova, A., Hedlund, B.P. (2014) *Thermoflexus hugenholtzii* gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the *Chloroflexi*, *Thermoflexia* classis nov., and description of *Thermoflexaceae* fam. nov. and *Thermoflexales* ord. nov. Int. J. Syst. Evol. Microbiol. 64, 2119–2127, http://dx.doi.org/10.1099/ijs.0. 055855-0 (PART 6).
[15] Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26 (19), 2460–2461, http://dx.doi.org/10.1093/bioinformatics/btq461.
[16] Eikelboom, D.H. 2000 Process Control of Activated Sludge Plants by Microscopic Investigation, 1st ed., IWA Publishing, London.
[17] Grady, C.P.L., Jr., Daigger, G.T., Love, N.G., Filipe, C.D.M. 2011 Biological Wastewater Treatment, 3rd ed., CRC Press.
[18] Jenkins, D., Richard, M.G., Daigger, G.T., Jenkins, D. 2004 Manual on the Causes and Control of Activated Sludge Bulking, Foaming, and Other Solids Separation Problems, 3rd ed., Lewis Publishers.
[19] Karst, S.M., Kirkegaard, R.H., Albertsen, M. (2016) Mmgenome: a toolbox for reproducible genome extraction from metagenomes. BioRxiv, http://dx.doi.org/10.1101/059121.
[20] Kawaichi, S., Ito, N., Kamikawa, R., Sugawara, T., Yoshida, T., Sako, Y. (2013) *Ardenticatena maritima* gen. nov., sp. nov., a ferric iron- and nitrate-reducing bacterium of the phylum "*Chloroflexi*" isolated from an iron-rich coastal hydrothermal field, and description of *Ardenticatenia* classis nov. Int. J.

Syst. Evol. Microbiol. 63, 2992–3002, http://dx.doi.org/10.1099/ijs.0. 046532-0 (PART8).
[21] Kohno, T., Sei, K., Mori, K. (2002) Characterization of type 1851 organism isolated from activated sludge samples. Water Sci. Technol. 46 (1–2), 111–114.
[22] Kragelund, C., Levantesi, C., Borger, A., Thelen, K., Eikelboom, D., Tandoi, V., Kong, Y., Van Der Waarde, J., Krooneman, J., Rossetti, S., Thomsen, T.R., Nielsen, P.H. (2007) Identity, abundance and ecophysiology of filamentous *Chloroflexi* species present in activated sludge treatment plants. FEMS Microbiol. Ecol. 59 (3), 671–682, http://dx.doi.org/10.1111/j.1574-6941.2006.00251.x.
[23] Kragelund, C., Thomsen, T.R., Mielczarek, A.T., Nielsen, P.H. (2011) Eikelboom's morphotype 0803 in activated sludge belongs to the genus *Caldilinea* in the phylum *Chloroflexi*. FEMS Microbiol. Ecol. 76 (3), 451–462, http://dx.doi.org/10.1111/j.1574-6941.2011.01065.x.
[24] Lane, D.J. (1991) 16S/23S rRNA sequencing. In: Stackebrandt, E., Goodfellow, M. (Eds.), Nucleic Acid Techniques in Bacterial Systematics, John Wiley & Sons New York, NY, pp. 115–175.
[25] Li, H. (2017) Minimap2: pairwise alignment for nucleotide sequences. ArXiv, http://dx.doi.org/10.1101/169557.
[26] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25 (16), 2078–2079, http://dx.doi.org/10.1093/bioinformatics/btp352.
[27] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, A., Buchner, A., Lai, T., Steppi, S., Jacob, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H. (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32 (4), 1363–1371, http://dx.doi.org/10.1093/nar/gkh293.
[28] Magoč, T., Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27 (21), 2957–2963, http://dx.doi.org/10.1093/bioinformatics/btr507.
[29] Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 17 (1), http://dx.doi.org/10.14806/ej.17.1.200.
[30] McIlroy, S.J., Karst, S.M., Nierychlo, M., Dueholm, M.S., Albertsen, M., Kirkegaard, R.H., Seviour, R.J., Nielsen, P.H. (2016) Genomic and *in situ* investigations of the novel uncultured *Chloroflexi* associated with 0092 morphotype filamentous bulking in activated sludge. ISME J. 10, 2223–2234, http://dx.doi.org/10.1038/ismej.2016.14.
[31] McIlroy, S.J., Kirkegaard, R.H., McIlroy, B., Nierychlo, M., Kristensen, J.M., Karst, S.M., Albertsen, M., Nielsen, P.H. (2017) Database update MiDAS 2. 0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database, 1–9, 10.1093/database/bax016.
[32] McIlroy, S.J., Onetto, C.A., McIlroy, B., Herbst, F.-A., Dueholm, M.S., Kirkegaard, R.H., Fernando, E., Karst, S.M., Nierychlo, M., Kristensen, J.M., Eales, K.L., Grbin, P.R., Wimmer, R., Nielsen, P.H. (2018) Genomic and *in situ* analyses reveal the *Micropruina* spp. as abundant fermentative glycogen accumulating organisms in enhanced biological phosphorus removal systems. Front. Microbiol 9 (May), 1–12, http://dx.doi.org/10.3389/fmicb.2018.01004.
[33] Mielczarek, A.T., Kragelund, C., Eriksen, P.S., Nielsen, P.H. (2012) Population dynamics of filamentous bacteria in Danish wastewater treatment plants with nutrient removal. Water Res. 46 (12), 3781–3795, http://dx.doi.org/10.1016/j.watres.2012.04.009.
[34] Morgan-Sagastume, F., Larsen, P., Nielsen, J.L., Nielsen, P.H. (2008) Characterization of the loosely attached fraction of activated sludge bacteria. Water Res. 42 (4-5), 843–845, http://dx.doi.org/10.1016/j.watres.2007.08.026.
[35] Murray, R.G.E., Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. Int. J. Syst. Bacteriol. 45 (1), 186–187, http://dx.doi.org/10.1099/00207713-45-1-186.
[36] 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 (3), 695–700.
[37] Nielsen, J.L. (2009) Protocol for fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotides. In: Nielsen, P.H., Daims, H., Lemmer, H. (Eds.), FISH Handbook for Biological Wastewater Treatment, 1st ed., IWA Publishing, London, pp. 73–84.
[38] Nielsen, P.H., Kragelund, C., Seviour, R.J., Nielsen, J.L. (2009) Identity and ecophysiology of filamentous bacteria in activated sludge. FEMS Microbiol. Rev. 33 (6), 969–998, http://dx.doi.org/10.1111/j.1574-6976.2009.00186.x.
[39] Nierychlo, M., Miłobędzka, A., Petriglieri, F., McIlroy, B., Nielsen, P.H., McIlroy, S.J. (2018) The ecology of the *Chloroflexi* in full-scale activated sludge wastewater treatment plants. BioRxiv, http://dx.doi.org/10.1101/335752.
[40] Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P. (2018) A proposal for a standardized bacterial taxonomy based on genome phylogeny. BioRxiv, http://dx.doi.org/10.1101/256800.
[41] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25 (7), 1043–1055, http://dx.doi.org/10.1101/gr.186072.114.
[42] Poock, S.R., Leach, E.R., Moir, J.W.B., Cole, J.A., Richardson, D.J. (2002) Respiratory detoxification of nitric oxide by the cytochrome *c* nitrite reductase of *Escherichia coli*. J. Biol. Chem. 277 (26), 23664–23669, http://dx.doi.org/10.1074/jbc.M200731200.
[43] Pruesse, E., Peplies, J., Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28 (14), 1823–1829, http://dx.doi.org/10.1093/bioinformatics/bts252.
[44] 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(D1, 590–596, http://dx.doi.org/10.1093/nar/gks1219.
[45] R Core Team. (2015) R: a language and environment for statistical computing.
[46] Richter, M., Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. 106 (45), 19126–19131, http://dx.doi.org/10.1073/pnas.0906412106.
[47] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J. Prepr. 4, e1900v1, http://dx.doi.org/10.7287/peerj.preprints.1900v1.
[48] Rodriguez-R, L.M., Konstantinidis, K.T. (2014) Bypassing cultivation to identify bacterial species. Microbe 9 (3), 111–118.
[49] Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H., Kamagata, Y. (2003) *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain *Bacteria* at the subphylum level. Int. J. Syst. Evol Microbiol. 53 (6), 1843–1851, http://dx.doi.org/10.1099/ijs.0. 02699-0.
[50] Speirs, L., Nittami, T., McIlroy, S., Schroeder, S., Seviour, R.J. (2009) Filamentous bacterium Eikelboom Type 0092 in activated sludge plants in Australia is a member of the phylum *Chloroflexi*. Appl. Environ. Microbiol. 75 (8), 2446–2452, http://dx.doi.org/10.1128/AEM.02310-08.
[51] Speirs, L.B.M., Dyson, Z.A., Tucci, J., Seviour, R.J. (2017) Eikelboom filamentous morphotypes 0675 and 0041 embrace members of the *Chloroflexi*: resolving their phylogeny, and design of fluorescence *in situ* hybridisation probes for their identification. FEMS Microbiol. Ecol. 93 (10), 1–13, http://dx.doi.org/10.1093/femsec/fix115.
[52] Vaser, R., Sovic, I., Nagarajan, N., Sikic, M. (2016) Fast and accurate de novo genome assembly from long uncorrected reads. BioRxiv, http://dx.doi.org/10.1101/068122.
[53] Wallner, G., Amann, R., Beisker, W. (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry 14 (2), 136–143, http://dx.doi.org/10.1002/cyto.990140205.
[54] Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73 (16), 5261–5267, http://dx.doi.org/10.1128/AEM.00062-07.
[55] Wick, R.R. Filtlong. https://github.com/rrwick/Filtlong. [Accessed 16 April 2018].
[56] Wick, R.R. Porechop. https://github.com/rrwick/Porechop. [Accessed 16 April 2018].
[57] Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E. (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13 (6), 1–22, http://dx.doi.org/10.1371/journal.pcbi.1005595.
[58] Yamada, T., Imachi, H., Ohashi, A., Harada, H., Hanada, S., Kamagata, Y., Sekiguchi, Y. (2007) *Bellilinea caldifistulae* gen. nov., sp. nov and *Longilinea arvoryzae* gen. nov., sp. nov., strictly anaerobic, filamentous bacteria of the phylum *Chloroflexi* isolated from methanogenic propionate-degrading consortia. Int. J. Syst. Evol. Microbiol. 57 (10), 2299–2306, http://dx.doi.org/10.1099/ijs.0. 65098-0.
[59] Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H., Kamagata, Y. (2006) *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae. Int. J. Syst. Evol. Microbiol. 56 (6), 1331–1340, http://dx.doi.org/10.1099/ijs.0. 64169-0.
[61] Yoon, D.N., Park, S.J., Kim, S.J., Jeon, C.O., Chae, J.C., Rhee, S.K. (2010) Isolation, characterization, and abundance of filamentous members of *Caldilineae* in activated sludge. J. Microbiol. 48 (3), 275–283, http://dx.doi.org/10.1007/s12275-010-9366-8.