MiDAS 5: Global diversity of bacteria and archaea in anaerobic digesters

Authors: Morten Kam Dahl Dueholm[1,*], Kasper Skytte Andersen[1], Anne-Kirstine C. Petersen[1], Vibeke Rudkjøbing[1], MiDAS Global Consortium for Anaerobic Digesters[**] & Per Halkjær Nielsen[1,*]

Affiliations:

[1]Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark. [**]A list of authors and their affiliations appears at the end of the paper.

[*]Correspondence to: Per Halkjær Nielsen, Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark; Phone: +45 2173 5089; Fax: Not available; E-mail: phn@bio.aau.dk or Morten Kam Dahl Dueholm, Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark; Phone: (+45) 9940 8508; Fax: Not available; E-mail: md@bio.aau.dk

Running title: Global microbiota of anaerobic digesters

Abstract

Anaerobic digestion represents a key biotechnology for the transformation of organic waste into renewable energy (biogas) and relies on complex microbial communities that work in concert to degrade the complex substrates into methane and carbon dioxide. Here, we sequenced more than half a million high-quality, full-length 16S rRNA gene sequences from 285 full-scale anaerobic digesters (ADs) across the world to expand our knowledge about diversity and function of the bacteria and archaea in ADs. The sequences were processed into full-length 16S rRNA amplicon sequence variants (FL-ASVs), which were added to the MiDAS 4 database for bacteria and archaea in wastewater treatment systems to create 'MiDAS 5'. The expansion of the MiDAS database significantly increased the coverage for bacteria and archaea in ADs worldwide, leading to an improved rate of genus and species-level classification. Using MiDAS 5, we carried out an amplicon-based, global-scale microbial community profiling of the sampled ADs using three common sets of primers targeting different regions of the 16S rRNA gene in bacteria and/or archaea. We revealed how environmental conditions and biogeography shape the AD microbiota. We also identify core and conditionally rare or abundant taxa, encompassing less than 1000 genera and species. These represent 83-99% and 10-64% of the accumulated read abundance respectively, across samples depending on the amplicon primers used. Finally, we examined the global diversity of functional groups with known importance for the anaerobic digestion process. Our online global MiDAS Field Guide presents the data generated in this study and summarizes present knowledge about all taxa.

Introduction:

Anaerobic digestion has gained attention as an important, sustainable biotechnology as it provides a number of benefits that align with the goals of sustainability. It can help to produce renewable energy (biogas) from organic waste such as manure, food waste, and sludge from wastewater treatment plants (WWTPs) [1,2]. The anaerobic digestion process also reduces pathogens and the amount of organic waste that is sent to landfills, thereby reducing methane emissions and supporting sustainable waste management practices [1]. Finally, the fertilizer that is produced as a byproduct of anaerobic digestion can be used to support sustainable agriculture, reducing the need for synthetic fertilizers that can have negative environmental impacts [3,4].

The anaerobic digestion process relies on the microbial degradation and conversion of organic matter, which requires a complex interplay between several functional guilds. These include hydrolyzing, acidogenic, and acetogenic syntrophic bacteria as well as methanogenic archaea [5]. The taxonomy is poorly characterized for many of the microorganisms in anaerobic digesters (ADs), and even among the most abundant taxa many lack genus- or species-level classifications [6]. To optimize performance, a comprehensive knowledge about microbial immigration/competition, environmental/operational conditions, and taxonomy is essential [7–9]. Recent microbial surveys have increased our knowledge about the anaerobic digestion process [7,10–14]. However, sharing knowledge across studies is still hindered by the absence of standardized protocols and a common reference database with a unifying taxonomy [15,16]. To facilitate collaboration and knowledge sharing, it is essential to establish these standard protocols and resources.

The Microbial Database for Activated Sludge and Anaerobic Digesters (MiDAS) project was established as an open-source platform for sharing updated knowledge about the physiology and ecology of the important microorganisms present in engineered ecosystems of activated sludge plants, ADs, and related WWTPs [15–18]. MiDAS provides standardized protocols for microbial profiling of microbes in wastewater treatment systems [19], an ecosystem-specific full-length 16S rRNA gene reference database [16,18], and a field guide where knowledge about the specific genera are stored and shared (https://www.midasfieldguide.org).

The MiDAS 16S rRNA gene reference database was created based on millions of high-quality, chimera-free, full-length 16S rRNA genes resolved into amplicon sequence variants (ASVs) and classified using automated taxonomy assignment (AutoTax) [6,16,18]. AutoTax provides a comprehensive seven-rank taxonomy (kingdom to species-level) to all reference sequences based on the most recent version of the SILVA SSURef 99 NR taxonomy, and adds a robust placeholder taxonomy for lineages without an official taxonomy [6]. The placeholder taxa can easily be distinguished by their names (midas_x_y, where x indicates the taxonomic rank and y is a numerical identifier) and

provides an important opportunity to study unclassified taxa along classified taxa at different taxonomic ranks. The placeholder taxonomy should not be seen as a replacement for proper taxonomic classifications but can pinpoint important lineages that should be studied in depth using phylogenomics [20–24].

The MiDAS 16S rRNA gene reference database (MiDAS 4.8.1) currently contains reference sequences from anaerobic digesters located at WWTPs in Denmark and WWTPs worldwide [18]. However, it may not provide comprehensive coverage for all important microbes found in ADs treating other types of waste or in other locations.

In this study, we introduce MiDAS 5, an updated version of MiDAS 4 expanded with more than half a million high-quality, full-length archaeal and bacterial 16S rRNA gene sequences from 285 ADs worldwide treating different types of biowaste. We carried out a global survey of ADs using three commonly used short-read amplicon primer sets targeting bacteria (V1-V3), archaea (V3-V5), and both (V4). This data was used in combination with MiDAS 5 to i) link the global diversity of bacteria and archaea to biogeography and environmental factors, ii) identify important core taxa, and iii) uncover the global diversity within selected functional guilds. The results provide a solid foundation for future research on AD microbiology.

Results and Discussion:

The MiDAS Global Consortium for Anaerobic Digesters was established in 2018 to coordinate the sampling and collection of metadata from ADs worldwide (Supplementary Data 1). Samples were obtained in duplicates from 285 ADs in 188 cities in 19 countries on five continents (Fig. 1a). The majority of the ADs treated surplus sludge from wastewater treatment plants (60.0%) (Fig. 1b). However, ADs treating food waste (6.7%), industrial waste (7.4%), and manure (4.9%) were also included in the survey. Most of the ADs were mesophilic (85.6%), few were thermophilic (6.0%), and some missed temperature data (8.4%). The main digester technology used was completely mixed reactors (67.7%) followed by two-stage reactors (12.6%). A few upflow anaerobic sludge blanket (UASB) and plug-flow reactors were also sampled to expand the diversity of digester types.

Expanding the MiDAS database with reference from global ADs

To expand the MiDAS database with sequences from ADs across the globe, we applied high-fidelity, full-length 16S rRNA gene sequencing on all samples collected in this study [25]. More than half a million full-length 16S rRNA gene sequences, representing both bacteria and archaea, were obtained after quality filtering and primer trimming. The sequences were processed with AutoTax to produce full-length 16S rRNA gene ASVs (FL-ASVs), and these were added to the MiDAS 4.8.1 database leading to an expanded database (MiDAS 5) with a total of 120,408 unique, chimera-free FL-ASV reference sequences.

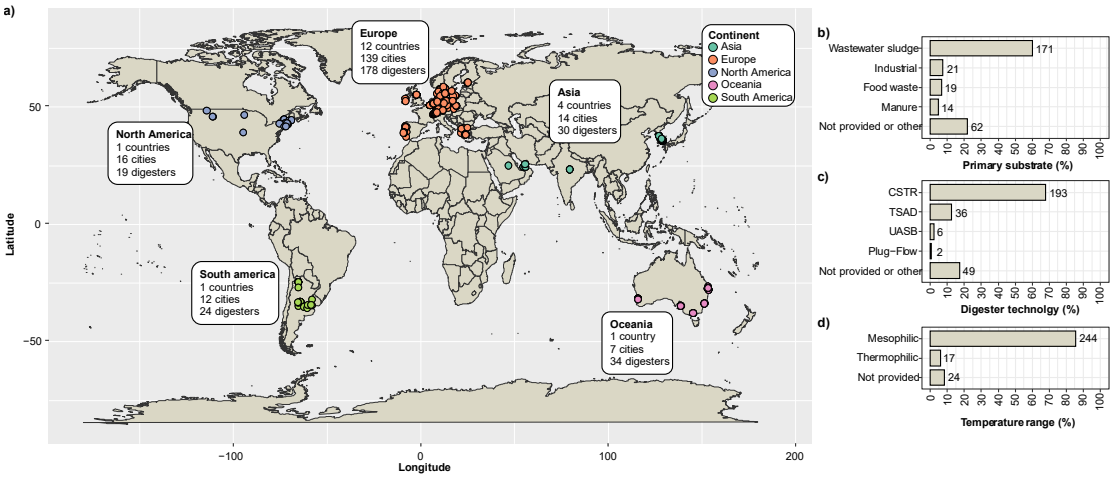

Fig. 1: Sampling of anaerobic digesters (ADs) across the world. a) Geographical distribution of ADs included. b) Distribution of primary substrates. c) Distribution of digester technologies. CSTR: Continuous stirred-tank reactor; TSAD: Two-stage anaerobic digestion; UASB: Upflow anaerobic sludge blanket: d) Distribution of digester temperatures. The values next to the bars are the number of ADs in each group.

Table 1: Sequence novelty of FL-ASVs obtained in this study. Sequence novelty was determined based on the percentage identity between each of the 30,246 new FL-ASV and their closest relative in the databases indicated and identity thresholds for each taxonomic rank proposed by Yarza *et al.* [26] shown in the parentheses.

	SILVA 138.1 SSURef NR99		MiDAS 4.8.1	
	Sequences	Percentage	Sequences	Percentage
Phylum (<75.0%)	28	0.09%	67	0.22%
Class (<78.5%)	55	0.18%	186	0.61%
Order (<82.0%)	112	0.37%	462	1.53%
Family (<86.5%)	354	1.17%	1,483	4.90%
Genus (<94.5%)	5,240	17.32%	9,419	31.14%
Species (<98.7%)	15,806	52.26%	16,863	55.75%

The novelty of the 30,246 new FL-ASVs were determined based on the percent identity shared with their closest relatives in the SILVA 138.1 SSURef NR99 and MiDAS 4.8.1 database using the threshold for each taxonomic rank proposed by Yarza *et al.* [26] (Table 1). 17% and 31% of the new FL-ASV lacked genus-level homologs (≥94.5% identity) and 52% and 56% were without species-level homologs (≥98.7% identity) in SILVA 138.1 and MiDAS 4, respectively. This suggests a substantial increase in the diversity within the MiDAS database.

147 ***MiDAS 5 introduces many new taxa***

148 To investigate how the new FL-ASVs affected the taxonomic diversity in the MiDAS
149 database, we determined the number of additional taxa introduced at different
150 taxonomic ranks (Table 2). A substantial increase in diversity was observed with the
151 addition of 2,770 new genera (29.2% increase) and 8,858 new species (28.3% increase).
152 However, many additional taxa were also introduced at higher taxonomic ranks as
153 highlighted with the introduction of six more bacterial and five more archaeal phyla
154 previously known from the SILVA taxonomy. In addition, we observed nine deep
155 branching lineages classified as novo phyla, however phylogenetic analysis revealed
156 that these likely represent mitochondrial sequences. The largest percentage of the new
157 FL-ASVs (42.8%) were found within the Firmicutes (Supplementary Fig. 1a).
158 Firmicutes often occur in high abundance in ADs, where they are involved in
159 fermentation and thereby directly stimulate biogas yields [7,10,13,25]. A closer look into the
160 expanded diversity within the Firmicutes revealed that new FL-ASVs were associated
161 with several families (Supplementary Fig. 1b), including Hungateiclostridiaceae (1,324
162 FL-ASVs), Lachnospiraceae (788 FL-ASVs), Peptostreptococcales-Tissierellales
163 Family_XI (763 FL-ASVs), Christensenellaceae (754 FL-ASVs),
164 Caldicoprobacteraceae (620 FL-ASVs), and Syntrophomonadaceae (555 FL-ASVs).
165 The Syntrophomonadaceae are of special relevance, as this family includes several
166 syntrophic fatty acid degrading bacteria, which are often the metabolic bottleneck in
167 the overall ADs process [27,28].

168 Table 2: New taxa introduced with MiDAS 5. The number of new taxa represent unique taxa at the
169 different taxonomic ranks that were not part of MiDAS 4.8.1 [18] and includes both official taxonomic
170 names and de novo placeholder names provided by AutoTax [6].

	Total taxa	New taxa	Increase (%)
Phylum	105	20	23.5%
Class	259	37	16.7%
Order	727	104	16.7%
Family	2,212	360	19.4%
Genus	12,254	2,770	29.2%
Species	40,207	8,858	28.3%

171

172 ***MiDAS 5 provides improved coverage and classifications for AD microbiota***
173 The performance of the MiDAS 5 database was evaluated based on three ASV-
174 resolved, short-read, 16S rRNA gene amplicon datasets generated from the AD samples
175 collected in this study (Fig. 2). The V1-V3 amplicons include only bacteria and provide
176 high phylogenetic resolution. However, the primers targeting this region have a lower
177 coverage for the known bacterial diversity according to in silico evaluations [6,29]. The
178 V4 amplicons include both bacteria and some archaeal lineages and are commonly used
179 due to a very good coverage of the known bacterial diversity. However, the amplicons
180 have a weaker phylogenetic resolution compared to V1-V3, which in many cases
181 prevent species-level classifications [6,29]. The V3-V5 amplicons cover mainly archaea
182 and have previously been used to describe their diversity in ADs [7,10].
183
184 Our first evaluation was based on stringent mapping of the short-read ASVs to MiDAS
185 5 and other commonly applied reference databases. This allowed us to determine the
186 percent identity shared between each ASV and their best hit in each reference database,
187 which was subsequently used to calculate the percentage of ASVs with high-identity
188 hits (≥99% identity) per sample and database after removing the rare biosphere (ASV
189 with <0.01% relative abundance), known to be enriched in non-growing organisms and
190 environmental DNA [10]. MiDAS 5 performed exceptionally well for bacteria with high-
191 identity hits of 94.5%±4.6% (mean±SD) for V1-V3 and 96.4%±2.8% for V4 ASVs,
192 compared to 69.9%±19.7% and 73.6%±16.1% for MiDAS 4, and 61.9%±9.2% and
193 78.8%±7.0% for SILVA v.138.1 (Fig. 2). For the V3-V5 archaeal dataset, an increase
194 in coverage was observed from 34.1%±8.7% with MiDAS 4 to 56.0%±11.8% with
195 MiDAS 5. However, the SILVA database provides even better coverage
196 (66.4%±12.0%). The 16S rRNA gene reference database derived from the genome
197 taxonomy database (GTDB) performed poorly, which relates to the fact that many
198 metagenome-assembled genomes (MAGs) created from short-read data lack 16S rRNA
199 genes due to assembly issues [30,31]. The lower coverage seen for archaea is likely related
200 to reduced sequencing efforts and difficulties designing universal primers for archaeal
201 full-length 16S rRNA gene sequencing [32,33]. Because the sampling of ADs was directed
202 towards mesophilic digesters treating surplus sludge from WWTPs, we also evaluated
203 the MiDAS 5 coverage for ADs treating different primary substrates and temperatures
204 (Supplementary Fig. 2). MiDAS 5 gave very good coverage for all sample types
205 supporting the general applicability of the reference database for ADs.

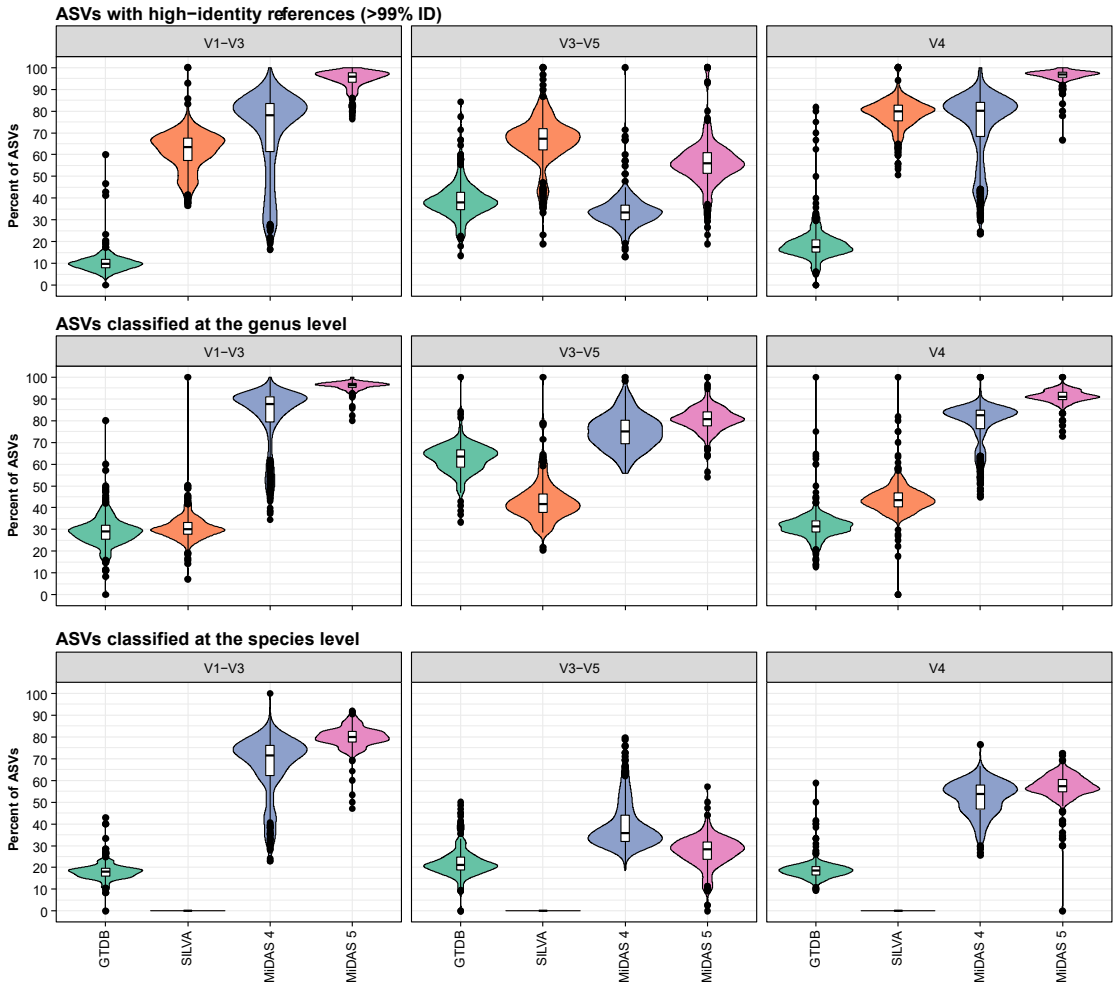

Fig. 2: Database evaluation based on short-amplicon data from this study. The ASVs for each of the samples were filtered based on their relative abundance (only ASVs with ≥0.01% relative abundance were kept) before the analyses. The percentage of the microbial community represented by the remaining ASVs after the filtering was 95.43% ± 2.38% (mean±SD) for V1-V3 amplicons (only bacteria), 99.69% ± 0.13% for V3-V5 amplicons (mainly archaea), and 97.68% ± 1.97% for V4 amplicons (bacteria and archaea) across samples. High-identity (≥99%) hits were determined by stringent mapping of ASVs to each reference database. Classification of ASVs was done using the SINTAX classifier. The violin and box plot represent the distribution of percent of ASVs with high-identity hits or genus/species-level classifications for each database across n=469 (V1-V3), n=273 (V3-V5), and n=470 (V4) duplicate, biologically independent samples duplicates. Box plots indicate median (middle line), 25th, 75th percentile (box), and the min and max values after removing outliers based on 1.5x interquartile range (whiskers). Outliers have been removed from the box plots to ease visualization. Different colors are used to distinguish the different databases: GTDB SSU_r202, SILVA 138.1 SSURef NR99, MiDAS 4.8.1, and MiDAS 5.0.

Our second database evaluation was based on the classification of ASVs from each amplicon dataset using the SINTAX classifier (Fig. 2). We found that MiDAS 5 greatly improved the rates of genus-level classification (96.1%±1.9% for V1-V3, 91.2%±2.9% for V4, and 80.6%±6.0% for V3-V5) compared to MiDAS 4 (81.7%±14.2 for V1-V3, 78.6%±9.7% for V4, and 75.0%±8.1% for V3-V5), and that these rates of classification were much higher than those obtained with SILVA (30.7%±6.0 for V1-V3, 43.8%±7.8% for V4, and 42.7%±9.2% for V3-V5) and GTDB (29.0%±6.9 for V1-V3, 31.7%±6.9% for V4, and 62.7%±7.2% for V3-V5). Analysis of species-level classifications revealed similar improvements with MiDAS 5 for bacteria (Fig. 2). However, a decrease in species-level classifications was observed between MiDAS 4 and 5 for the archaeal V3-V5 dataset. We hypothesize that this effect relates to over-classifications with MiDAS 4 due to the lack of appropriate reference sequences in MiDAS 4.

Finally, we investigated if the additional reference sequences introduced in MiDAS 5 could improve classification of amplicon data from WWTPs based on data from the MiDAS global sampling of WWTPs [18] and the Global Water Microbiome Consortium project [34] (Supplementary Fig. 3). The improvements were insignificant, highlighting that most of the added references originated from AD-specific taxa.

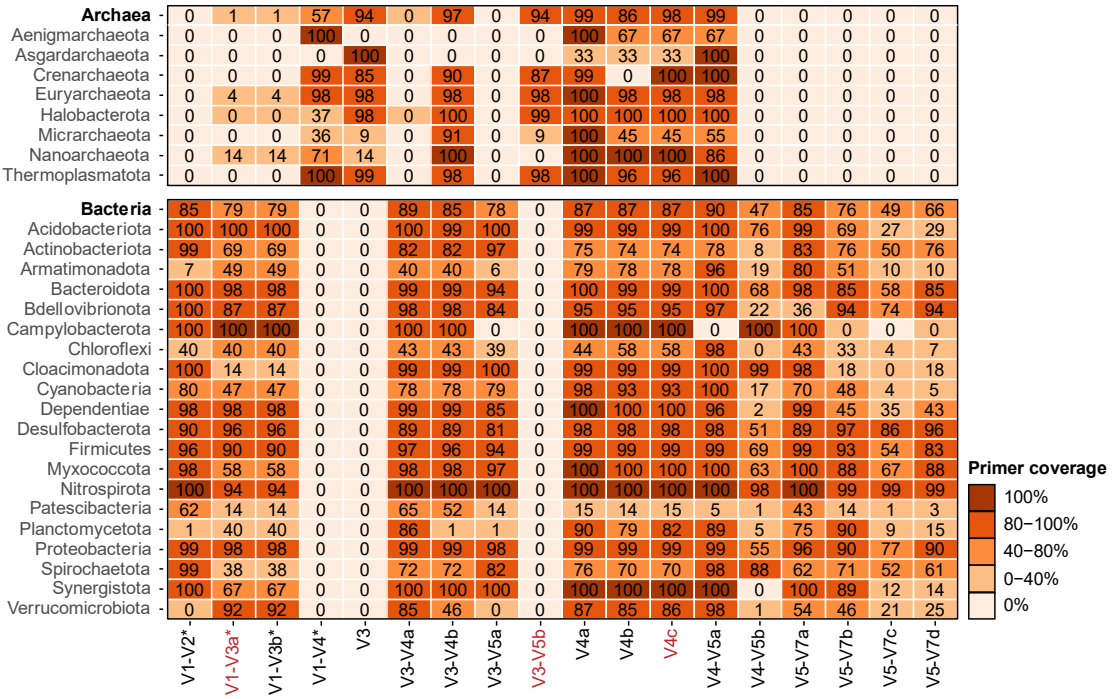

Fig. 3: Ecosystem-specific primer coverage for all archaea and bacteria and the 20 most diverse bacterial and all archaeal phyla based on unique FL-ASVs. The coverage was determined as the percentage of FL-ASVs in the MiDAS 5 database with perfect hits for both forward and reverse primers. The primer pairs marked in red were used in the current study. Detailed information of all primer pairs and coverage information for all taxa in MiDAS 5 are provided in Supplementary Data 2. *Only the reverse primer was evaluated for these primer pairs because the forward primer was used to create the MiDAS 5 reference sequences. The coverage might therefore be overestimated.

Evaluation of 16S rRNA gene amplicon primers for community profiling of ADs

The comprehensive ASV-resolved MiDAS 5 database provides a unique opportunity to determine the theoretical coverage of commonly applied 16S rRNA gene amplicon primer pairs for bacteria and archaea in ADs (Fig. 3). This information is highly valuable when designing experiments, especially if targeting specific taxa. Accordingly, we have determined the theoretical coverage for several commonly applied primer pairs for all kingdom to species-level taxa in MiDAS 5 (Supplementary Data 2). We found a fairly low coverage of the V1-V3 primer pair (perfect hits for ≤79% of the bacterial FL-ASVs), which we commonly use due to its high phylogenetic resolution [6,18]. We should therefore expect a bias when using this primer pair, although it has previously shown good correlation with quantitative FISH results [19]. The V4c primer used here and in the Earth Microbiome project [35] showed good coverage for both bacteria (perfect hits for 87% of the FL-ASVs) and archaea (perfect hits for 98% of the FL-ASVs). It may therefore be applied for profiling the entire prokaryotic diversity in ADs. However, it should be noted that the V4 amplicon has a lower phylogenetic resolution which prevents species-level classification for many genera [6,29]. Interestingly, we observed an even higher coverage for both bacteria (perfect hits for 90% of the FL-ASVs) and archaea (perfect hits for 99% of the FL-ASVs) with the V4-V5a primer pair. Because this longer amplicon also carries a stronger phylogenetic resolution, we would recommend this primer pair for general profiling of ADs in future studies. The V3-V5b primer pair, used here to target archaea, had a good coverage for archaea, although not as good as the V4c and V4-V5a universal primers. In most cases, we would therefore recommend the application of universal primers.

Effect of process and environmental factors on the AD microbiota

Alpha diversity analyses showed that the rarefied (10,000 read per sample) ASV richness and inverse Simpsons diversity in ADs were affected mainly by the primary substrate type and the temperature in the ADs (Supplementary Fig. 4). Significantly higher bacterial richness and diversity were observed for ADs treating surplus sludge from WWTPs compared to the other types of substrates. This effect likely reflects the extensive immigration of bacteria into the ADs with the surplus sludge [7,10,36]. A higher richness and diversity were observed for bacteria in mesophilic ADs compared to thermophilic ADs. A similar trend has previously been observed for full-scale ADs treating manure [37,38], household waste [39], and surplus sludge from WWTPs [7].

Genus-level taxonomic beta-diversity was used to investigate the effect of process conditions and geography on the overall microbiota in ADs using principal coordinate analysis (PCoA) and permutational multivariate analysis of variance (PERMANOVA) (Fig. 4). We used this approach because many of the important traits are categorical (yes/no) and only conserved at lower taxonomic ranks (genus/species). Furthermore, MiDAS 5 enabled us to classify almost all our ASVs at the genus-level, thereby providing a comprehensive description of the microbiota. We found that the overall

microbial community could be explained mainly by the primary substrate and to a lesser extent temperature, continent, and digester technology (Fig. 4). This trend was observed for both bacteria and archaea. The percentage of total variation explained by each parameter was, except for the primary substrate, low, suggesting that the global AD microbiota represents a continuous distribution rather than distinct states, as also observed for the human gut microbiota [40] and WWTPs [18]. The pronounced effect of the primary substrates highlights that these substrates are distinct and abundant in microbes, particularly in the case of manure and wastewater sludge.

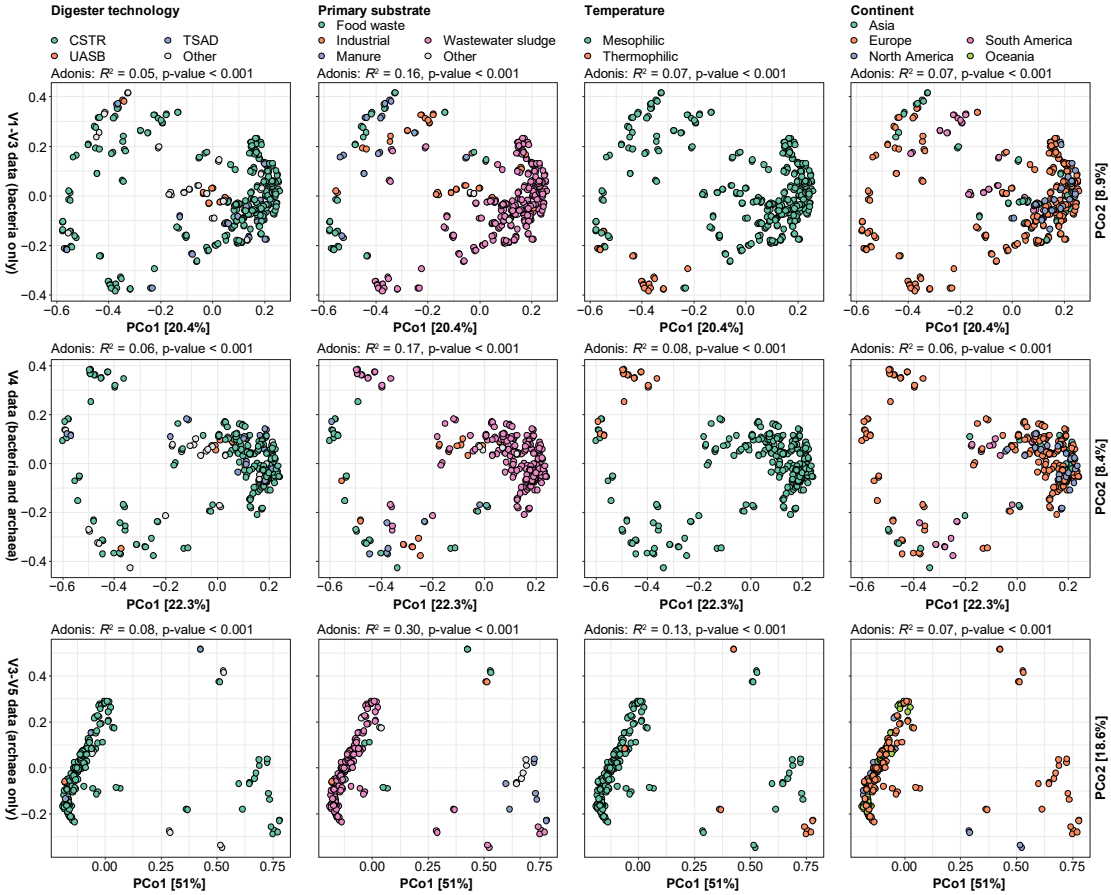

Fig. 4: Effects of process and environmental factors on the anaerobic digester microbiota. Principal coordinate analyses of Bray-Curtis beta-diversity for genera based on V1-V3 (bacteria), V4 (archaea and bacteria), and V3-V5 (archaea) amplicon data. Samples are colored based on metadata. The fraction of variation in the microbial community explained by each variable in isolation was determined by PERMANOVA (Adonis R^2-values). Exact P-values <0.001 could not be confidently determined due to the permutational nature of the test. CSTR: Continuous stirred-tank reactor; TSAD: Two-stage anaerobic digestion; UASB: Upflow anaerobic sludge blanket.

308 ***Core and conditional rare or abundant taxa in the global AD microbiota***
309 The global AD microbiota represents a huge microbial diversity. However, most
310 organisms only occur in very low abundance and are therefore unlikely to have any
311 quantitative impact on the overall metabolism and the process performance in ADs.
312 Analysis of core and conditionally rare or abundant taxa (CRAT) is a powerful
313 approach to identify the most important genera and species within a specific ecosystem
314 [18,25,41]. Core taxa are commonly defined based on how frequently specific taxa are
315 observed in samples from a well-defined habitat [42], whereas CRAT are typically present
316 in low abundance but can occasionally become very abundant [43,44]. The CRAT may
317 include taxa related to process disturbances, such as filamentous microbes associated
318 with foam formation, or taxa associated with the degradation of special substrates found
319 in, e.g., industrial waste.
320
321 We recently introduced and applied the following core and CRAT definitions in our
322 survey of the global microbiota of wastewater treatment plants: strict core (>0.1%
323 relative abundance in >80% of samples), general core (>0.1% relative abundance in
324 >50% of samples), loose core (>0.1% relative abundance in >20% of samples), and
325 CRAT (not part of the core, but present in at least one sample with a relative abundance
326 >1%) [18]. Here, we applied the same criteria to identify core and CRAT genera and
327 species in our global AD dataset. Because the primary substrate showed a strong effect
328 on the overall microbial community (Fig. 4), we determined the core and CRAT for
329 each individual substrate separately (Supplementary Data 3). Only mesophilic ADs
330 were examined for ADs treating food waste, industrial waste, and manure due to the
331 low number of thermophilic ADs sampled. Both mesophilic and thermophilic digesters
332 were examined for ADs treating wastewater sludge. All three amplicon datasets were
333 examined, and the results combined to reduce the effect of primer bias.
334
335 The core analysis revealed that most core genera were uniquely associated with specific
336 primary substrates and temperature range (Fig. 5a). However, there were also a
337 significant number of core genera shared across substrates (Fig. 5a). In contrast, very
338 few core species were shared between ADs treating different primary substrates (Fig.
339 5b). This fits well with similar results from a study of ADs in Belgium and Luxemburg
340 [13]. To define a most wanted list for bacteria and archaea in ADs globally, we linked
341 each core and CRAT with the best classification across primary substrates and process
342 temperature (Supplementary Data 3). The resulting list contained 506 core genera (71
343 strict, 122 general, and 313 loose) and 152 CRAT genera. The strict core genera
344 included 11 known methanogens and four known syntrophs (*Smithella*,
345 *Syntrophaceticus*, *Syntrophomonas*, *Syntrophorhabdus*). At the species-level, we
346 identified 597 core (27 strict, 127 general, and 443 loose) and 339 CRAT species. The
347 strict core species included three methanogens (*Methanobrevibacter smithii*,
348 *Methanosphaera stadtmanae* and *Methanothermobacter* midas_s_3958) and one
349 syntroph (*Syntrophomonas* midas_s_90707). It is worth noting that a large fraction

of the taxa observed in ADs is unable to grow, but only occurs because the taxa are in high abundance in the feed [7,10,36]. Data from Danish ADs treating wastewater sludge [7] showed that 140 (28%) of the core genera and 21 (3.5%) of the core species were non-growing (<20% of ASVs belonging to the specific taxa were classified as growing), but it is not known whether this also translates to global ADs.

Many core and CRAT are only identified due to the MiDAS placeholder taxonomy

A large proportion of core and CRAT identified was classified as MiDAS *de novo* taxa. At the genus-level, 281/506 (56%) of the core genera and 100/152 (66%) of the CRAT genera had MiDAS placeholder names, and at the species-level, the proportion was even higher. Here placeholder names were assigned to 552/597 (92%) of the core species and 322/339 (95%) CRAT species. These proportions are similar to those observed for the global microbiota in activated sludge [18] and reveals the importance of a taxonomic framework that can handle uncultured taxa which have not yet been officially classified.

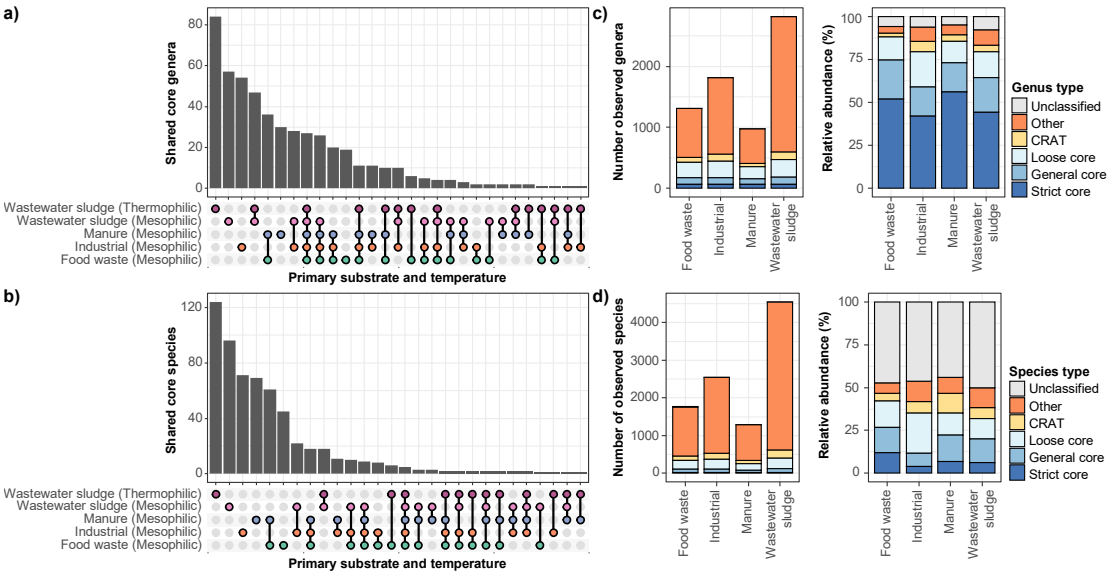

Fig. 5: Core and conditionally rare or abundant taxa (CRAT) in anaerobic digesters globally. a) and b) UpSet plots displaying the number of shared core genera and species, respectively, across ADs treating different primary substrates and operating at different temperatures. c) and d) Number of observed genera and species, respectively, and their abundance in mesophilic ADs treating different primary substrates based on V4 amplicon data (bacteria and archaea). Values for genera and species are divided into strict core, general core, loose core, CRAT, other taxa, and unclassified ASVs based on the most wanted list (Supplementary Data 3). The relative abundance of different groups was calculated based on the mean relative abundance of individual genera or species across samples. Similar figures for V1-V3 (bacteria only) and V3-V5 (archaea only) amplicons data can be found in Supplementary Fig. 5.

376 ***The global AD microbiota is dominated by core and CRAT taxa***
377 Despite only accounting for a minor fraction of the total diversity in the global AD
378 microbiota, the core and CRAT represented most of the microbes in the ecosystem
379 according to relative amplicon read abundance (Fig. 5c, 5d, Supplementary Fig. 5). The
380 core and CRAT genera accounted for 86-93% (V1-V3), 83-90% (V4), and 95-99%
381 (V3-V5) of the accumulated read abundance in mesophilic ADs depending on primary
382 substrates. The remaining fractions consisted mainly of ASVs unclassified at the genus
383 level, and genera present in very low abundance, presumably with minor importance
384 for the AD performance.
385
386 For the species level, the core and CRAT represented 54-64% (V1-V3), 38-47% (V4),
387 and 10-58% (V3-V5) accumulated read abundance depending on the primary substrate.
388 The remaining fractions were mainly composed of ASVs, which could not be classified
389 at the species level, probably due to insufficient phylogenetic resolution in the short-
390 read amplicons [6,29]. The lack of species-level classification was especially pronounced
391 for the archaeal V3-V5 ASVs in ADs treating industrial waste, manure, and wastewater
392 sludge (Fig. 5).
393
394 The large relative abundance of core and CRAT in the global AD microbiota suggests
395 that we can explain a substantial part of the metabolic processes in ADs, if we
396 understand the physiology and metabolic potential of these taxa.
397
398 ***Global diversity of archaea reveals new potential methanogens***
399 As methanogenic archaea are ultimately responsible for the generation of methane in
400 ADs, we examined the global diversity of archaea in all samples based on V4 (Fig. 6)
401 and V3-V5 amplicon data (Supplementary Fig. 6). The V4 amplicon data,
402 encompassing both archaea and bacteria, showed that the archaeal reads constituted
403 3.9%±3.0% for ADs treating food waste, 4.6%±2.8% for manure, 6.0%±2.4% for
404 wastewater sludge, and 14.2%±12.2% for industrial waste. Many of the abundant
405 archaea represented well-known methanogens. However, we also observed several
406 abundant genera, only classified based on the MiDAS placeholder taxonomy, affiliated
407 to orders and families of known methanogens. These include midas_g_90791 and
408 midas_g_8154, which represent new families within the orders Methanomicrobiales
409 and Methanofastidiosales, respectively, and midas_g_90473 and midas_g_93310,
410 representing new genera within Methanomassiliicoccaceae and Methanospirillaceae,
411 respectively. In addition, we observed two abundant MiDAS placeholder genera
412 (midas_g_90791 and midas_g_97217) that represent a new order within the class *Ca.*
413 Bathyarchaeia. Members of his class can have a versatile metabolism, and some encode
414 the key methanogenic enzyme methyl-coenzyme M reductase (MCR) [45,46]. Targeted
415 metagenomics and assembly of MAGs should be applied to confirm the methanogenic
416 potential of these new potential methanogens, and our amplicon datasets provide insight
417 into where these taxa occur in high abundance.

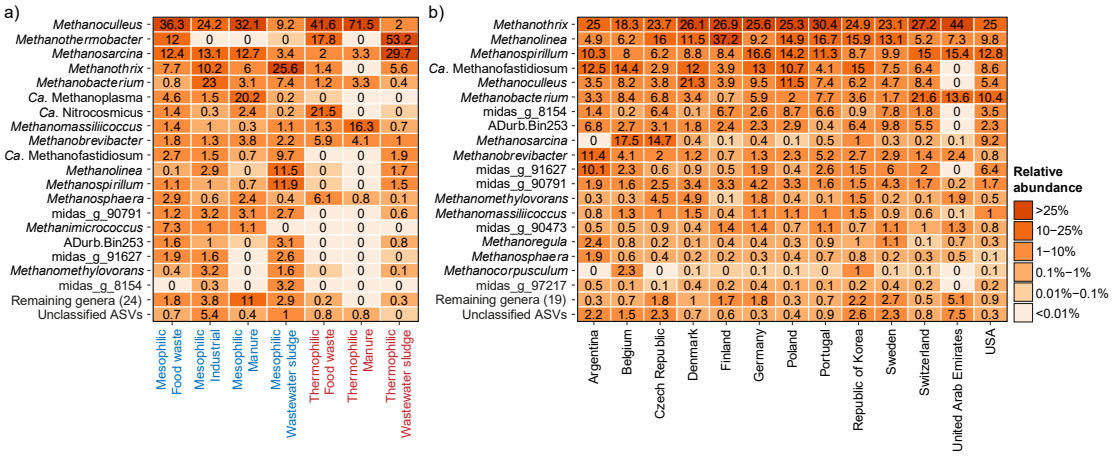

Fig. 6: Top 25 archaeal genera based on V4 amplicon data. The percent relative abundance represents the mean abundance relative to all archaea across a) different primary substrates and temperature range, and b) different countries considering only mesophilic ADs treating mainly wastewater sludge.

The methanogenic community composition was clearly affected by the primary substrate and whether the digestion was performed at mesophilic or thermophilic conditions (Fig. 6a, Supplementary Fig. 6a). The most common methanogens across substrates and temperatures were *Methanoculleus*, *Methanosarcina*, *Methanothrix*, and *Methanothermobacter*, although the latter was only abundant in thermophilic ADs treating wastewater sludge and food waste and in a single mesophilic reactor treating food waste, indicating that this digester might use a two-stage thermophilic-mesophilic configuration.

Because most of our samples originated from mesophilic reactors treating wastewater sludge, we examined the diversity of methanogens across countries in these ADs (Fig. 6b, Supplementary Fig. 6b). This analysis revealed that the same genera were dominating across the world. The most common methanogens in these ADs were *Methanothrix*, *Methanolinea*, *Methanospirillum*, *Methanoculleus*, and the recently discovered *Ca*. Methanofastidiosum [47] (Fig. 6b, Supplementary Fig. 6b). Next, we examined if the methanogens were also conserved at higher phylogenetic resolution. As many archaeal ASVs could not be classified at the species-level, we examined the global diversity at the ASV-level (Supplementary Fig.7). We found that the vast majority of the abundant ASVs occurred globally. This high similarity among methanogens across the world suggests a great potential for knowledge transfer between countries.

Among the highly abundant archaea, we also observed an ammoniaoxidizing archaeon (AOA) from the genus *Ca*. Nitrosocosmicus [48], which was especially abundant in thermophilic ADs treating food waste. This is very surprising and may indicate that they also have an anaerobic physiology which should be investigated further. Another abundant archaeon was the *Ca*. Diapherotrites ADurb.bin253 belonging to the order

Woesearchaeales which are characterized by ultra-small genomes and an anaerobic and parasitic/fermentation-based lifestyle [49].

Global diversity of syntrophic bacteria

Syntrophic bacteria play a vital role in ADs by converting substrates, such as short-chain fatty acids to acetate, H_2, and formate, which act as substrates or reducing equivalents for methanogens to produce methane and CO_2 [27,50,51]. This obligately mutualistic metabolism between the syntrophs and methanogens is critical because substrate oxidation can only occur and sustain growth of the syntrophs under anaerobic conditions if the products are kept at a very low concentration by rapid consumption by the methanogens [50,52]. Due to the fastidious metabolism, syntrophs are usually present in low abundance, and can easily become the bottleneck in the anaerobic digestion process [7,8]. Accordingly, we investigated the global diversity of this functional guild in the ADs sampled (Fig. 7, Supplementary Fig. 8).

A clear effect of the primary substrates and digester temperature was observed on the composition and abundance of syntrophic genera in the digesters (Fig. 7a, Supplementary Fig. 8a). The most abundant genus across substrates and temperature was *Syntrophaceticus*, despite being barely detected in ADs treating wastewater sludge. The type strain of this genus, *S. schinkii* Sp3^T, is an acetate-oxidizing syntroph that thrives, and has a competitive advantage, under high ammonium concentrations (up to 8,400 mgN/L) [53,54]. The lack of *Syntrophaceticus* in ADs treating wastewater sludge may therefore be explained by lower ammonium concentrations in these ADs (1,617±4,312 mgN/L, n=145) compared to those treating food waste (2,913±1,681 mgN/L, n=33), and manure (3,449±933 mgN/L, n=18).

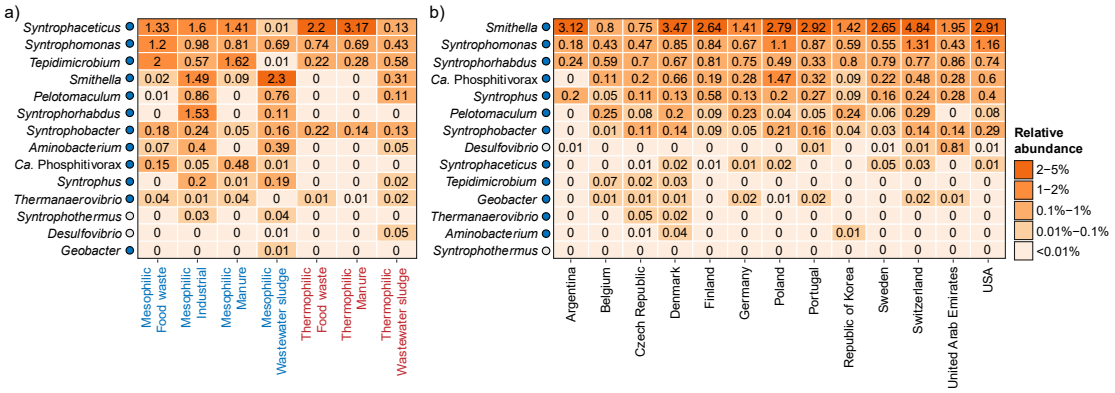

Fig. 7: Global diversity of syntrophs based on V1-V3 amplicon data. The percent relative abundance represents the mean for genera across a) different primary substrates and temperature range, and b) different countries considering only mesophilic ADs treating mainly wastewater sludge. Colored circles next to the genus labels indicate whether the genera have previously been identified as growing in ADs at Danish WWTPs according to Jiang *et al.* [7]. Blue: >50% of ASVs classified as growing; Yellow: 20-50% of ASVs classified as growing. Red: <20% of ASVs classified as growing. Gray: No information available for the specific genus.

484 *Syntrophomonas*, the second most abundant genus, was common in all AD types
485 investigated, indicating a broader ecological niche. Isolated representatives from this
486 genus can grow syntrophically via β-oxidation of saturated fatty acids of various lengths
487 (C4-C18, depending on strain) [55–58], and they are therefore likely important for the
488 conversion of long-chain fatty acids in ADs. Among the abundant syntrophs,
489 *Tepidimicrobium,* a member of the order Clostridiales, was also observed in all AD
490 types except mesophilic ADs treating wastewater sludge. The exact metabolism of
491 *Tepidimicrobium* in ADs remains to be determined, however all isolated representatives
492 can degrade proteinaceous compounds and some species can also use carbohydrates [59].
493 Furthermore, *Tepidimicrobium* has been proposed to grow syntrophically by direct
494 interspecies electron transfer (DIET) with *Methanothermobacter* in a process like that
495 observed for *Geobacter* [60]. Accordingly, it is likely that the *Tepidimicrobium* acts as a
496 syntrophic primary degrader in the ADs targeting mainly proteins, carbohydrates, and
497 derivatives.
498
499 Finally, we observed a high abundance of the genus *Smithella* in mesophilic ADs
500 treating industrial waste, manure, and wastewater sludge. The type strain *S. propionica*
501 LYP^T is a propionate oxidizing syntroph, which uses a unique dismutation pathway in
502 which propionate is first converted to acetate and butyrate, and the latter is hereafter β-
503 oxidized syntrophically to acetate and hydrogen [61,62]. Calculations of Gibbs free energy
504 for this special propionate metabolism indicates a higher tolerance toward elevated
505 hydrogen concentrations [63], which could explain why some *Smithella* prevail in certain
506 ADs. However, *Smithella* has also been implicated in the syntrophic degradation of
507 long-chain alkanes [64,65], which could reflect a more versatile metabolism.
508
509 When investigating geographical diversity of syntrophic fatty acid oxidizing bacteria
510 in mesophilic ADs treating wastewater sludge, a similar pattern was observed across
511 countries (Fig. 7b, Supplementary Fig. 8b). *Smithella*, was generally the dominating
512 syntroph. However, *Syntrophomonas*, *Syntrophorhabdus*, *Ca*. Phosphitivorax, and
513 *Syntrophus* also occurred at a fairly high relative abundance in almost all countries.
514 Isolates of *Syntrophorhabdus,* including the type strain *S. aromaticus* UI^T, are
515 syntrophic fermenters of aromatic compounds and may accordingly play an important
516 role in the detoxification of these substrates in ADs [66,67]. *Ca*. Phosphitivorax was
517 recently discovered as a butyrate degrading syntroph by genome-resolved meta-
518 transcriptomics in a digester treating wastewater sludge [51], and *Syntrophus* participates
519 in the degradation of fatty acids and aromatics [68,69]. Overall, the results suggest a
520 complex syntrophic degradation process, which involves multiple genera with different
521 substrate specificities.
522
523 To gain additional insight into the global diversity of syntrophs, we also investigated
524 the species-level diversity across mesophilic digesters treating wastewater sludge
525 (Supplementary Fig. 9). We observed a large species diversity among most of the

526 abundant syntrophic genera. Furthermore, we found that the most abundant species in
527 the ADs were often distinct from the isolated representatives, which prompts for further
528 investigations into the metabolic potential of syntrophs *in situ*.
529
530 ***Global diversity of filamentous bacteria***
531 Foaming is a common operational problem in ADs and has a strong negative impact on
532 process performance resulting in considerable costs. Both abiotic and biotic factors are
533 involved in foaming [70]. The abiotic factors include high loading rates of surfactants (oil,
534 grease, fatty acids, detergent, proteins, and particulate matter) and biosurfactants
535 produced by microbes in the digester [71]. The biotic factors cover increased abundance
536 of hydrophobic, filamentous microorganisms that can interact with, and stabilize, gas
537 bubbles in the foam [70,72]. To gain further insight into potential foam forming microbes,
538 we examined the global diversity of known filamentous bacteria in ADs (Fig. 8,
539 Supplementary Fig. 10).
540
541 The diversity and mean relative abundance of known filamentous organisms were
542 generally low in the ADs examined except for those treating wastewater sludge (Fig.
543 8a, Supplementary Fig. 10). However, the increased diversity and abundance in the
544 latter are to a large extent the result of passive immigration from the fed surplus sludge.
545 However, most of these are unable to grow in the ADs [7] (Fig. 8a, Supplementary Fig.
546 10). *Anaerolinea*, *Ca*. Brevefilum, and *Trichococcus* were common across ADs treating
547 all primary substrates (Fig. 8a, Supplementary Fig. 10), whereas *Ca*. Microthrix and
548 *Ca.* Promineofilum were mainly observed in ADs treating wastewater sludge. Many of
549 the Chloroflexi genera found here were also observed in a recent meta-analysis of
550 amplicon data from 17 studies representing 62 ADs [73]. Several of the abundant
551 filamentous genera, including *Ca*. Microthrix and *Ca*. Brevefilum, were previously
552 found to correlate with the foaming potential of full-scale digester sludge from
553 mesophilic ADs at WWTPs [72]. *Ca*. Brevefilum seems especially interesting as it grows
554 well in ADs [7,74].
555
556 The species-level diversity was generally low for the filamentous bacteria
557 (Supplementary Fig. 11). *Ca*. Brevefilum was dominated by *Ca*. B. fermentans,
558 *Trichococcus* by midas_s_4, *Ca*. Microthrix by *Ca.* M. parvicella and *Ca.* M.
559 subdominans, and *Gordonia* by *G. amarae* and *G. defluvii*. *Ca.* Promineofilum was
560 more diverse with five common species. The low species-level diversity of potential
561 foam-forming bacteria suggests that it may be feasible to develop and implement
562 universal mitigation strategies for these bacteria in ADs worldwide.

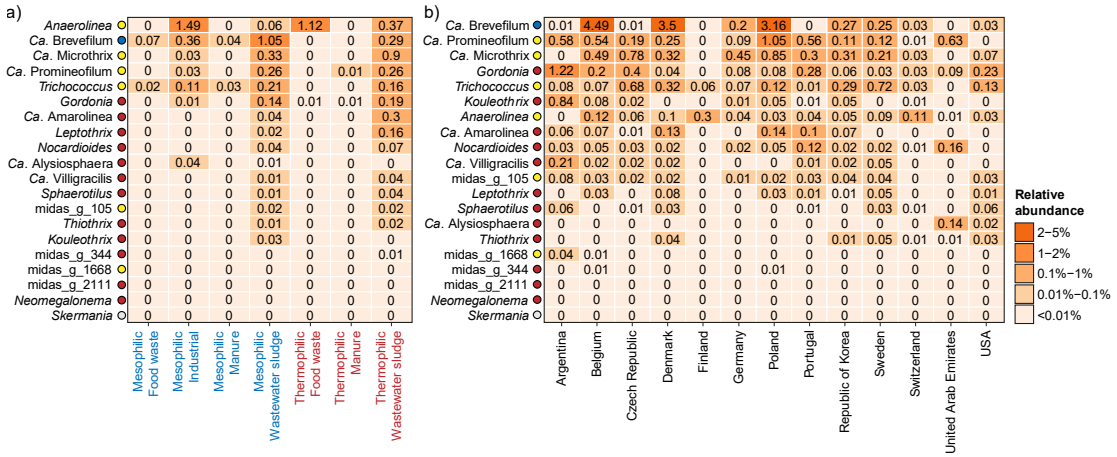

Fig. 8: Global diversity of known filamentous organisms based on V1-V3 amplicon data. The percent relative abundance represents the mean for genera across a) different primary substrates and temperature range, and b) different countries considering only mesophilic ADs treating mainly wastewater sludge. Colored circles next to the genus labels indicate whether the genera have previously been identified as growing in ADs at Danish WWTPs according to Jiang *et al.* [7]. Blue: >50% of ASVs classified as growing; Yellow: 20-50% of ASVs classified as growing. Red: <20% of ASVs classified as growing. Gray: No information available for the specific genus.

Conclusion and perspectives

MiDAS 5 was made possible thanks to a huge collaborative effort from experts worldwide, who contributed to the project by sampling and providing metadata for ADs in their respective countries. Building on the success of its predecessor, MiDAS 4, this latest expansion covers ASV-resolved, full-length 16S rRNA gene references from numerous ADs from all parts of the globe covering different operations parameters and different substrates. This expanded database provides greatly improved coverage for AD-specific taxa and a strongly needed taxonomy for uncultured lineages, which lacks official taxonomic classification. As such, it will be an invaluable resource for researchers and AD professionals, providing them with a common point of reference to facilitate knowledge sharing and pave the way for a comprehensive understanding of the AD microbiome.

Our in silico 16S rRNA gene primer evaluation based on the MiDAS 5 database revealed that the coverage of commonly applied primer pairs varies significantly, with some having low coverage and potential bias towards certain taxa. Because the primer coverage was evaluated for all taxa in the MiDAS 5 database at all taxonomic rank, it provides a solid foundation for designing experiments and targeting specific taxa in future studies. For general microbial profiling of ADs, we would recommend the use of the V4a or the V4-V5a primer pairs, as they both show excellent coverage for archaea and bacteria in the ecosystem.

593 Although the total microbial diversity in ADs is huge, importantly, we showed that less
594 than 1000 genera and species accounted for most of the microbes in the AD ecosystem.
595 By focusing on these abundant and common microbes, we will be able to explain most
596 of the microbial processes that occur in the anaerobic digestion process. This list, the
597 "Most wanted" organisms, contain species that should be prime targets for future in situ
598 studies and genome reconstruction projects.
599
600 The global survey of the ADs microbiota using three different primer pairs provided a
601 unique insight into the global diversity of individual AD taxa and clues into the
602 environmental and operational factors that define their ecological niches. This
603 information will be invaluable in the development of future microbiome management
604 strategies and improved sustainability of the field of anaerobic digestion.
605
606 To enhance knowledge dissemination, we have updated the MiDAS Field Guide
607 available at www.midasfieldguide.org. This dynamic resource allows users to delve
608 into specifics related to the physiology, morphology, and ecology of genera listed in
609 the MiDAS database. Additionally, it offers country-specific data on the prevalence of
610 all MiDAS genera and species in WWTPs and ADs. Finally, it provides information on
611 the availability of fluorescence in situ hybridization probes and reference genomes,
612 paving the way for subsequent research endeavors.

613 **Methods:**

614 ***Sampling and metadata collection***

615 To facilitate sampling of ADs worldwide, we established the MiDAS Global
616 Consortium for Anaerobic Digesters, which consists of 25 anaerobic digestion experts
617 in 19 countries. Members of the consortium acted as national sampling coordinators
618 and were in direct contact with the ADs. Two samples were obtained from each AD
619 and shipped on ice to the sampling coordinators. For each replicate, 2 mL sample was
620 preserved in 2 mL RNAlater (Invitrogen), stored at 4°C until all national samples were
621 collected (usually within a few days), and then shipped to Aalborg University with
622 cooling elements. Upon arrival, the samples were separated into aliquots that were
623 prepared for nucleic acid purification. Metadata associated with each AD was also
624 obtained by the sampling coordinators and is provided as Supplementary Data 1.
625 Minimum information from all ADs included continent, country, GPS coordinates,
626 sampling date, temperature in the digester ("Mesophilic" (≤45°C) or "Thermophilic"
627 (50-60°C)), primary substrate ("Wastewater sludge", "Industrial", "Food waste",
628 "Manure", or "Other"), and digester technology ("Two-stage digester (TSAD)",
629 "Continuous Stirred Tank Reactors (CSTR)", "Plug-flow", "Upflow anaerobic sludge
630 blanket (UASB)", or "Other").

631

632 ***General molecular methods***

633 All commercial kits were used according to the protocols provided by the manufacturer
634 unless otherwise stated. The concentration and quality of nucleic acids were determined
635 using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and an Agilent 2200
636 Tapestation (Agilent Technologies), respectively.

637

638 ***Nucleic acid purification***

639 DNA was purified using a custom plate-based extraction protocol based on the
640 FastDNA spin kit for soil (MP Biomedicals). The protocol is available at
641 www.midasfieldguide.org (aau_ad_dna_v 2.0). RNAlater preserved samples were
642 thawed and homogenized using a Heidolph RZR 2020 laboratory stirrer. 20 µL of
643 sample was resuspended in 300 µL PBS and transferred to Lysing Matrix E barcoded
644 tubes (MP Biomedicals). 40 µL of MT buffer was added and lysis was performed by
645 bead beating in a FastPrep-96 bead beater (MP Biomedicals) (3x 120 s, 1800 rpm with
646 2 min incubation on ice between cycles). The samples were centrifuged (3,486 ×g, 10
647 min) and 200 µL supernatant was transferred to a 96-well PCR-plate. 50 µL Protein
648 Precipitation Solution (PPS) was mixed with each sample, which was then centrifuged
649 again. 150 µL supernatant was cleaned-up using 100 µL CleanNGS beads with elution
650 into 60 µL of nuclease-free water. 40 µL of the purified DNA was transferred to a new
651 96-well plate and stored at -80°C.

652 ***Full-length 16S rRNA gene library preparation, sequencing, and processing***
653 Full-length 16S rRNA gene sequencing was carried out using high-accuracy, long-read
654 amplicon sequencing using UMIs and PacBio CCS [75]. Oligonucleotides used can be
655 found in Supplementary Table 1. Bacterial and archaeal 16S rRNA genes were UMI-
656 tagged using overhang primers based on the 27F and 1391R [76] and SSU1ArF and
657 SSU1000ArR [32] primer pairs, respectively. These primers have shown excellent
658 coverage for the known bacterial and archaeal diversity in silico [32,76].
659
660 *Addition of UMI-tags by overhang PCR:* Adaptors containing UMIs, and defined
661 primer binding sites were added to each end of the bacterial and archaeal 16S rRNA
662 genes by PCR. The reaction contained 20 µL of 5x SuperFi Buffer (Invitrogen), 2 µL
663 of 10 mM dNTP mix, 5 µL of 10 µM f16S_pcr1_fw, 5 µL of 10 µM f16S_pcr1_rv, 1
664 µL of 2 U/µL Platinum SuperFi DNA polymerase (Invitrogen), 100 ng of pooled
665 template DNA (from all ADs), and nuclease-free water to 100 µL. The reaction was
666 incubated with an initial denaturation at 98°C for 30 s followed by 2 cycles of
667 denaturation at 98°C for 20 s, annealing at 55°C for 30 s, and extension at 72°C for 45
668 s, and then a final extension at 72°C for 5 min. The sample was purified using 0.6x
669 CleanNGS beads and eluted in 20 µL nuclease-free water.

670 *Primary library amplification:* The tagged 16S rRNA gene amplicons were amplified
671 using PCR to obtain enough product for quantification. The reaction contained 19 µL
672 of UMI-tagged sample, 20 µL 5x SuperFi buffer (Invitrogen), 2 µL of 10 mM dNTP, 5
673 µL of 10 µM f16S_pcr2_fw, 5 µL of 10 µM f16S_pcr2_rv, 48 µL nuclease-free water,
674 and 1 µL 2U/µL Platinum SuperFi DNA polymerase (Invitrogen). The reaction was
675 incubated with an initial denaturation at 98°C for 30 s followed by 15 cycles of
676 denaturation at 98°C for 20 s, annealing at 60°C for 30 s, and extension at 72°C for 45
677 s and then a final extension at 72°C for 5 min. The PCR product was purified using
678 0.6x CleanNGS beads and eluted in 11 µL nuclease-free water. The amplicons were
679 validated on a Genomic screentape and quantified with the Qubit dsDNA HS assay kit.

680 *Clonal library amplification:* Tagged amplicon libraries were diluted to approximately
681 250,000 molecules/µL and amplified by PCR to obtain clonal copies of each uniquely
682 tagged amplicon molecule. Three libraries were made for the bacterial 16S rRNA genes
683 and one for archaea. The PCR reactions contained 1 µL diluted primary library, 20 µL
684 5x SuperFi buffer (Invitrogen), 2 µL of 10 mM dNTP, 5 µL of 10 µM f16S_pcr2_fw,
685 5 µL of 10 µM f16S_pcr2_rv, 66 µL nuclease-free water, and 1 µL 2U/µL Platinum
686 SuperFi DNA polymerase (Invitrogen). The reaction was incubated with an initial
687 denaturation at 98°C for 30 s followed by 25 cycles of denaturation at 98°C for 20 s,
688 annealing at 60°C for 30 s, and extension at 72°C for 45 s and then a final extension at
689 72°C for 5 min. The PCR product was purified using 0.6x CleanNGS beads and eluted
690 in 20 µL nuclease-free water. The amplicons were validated on a Genomic screentape
691 and quantified with the Qubit dsDNA HS assay kit.

692 *PacBio CCS sequencing:* The four clonal libraries were sent to Admera Health
693 (Plainfield, NJ, USA) for PacBio library preparation and sequencing. Here amplicons
694 were incubated with T4 polynucleotide kinase (New England Biolabs) following the
695 manufacturer's instructions, and sequencing library prepared using SMRTbell Express
696 Template Preparation kit 1.0 following the standard protocol. Sequencing was
697 performed using 4x SMRT cells on a Sequel II using a Sequel II Sequencing kit 1.0,
698 Sequel II Binding and Int Ctrl kit 1.0 and Sequel II SMRT Cell 8M, following the
699 standard protocol with 1 h pre-extension and 15 h collection time (Pacific Biosciences).

700 *Bioinformatic processing:* CCS reads were generated from raw PacBio data using CCS
701 v.3.4.1 (https://github.com/PacificBiosciences/ccs) with default settings. UMI
702 consensus sequences (consensus_raconx3.fa) were obtained using the longread_umi
703 script (https://github.com/SorenKarst/longread_umi) [75] using the following options:
704 pacbio_pipeline, -v 3, -m 1000, -M 2000, -s 60, -e 60, -f
705 CAAGCAGAAGACGGCATACGAGAT, -F AGRGTTYGATYMTGGCTCAG
706 (bacteria) or TCCGGTTGATCCYGCBRG (archaea), -r
707 AATGATACGGCGACCACCGAGATC, -R GACGGGCGGTGWGTRCA (bacteria)
708 or GGCCATGCAMYWCCTCTC (archaea), and -c 3. The UMI-consensus reads were
709 oriented based on the SILVA 138.1 SSURef NR99 database using the usearch
710 v.11.0.667 -orient command and trimmed between the 27f and 1391r (bacteria) or
711 SSU1ArF and SSU1000ArR (archaea) primer binding sites using the trimming function
712 in CLC genomics workbench v. 20.0. Sequences without both primer binding sites were
713 discarded. The trimmed high-fidelity reads were processed with AutoTax v. 1.7.4 [6] to
714 create full-length 16S rRNA gene amplicon sequence variants (FL-ASV) and these
715 were added to the MiDAS 4.8.1 reference database [18] to create MiDAS 5.0.

716 ***Short-read amplicon sequencing***
717 V1-V3 amplicons were made using the 27F (5'-AGAGTTTGATCCTGGCTCAG-3')
718 [77] and 534R (5'-ATTACCGCGGCTGCTGG-3') [78] primers with barcodes and Illumina
719 adaptors (IDT) [79]. 25 µL PCR reactions in duplicate were run for each sample using 1X
720 PCRBIO Ultra Mix (PCR Biosystems), 400 nM of both forward and reverse primer,
721 and 10 ng template DNA. PCR conditions were 95°C, for 2 min followed by 20 cycles
722 of 95°C for 20 s, 56°C for 30 s, and 72°C for 60 s, followed by a final elongation at
723 72°C for 5 min. PCR products were purified using 0.8x CleanNGS beads and eluted in
724 25 µL nuclease-free water.
725
726 V3-V5 amplicons were made using the Arch-340F (5'-CCCTAHGGGGYGCASCA-
727 3') and Arch-915R (5'-GWGCYCCCCCGYCAATTC-3') primers [80]. 25 µL PCR
728 reactions in duplicate were run for each sample using 1X PCRBIO Ultra Mix (PCR
729 Biosystems), 400 nM of both forward and reverse primer, and 10 ng template DNA.
730 PCR conditions were 95°C, for 2 min followed by 30 cycles of 95°C for 15 s, 55°C for
731 15 s, and 72°C for 50 s, followed by a final elongation at 72°C for 5 min. PCR products
732 were purified using 0.8x CleanNGS beads and eluted in 25 µL nuclease-free water. 2

733 µL of purified PCR product from above was used as template for a 25 µL Illumina
734 barcoding PCR reaction containing 1x PCRBIO Reaction buffer, 1 U PCRBIO HiFi
735 Polymerase (PCR Biosystems) and 10 µL of Nextera adaptor mix (Illumina). PCR
736 conditions were 95°C, for 2 min, 8 cycles of 95°C for 20 s, 55°C for 30 s, and 72°C for
737 60 s, followed by a final elongation at 72°C for 5 min. PCR products were purified
738 using 0.8x CleanNGS beads and eluted in 25 µL nuclease-free water.
739
740 V4 amplicons were made using the 515F (5'-GTGYCAGCMGCCGCGGTAA-3') [78]
741 and 806R (5'-GGACTACNVGGGTWTCTAAT-3') [81] primers. 25 µL PCR reactions
742 in duplicate were run for each sample using 1X PCRBIO Ultra Mix (PCR Biosystems),
743 400 nM of both forward and reverse primer, and 10 ng template DNA. PCR conditions
744 were 95°C, for 2 min followed by 30 cycles of 95°C for 15 s, 55°C for 15 s, and 72°C
745 for 50 s, followed by a final elongation at 72°C for 5 min. PCR products were purified
746 using 0.8x CleanNGS beads and eluted in 25 µL nuclease-free water. 2 µL of purified
747 PCR product from above was used as template for a 25 µL Illumina barcoding PCR
748 reaction as described for the V3-V5 amplicons.
749
750 16S rRNA gene V1-V3, V3-V5, and V4 amplicon libraries were pooled separately in
751 equimolar concentrations and diluted to 4 nM. The amplicon libraries were paired-end
752 sequenced (2 × 300 bp) on the Illumina MiSeq using v3 chemistry (Illumina, USA). 10
753 to 20% PhiX control library was added to mitigate low diversity library effects.
754
755 ***Processing of short-read amplicon data***
756 Usearch v.11.0.667 [82] was used for processing of 16S rRNA gene amplicon data and
757 for read mapping. V1-V3 forward and reverse reads were merged using the usearch -
758 fastq_mergepairs command, filtered to remove phiX sequences using usearch -
759 filter_phix, and quality filtered using usearch -fastq_filter with -fastq_maxee 1.0.
760 Dereplication was performed using -fastx_uniques with -sizeout, and amplicon
761 sequence variants (ASVs) were resolved using the usearch -unoise3 command. An
762 ASV-table was created by mapping the quality filtered reads to the ASVs using the
763 usearch -otutab command with the -zotus and -strand plus options. Taxonomy was
764 assigned to ASVs using the usearch -sintax command with -strand both and -
765 sintax_cutoff 0.8 options. Mapping of ASVs to reference databases was done with the
766 usearch -usearch_global command and the -id 0, -maxaccepts 0, -maxrejects 0, -
767 top_hit_only, and -strand plus options.
768
769 16S rRNA gene V3-V5 forward reads (reverse reads in relation the 16S rRNA gene)
770 were filtered to remove phiX sequences using usearch -filter_phix, trimmed to remove
771 primers and obtain a fixed length of 250 bp using -fastx_truncate with -stripleft -17 and
772 trunclen 250, reverse complemented with usearch -fastx_revcomp, and quality filtered
773 using usearch -fastq_filter with -fastq_maxee 1.0. Subsequent processing was like that
774 for the V1-V3 amplicons.

775 16S rRNA gene V4 forward reads (reverse reads in relation the 16S rRNA gene) were
776 trimmed with cutadapt v.2.8 [83] based on the V4 primers with the -g
777 ^GGACTACHVGGGTWTCTAAT...TTACCGCGGCKGCTGGCAC and --discard-
778 untrimmed options. The trimmed reads, which span the entire V4 amplicon, were
779 reverse complemented with usearch -fastx_revcomp, and quality filtered using usearch
780 -fastq_filter with -fastq_maxee 1.0. Subsequent processing was like that for the V1-V3
781 amplicons.
782
783 **In silico *primer evaluation***
784 The specificity of commonly used amplicon primers was determined for each FL-ASV
785 using the analyze_primers.py script from Primer Prospector v. 1.0.1 [84]. The specificity
786 of primer sets was defined based on the overall weighted scores (OWS) for the primer
787 with the highest score as follows: Perfect hit (OWS = 0), partial hit (OWS > 0, and ≤
788 1), poor hit (OWS > 1). The percentage of perfect hits were calculated in R for all taxa
789 in MiDAS 5.
790
791 ***Microbial community analyses***
792 Short-read amplicon data was analyzed with R v.4.2.0 [85] through RStudio IDE [86], with
793 the tidyverse v.1.3.1 (https://www.tidyverse.org/), vegan v.2.6-4 [87], maps v.3.4.0 [88],
794 data.table v.1.14.4 [89], FSA v.0.9.3 [90], rcompanion v. 2.4.18 [91], patchwork v.1.1.1 [92],
795 ggupset v.0.3.0 [93] and Ampvis2 v.2.7.34 [94] packages.
796
797 The microbial community analyses were performed based on all three 16S rRNA gene
798 short-read amplicon dataset (V1-V3, V3-V5, and V4). Samples with less than 10,000
799 reads and those lacking information about digester technology, primary substrate, and
800 temperature in the digester were discarded from the analyses. After filtration, 365 V1-
801 V3, 200 V3-V5, and 269 V4 samples remained.

802 Associations between the AD microbiota and the following process-related or
803 environmental variables were investigated: Digester technology, primary substrate,
804 temperature in the digester, and continent (see definitions above). All variables were
805 treated as factors.

806 For alpha diversity analyses, samples were rarefied to 10,000 reads, and alpha diversity
807 (observed ASVs and inverse Simpson) was calculated using the ampvis2 package. The
808 Kruskal-Wallis with Dunn's post-hoc test (Bonferroni correction with α=0.01 before
809 correction) was used to determine statistically significant differences in alpha diversity
810 between samples grouped by process and environmental variables.

811 Beta diversity distances based on Bray-Curtis (abundance-based) for genera was
812 calculated using the vegdist function in the vegan R package and visualized by PCoA
813 plots with the ampvis2 package. To determine how much individual parameters
814 affected the structure of the microbial community across the ADs, a permutational

815 multivariate analysis of variance (PERMANOVA) test was performed on the beta-
816 diversity matrices using the adonis function in the vegan package with 999
817 permutations.

818 Core taxa (genera and species) were determined separately for ADs treating different
819 primary substrates and operating at different temperatures (mesophilic and
820 thermophilic) based on their relative abundances in individual ADs according to the
821 three short-read amplicon datasets. Core taxa definitions were identical to those applied
822 in the MiDAS global survey of WWTPs [18]. Taxa were considered abundant when
823 present at >0.1% relative read abundance in individual ADs. Based on how frequently
824 taxa were observed to be abundant, we defined the following core communities: loose
825 core (>20% of ADs), general core (>50% of ADs), and strict core (>80% of ADs).
826 Additionally, we defined conditionally rare or abundant taxa (CRAT) [42] composed of
827 taxa present in one or more ADs at >1% relative abundance, but not belonging to the
828 core taxa.
829
830 **Data Availability:**
831 The raw and assembled sequencing data generated in this study have been deposited in
832 the NCBI SRA database and will be made available when the final article is accepted
833 for peer-reviewed publication. The MiDAS 5 reference database in SINTAX, QIIME
834 and Dada2 format is available at the MiDAS fieldguide website
835 [https://www.midasfieldguide.org/guide/downloads].
836
837 **Code Availability:**
838 R scripts and raw data used for data analyses and figures will be made available at
839 GitHub and Figshare when the final article is accepted for peer-reviewed publication.

References:

1. Tiwary, A., Williams, I. D., Pant, D. C. & Kishore, V. V. N. Emerging perspectives on environmental burden minimisation initiatives from anaerobic digestion technologies for community scale biomass valorisation. *Renewable and Sustainable Energy Reviews* **42**, 883–901 (2015).
2. Achinas, S., Achinas, V. & Euverink, G. J. W. A technological overview of biogas production from biowaste. *Engineering* **3**, 299–307 (2017).
3. Samoraj, M. *et al.* The challenges and perspectives for anaerobic digestion of animal waste and fertilizer application of the digestate. *Chemosphere* **295**, 133799 (2022).
4. Czekała, W., Jasiński, T., Grzelak, M., Witaszek, K. & Dach, J. Biogas plant operation: Digestate as the valuable product. *Energies* **15**, 8275 (2022).
5. Briones, A. & Raskin, L. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Current Opinion in Biotechnology* **14**, 270–276 (2003).
6. Dueholm, M. S. *et al.* Generation of comprehensive ecosystem-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 (2020).
7. Jiang, C. *et al.* Characterizing the growing microorganisms at species level in 46 anaerobic digesters at Danish wastewater treatment plants: A six-year survey on microbial community structure and key drivers. *Water Research* **193**, 116871 (2021).
8. Vanwonterghem, I. *et al.* Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME Journal* **8**, 2015–2028 (2014).
9. Ofiţeru, I. D. *et al.* Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences* **107**, 15345–15350 (2010).
10. Kirkegaard, R. H. *et al.* The impact of immigration on microbial community composition in full-scale anaerobic digesters. *Scientific Reports* **7**, 9343 (2017).
11. Werner, J. J. *et al.* Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4158–4163 (2011).
12. Campanaro, S. *et al.* Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. *Biotechnology for biofuels* **9**, 26 (2016).
13. Calusinska, M. *et al.* A year of monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but different core microbiomes in bio-waste and wastewater anaerobic digestion systems. *Biotechnology for Biofuels* **11**, 196 (2018).

14. Mei, R. *et al.* Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Research* **124**, 77–84 (2017).
15. McIlroy, S. J. *et al.* MiDAS 2.0: An ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. *Database* **2017**, bax016 (2017).
16. Nierychlo, M. *et al.* 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 Research* **182**, 115955 (2020).
17. McIlroy, S. J. *et al.* MiDAS: The field guide to the microbes of activated sludge. *Database : the journal of biological databases and curation* **2015**, bav062 (2015).
18. Dueholm, M. K. D. *et al.* MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. *Nat Commun* **13**, 1908 (2022).
19. Albertsen, M., Karst, S. M., Ziegler, A. S., Kirkegaard, R. H. & Nielsen, P. H. Back to basics - the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PloS one* **10**, e0132783 (2015).
20. Kristensen, J. M., Singleton, C., Clegg, L.-A., Petriglieri, F. & Nielsen, P. H. High diversity and functional potential of undescribed "Acidobacteriota" in Danish wastewater treatment plants. *Front. Microbiol.* **12**, 906 (2021).
21. Nierychlo, M. *et al.* Low global diversity of *Candidatus* Microthrix, a troublesome filamentous organism in full-scale WWTPs. *Front Microbiol* **12**, 690251 (2021).
22. Petriglieri, F. *et al.* "*Candidatus* Dechloromonas phosphoritropha" and "*Ca.* D. phosphorivorans", novel polyphosphate accumulating organisms abundant in wastewater treatment systems. *ISME J* **15**, 3605–3614 (2021).
23. Petriglieri, F. *et al.* Reevaluation of the phylogenetic diversity and global distribution of the genus "*Candidatus* Accumulibacter". *mSystems* **7**, e00016-22 (2022).
24. Singleton, C. M. *et al.* The novel genus, '*Candidatus* Phosphoribacter', previously identified as *Tetrasphaera*, is the dominant polyphosphate accumulating lineage in EBPR wastewater treatment plants worldwide. *ISME J* **16**, 1605–1616 (2022).
25. Rivière, D. *et al.* Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME Journal* **3**, 700–714 (2009).
26. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**, 635–645 (2014).
27. Fujimoto, M., Carey, D. E., Zitomer, D. H. & McNamara, P. J. Syntroph diversity and abundance in anaerobic digestion revealed through a comparative core microbiome approach. *Appl Microbiol Biotechnol* **103**, 6353–6367 (2019).

28. Westerholm, M., Calusinska, M. & Dolfing, J. Syntrophic propionate-oxidizing bacteria in methanogenic systems. *FEMS Microbiology Reviews* **46**, fuab057 (2022).
29. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* **10**, 1–11 (2019).
30. Yuan, C., Lei, J., Cole, J. & Sun, Y. Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* **31**, i35–i43 (2015).
31. Singleton, C. M. *et al.* Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nature Communications* **12**, 2009 (2021).
32. Bahram, M., Anslan, S., Hildebrand, F., Bork, P. & Tedersoo, L. Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environmental Microbiology Reports* **11**, 487–494 (2018).
33. Pausan, M. R. *et al.* Exploring the archaeome: Detection of archaeal signatures in the human body. *Frontiers in Microbiology* **10**, 2796 (2019).
34. Wu, L. *et al.* Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nature Microbiology* **4**, 1183–1195 (2019).
35. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
36. Mei, R., Narihiro, T., Nobu, M. K., Kuroda, K. & Liu, W.-T. Evaluating digestion efficiency in full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial activity. *Scientific Reports* **6**, 34090 (2016).
37. Sun, L., Pope, P. B., Eijsink, V. G. H. & Schnürer, A. Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure. *Microbial Biotechnology* **8**, 815–827 (2015).
38. Karakashev, D., Batstone, D. J. & Angelidaki, I. Influence of environmental conditions on methanogenic compositions in anaerobic biogas reactors. *Applied and Environmental Microbiology* **71**, 331–338 (2005).
39. Levén, L., Eriksson, A. R. B. & Schnürer, A. Effect of process temperature on bacterial and archaeal communities in two methanogenic bioreactors treating organic household waste. *FEMS Microbiology Ecology* **59**, 683–693 (2007).
40. Knights, D. *et al.* Rethinking "enterotypes". *Cell Host & Microbe* **16**, 433–437 (2014).
41. Saunders, A. M., Albertsen, M., Vollertsen, J. & Nielsen, P. H. The activated sludge ecosystem contains a core community of abundant organisms. *The ISME journal* **10**, 11–20 (2016).
42. Dai, T. *et al.* Identifying the key taxonomic categories that characterize microbial community diversity using full-scale classification: a case study of microbial communities in the sediments of Hangzhou Bay. *FEMS Microbiology Ecology* **92**, fiw150 (2016).

962 43. Astudillo-García, C. *et al.* Evaluating the core microbiota in complex communities:
963 A systematic investigation. *Environmental Microbiology* **19**, 1450–1462 (2017).
964 44. Shade, A. *et al.* Conditionally rare taxa disproportionately contribute to temporal
965 changes in microbial diversity. *mBio* **5**, e01371-14 (2014).
966 45. Schorn, S. *et al.* Diverse methylotrophic methanogenic archaea cause high methane
967 emissions from seagrass meadows. *Proceedings of the National Academy of*
968 *Sciences* **119**, e2106628119 (2022).
969 46. Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota
970 revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
971 47. Nobu, M. K., Narihiro, T., Kuroda, K., Mei, R. & Liu, W. T. Chasing the elusive
972 Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing
973 methanogen. *The ISME Journal* **10**, 2478–2487 (2016).
974 48. Jung, M.-Y. *et al.* A hydrophobic ammonia-oxidizing archaeon of the
975 *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment.
976 *Environmental Microbiology Reports* **8**, 983–992 (2016).
977 49. Huang, W.-C. *et al.* Comparative genomic analysis reveals metabolic flexibility of
978 Woesearchaeota. *Nat Commun* **12**, 5281 (2021).
979 50. Morris, B. E. L., Henneberger, R., Huber, H. & Moissl-Eichinger, C. Microbial
980 syntrophy: Interaction for the common good. *FEMS Microbiology Reviews* **37**,
981 384–406 (2013).
982 51. Hao, L. *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed
983 by genome-centric metatranscriptomics. *ISME J* **14**, 906–918 (2020).
984 52. Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy:
985 the paradigm for anaerobic metabolic cooperation. *Annual Review of Microbiology*
986 **66**, 429–452 (2012).
987 53. Westerholm, M., Roos, S. & Schnürer, A. *Syntrophaceticus schinkiigen* nov., sp.
988 nov., an anaerobic, syntrophic acetate-oxidizing bacterium isolated from a
989 mesophilic anaerobic filter. *FEMS Microbiology Letters* **309**, 100–104 (2010).
990 54. Schnürer, A., Müller, B. & Westerholm, M. *Syntrophaceticus*. in *Bergey's Manual*
991 *of Systematics of Archaea and Bacteria* 1–10 (John Wiley & Sons, Ltd, 2018).
992 doi:10.1002/9781118960608.gbm01452.
993 55. McInerney, M. J., Bryant, M. P., Hespell, R. B. & Costerton, J. W. *Syntrophomonas*
994 *wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium.
995 *Applied and Environmental Microbiology* **41**, 1029–1039 (1981).
996 56. Wu, C., Liu, X. & Dong, X. *Syntrophomonas cellicola* sp. nov., a spore-forming
997 syntrophic bacterium isolated from a distilled-spirit-fermenting cellar, and
998 assignment of *Syntrophospora bryantii* to *Syntrophomonas bryantii* comb. nov. *Int*
999 *J Syst Evol Microbiol* **56**, 2331–2335 (2006).
1000 57. Sousa, D. Z., Smidt, H., Alves, M. M. & Stams, A. J. M. *Syntrophomonas zehnderi*
1001 sp. nov., an anaerobe that degrades long-chain fatty acids in co-culture with
1002 *Methanobacterium formicicum*. *Int J Syst Evol Microbiol* **57**, 609–615 (2007).

1003 58. Hatamoto, M., Imachi, H., Fukayo, S., Ohashi, A. & Harada, H. *Syntrophomonas*
1004 *palmitatica* sp. nov., an anaerobic, syntrophic, long-chain fatty-acid-oxidizing
1005 bacterium isolated from methanogenic sludge. *International Journal of Systematic*
1006 *and Evolutionary Microbiology* **57**, 2137–2142 (2007).
1007 59. Niu, L., Song, L., Liu, X. & Dong, X. *Tepidimicrobium xylanilyticum* sp. nov., an
1008 anaerobic xylanolytic bacterium, and emended description of the genus
1009 *Tepidimicrobium*. *International Journal of Systematic and Evolutionary*
1010 *Microbiology* **59**, 2698–2701 (2009).
1011 60. Wang, G., Li, Q., Gao, X. & Wang, X. C. Sawdust-derived biochar much mitigates
1012 VFAs accumulation and improves microbial activities to enhance methane
1013 production in thermophilic anaerobic digestion. *ACS Sustainable Chem. Eng.* **7**,
1014 2141–2150 (2019).
1015 61. Liu, Y., Balkwill, D. L., Aldrich, H. C., Drake, G. R. & Boone, D. R.
1016 Characterization of the anaerobic propionate-degrading syntrophs *Smithella*
1017 *propionica* gen. nov., sp. nov. and *Syntrophobacter wolinii*. *International Journal*
1018 *of Systematic and Evolutionary Microbiology* **49**, 545–556 (1999).
1019 62. de Bok, F. A. M., Stams, A. J. M., Dijkema, C. & Boone, D. R. Pathway of
1020 propionate oxidation by a syntrophic culture of *Smithella propionica* and
1021 *Methanospirillum hungatei*. *Applied and Environmental Microbiology* **67**, 1800–
1022 1804 (2001).
1023 63. Dolfing, J. Syntrophic propionate oxidation via butyrate: a novel window of
1024 opportunity under methanogenic conditions. *Applied and Environmental*
1025 *Microbiology* **79**, 4515–4516 (2013).
1026 64. Embree, M., Nagarajan, H., Movahedi, N., Chitsaz, H. & Zengler, K. Single-cell
1027 genome and metatranscriptome sequencing reveal metabolic interactions of an
1028 alkane-degrading methanogenic community. *ISME J* **8**, 757–767 (2014).
1029 65. Tan, B., Nesbø, C. & Foght, J. Re-analysis of omics data indicates *Smithella* may
1030 degrade alkanes by addition to fumarate under methanogenic conditions. *ISME*
1031 *Journal* **8**, 2353–2356 (2014).
1032 66. Nobu, M. K. *et al.* The genome of *Syntrophorhabdus aromaticivorans* strain UI
1033 provides new insights for syntrophic aromatic compound metabolism and electron
1034 flow. *Environmental Microbiology* **17**, 4861–4872 (2015).
1035 67. Qiu, Y.-L. *et al. Syntrophorhabdus aromaticivorans* gen. nov., sp. nov., the first
1036 cultured anaerobe capable of degrading phenol to acetate in obligate syntrophic
1037 associations with a hydrogenotrophic methanogen. *Appl Environ Microbiol* **74**,
1038 2051–2058 (2008).
1039 68. McInerney, M. J. *et al.* The genome of *Syntrophus aciditrophicus*: Life at the
1040 thermodynamic limit of microbial growth. *Proceedings of the National Academy of*
1041 *Sciences* **104**, 7600–7605 (2007).
1042 69. Jackson, B. E., Bhupathiraju, V. K., Tanner, R. S., Woese, C. R. & McInerney, M.
1043 J. *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty

1044 acids and benzoate in syntrophic association with hydrogen-using microorganisms.
1045 *Arch Microbiol* **171**, 107–114 (1999).
1046 70. Ganidi, N., Tyrrel, S. & Cartmell, E. Anaerobic digestion foaming causes – A
1047 review. *Bioresource Technology* **100**, 5546–5554 (2009).
1048 71. Duan, J.-L. *et al.* Unraveling anaerobic digestion foaming via association between
1049 bacterial metabolism and variations in microbiota. *ACS EST Eng.* **1**, 978–988
1050 (2021).
1051 72. Jiang, C. *et al.* Identification of microorganisms responsible for foam formation in
1052 mesophilic anaerobic digesters treating surplus activated sludge. *Water Research*
1053 **191**, 116779 (2021).
1054 73. Bovio-Winkler, P., Cabezas, A. & Etchebehere, C. Database mining to unravel the
1055 ecology of the phylum Chloroflexi in methanogenic full scale bioreactors. *Frontiers*
1056 *in Microbiology* **11**, 603234 (2021).
1057 74. McIlroy, S. J. *et al.* Culture-independent analyses reveal novel Anaerolineaceae as
1058 abundant primary fermenters in anaerobic digesters treating waste activated sludge.
1059 *Frontiers in Microbiology* **8**, 1134 (2017).
1060 75. Karst, S. M. *et al.* High-accuracy long-read amplicon sequences using unique
1061 molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods* **18**, 165–
1062 169 (2021).
1063 76. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers
1064 for classical and next-generation sequencing-based diversity studies. *Nucleic Acids*
1065 *Research* **41**, e1 (2013).
1066 77. Lane, D. J. 16S/23S rRNA sequencing. in *Nucleic Acid Techniques in Bacterial*
1067 *Systematics* (eds. Stackebrandt, E. & Goodfellow, M.) 115–175 (John Wiley and
1068 Sons, 1991). doi:10.1007/s00227-012-2133-0.
1069 78. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: Assessing
1070 small subunit rRNA primers for marine microbiomes with mock communities, time
1071 series and global field samples. *Environmental Microbiology* **18**, 1403–1414
1072 (2016).
1073 79. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the
1074 Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**, 1621–1624 (2012).
1075 80. Pinto, A. J. & Raskin, L. PCR biases distort bacterial and archaeal community
1076 structure in pyrosequencing datasets. *PLoS One* **7**, e43093 (2012).
1077 81. Apprill, A., Mcnally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU
1078 rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton.
1079 *Aquatic Microbial Ecology* **75**, 129–137 (2015).
1080 82. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
1081 *Bioinformatics* **26**, 2460–2461 (2010).
1082 83. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1083 reads. *EMBnet.journal* **17**, 10–12 (2011).

84. Walters, W. A. *et al.* PrimerProspector: De novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27**, 1159–1161 (2011).
85. R Development Core Team, R. F. F. S. C. & R Core Team. *R: A Language and Environment for Statistical Computing*. *Vienna Austria R Foundation for Statistical Computing* vol. 1 (R Foundation for Statistical Computing, Vienna, 2008).
86. RStudio Team. *RStudio: Integrated Development Environment for R*. (RStudio, PBC., 2020).
87. Oksanen, J. *et al. vegan: Community Ecology Package*. (2019).
88. Richard A. Becker, Allan R. Wilks, Ray Brownrigg, Thomas P Minka, & Alex Deckmyn. maps: Draw Geographical Maps. (2021).
89. Dowle, M. & Srinivasan, A. data.table: Extension of `data.frame`. (2019).
90. Ogle, D. H., Doll, J. C., Wheeler, A. P. & Dinno, A. FSA: Simple fisheries stock assessment methods. (2023).
91. Mangiafico, S. rcompanion: Functions to support extension education program evaluation. (2023).
92. Pedersen, T. L. Patchwork: the composer of plots. (2020).
93. Constantin Ahlmann-Eltze. ggupset: Combination matrix axis for 'ggplot2' to create 'UpSet' plots. (2020).
94. Andersen, K. S. S., Kirkegaard, R. H., Karst, S. M. & Albertsen, M. Ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* 299537 (2018) doi:10.1101/299537.

Acknowledgements:
The project has been funded by the Danish Research Council (grant 6111-00617A, P.H.N.) and the Villum Foundation (Dark Matter and grant 13351, P.H.N.). We thank all the involved anaerobic digester plants for providing samples and plant metadata.

Author Contributions:
P.H.N. and M.K.D.D. designed the study. M.K.D.D. and P.H.N. wrote the manuscript and all authors reviewed and approved the final manuscript. The MiDAS Global Consortium for Anaerobic Digesters provided samples and metadata. V.R. handled sampling, DNA extraction and library preparation for DNA sequencing. K.S.A. and M.K.D.D. performed the bioinformatics analyses. M.K.D.D., A-K.C.P. and K.S.A. curated metadata and carried out statistical analyses.

MiDAS Global Consortium for Anaerobic Digesters:
Madalena Alves[2], Yadira Bajón-Fernández[3], Damien Batstone[4], Caitlyn Butler[5], Mercedes Cecilia Cruz[6], Åsa Davidsson[7], Leonardo Erijman[8], Christof Holliger[9], Konrad Koch[10], Norbert Kreuzinger[11], Changsoo Lee[12], Gerasimos Lyberatos[13], Srikanth Mutnuri[14], Vincent O'Flaherty[15], Piotr Oleskowicz-Popiel[16], Dana Pokorna[17], Veronica Rajal[6], Michael Recktenwald[18], Jorge Rodríguez[19], Pascal E. Saikaly[20], Nick Tooker[5], Julia Vierheilig[11], Jo De Vrieze[21], Christian Wurzbacher[10a10] & Per Halkjær Nielsen[1]

[2]Centre of Biological Engineering, University of Minho, Portugal. [3]School of Water, Energy and Environment, Cranfield University. Cranfield, United Kingdom. [4]Australian Centre for Water and Environmental Biotechnology (ACWEB), The University of Queensland, Australia. [5]Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA, USA. [6]National University of Salta, Salta, Argentina. [7]Department of Chemical Engineering, Lund University, Sweden. [8]INGEBI-CONICET, University of Buenos Aires, Argentina. [9]Laboratory for Environmental Biotechnology, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. [10]Chair of Urban Water Systems Engineering, Technical University of Munich (TUM), Garching, Germany. [11]Institute of Water Quality and Resource Management, Vienna University of Technology, Austria. [12]Department of Urban and Environmental Engineering & Graduate School of Carbon Neutrality, Ulsan National Institute of Science and Technology (UNIST), South Korea. [13]School of Chemical Engineering, National Technical University of Athens, Greece. [14]Applied Environmental Biotechnology Laboratory, Birla Institute of Technology and Science (BITS-Pilani), India. [15]School of Biological and Chemical Sciences and Ryan Institute, University of Galway, Ireland. [16]Water Supply and Bioeconomy Division, Faculty of Environmental Engineering and Energy, Poznan University of Technology, Poland. [17]Department of Water Technology and Environmental Engineering, University of Chemistry and Technology Prague, Czech Republic. [18]Research Scientist at Kemira

Oyj, Espoo R&D Center, Finland. [19]Chemical Engineering Department, Khalifa University, United Arab Emirates. [20]Environmental Science and Engineering Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia. [21]Center for Microbial Ecology and Technology (CMET), Ghent University, Belgium.

Competing Interests:

The authors declare no competing interests.