Database update

MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups

Simon Jon McIlroy, Rasmus Hansen Kirkegaard, Bianca McIlroy, Marta Nierychlo, Jannie Munk Kristensen, Søren Michael Karst, Mads Albertsen and Per Halkjær Nielsen*

Department of Chemistry and Bioscience, Center for Microbial Communities, Aalborg University, Aalborg DK-9220, Denmark

***Corresponding author:** Tel: +45 99408503; Email: phn@bio.aau.dk

Abstract

Wastewater is increasingly viewed as a resource, with anaerobic digester technology being routinely implemented for biogas production. Characterising the microbial communities involved in wastewater treatment facilities and their anaerobic digesters is considered key to their optimal design and operation. Amplicon sequencing of the 16S rRNA gene allows high-throughput monitoring of these systems. The MiDAS field guide is a public resource providing amplicon sequencing protocols and an ecosystem-specific taxonomic database optimized for use with wastewater treatment facility samples. The curated taxonomy endeavours to provide a genus-level-classification for abundant phylotypes and the online field guide links this identity to published information regarding their ecology, function and distribution. This article describes the expansion of the database resources to cover the organisms of the anaerobic digester systems fed primary sludge and surplus activated sludge. The updated database includes descriptions of the abundant genus-level-taxa in influent wastewater, activated sludge and anaerobic digesters. Abundance information is also included to allow assessment of the role of emigration in the ecology of each phylotype. MiDAS is intended as a collaborative resource for the progression of research into the ecology of wastewater treatment, by providing a public repository for knowledge that is accessible to all interested in these biotechnologically important systems.

Database URL: http://www.midasfieldguide.org

Introduction

Wastewater treatment is one of the largest biotechnological industries in the world. Wastewater itself is increasingly viewed as a resource with a growing focus on the recovery of inorganic resources, such as nitrogen and phosphorus, and bioenergy. As such, anaerobic digester technology is increasingly being implemented at full-scale treatment plants, primarily for the conversion of solid waste to methane gas for energy, giving the potential for net-energy production from wastewater treatment (1). As with activated sludge systems, anaerobic digestion is mediated by complex microbial communities, with the function, stability and efficiency reliant on the tightly coupled synergistic activities of specialized microbial sub-populations. As such, an understanding of the ecology of these systems will be important for optimal process design and operation (2,3).

Recent advances in sequencing technology now make high throughput analyses of complex microbial communities with 16S rRNA gene sequencing possible. This allows high-resolution monitoring over time, permitting correlation analyses with operational parameters for the identification of key phylotypes in full-scale wastewater treatment and biogas systems (3,4). Of subsequent importance to our understanding of these systems is the ability to estimate the function of these phylotypes, which becomes unreliable for classification at taxonomic levels higher than the species or genus (5). Classification of 16S rRNA gene sequences is mediated by curated public databases: including Silva, the Ribosomal Databases Project (RDP) and Greengenes (6–8). However, a large proportion of communities in wastewater treatment and biogas systems are uncultured and unclassified in these databases (9–14). The onerous task of manual taxonomic annotation has led to the development of ecosystem-specific taxonomies where annotation efforts are focussed on abundant and pertinent groups (15–18)—including the MiDAS database for the organisms of activated sludge (4). The MiDAS initiative initially provided a taxonomic database curated for abundant and process important phylotypes for activated sludge wastewater treatment systems with biological nutrient removal. It also incorporated a referenced online database that allows phylotype identity to be linked to information about the morphology, ecophysiology, abundance and distribution of genus members in full-scale treatment systems. The project was made feasible by the observation that relatively few genera make up the majority of organisms in full-scale wastewater treatment systems (12). Surveys of anaerobic digesters located at wastewater treatment plants reveal that these microbial communities also have a common set of abundant genera (9,13).

In this report, we present the expansion of the MiDAS field guide to include the organisms of the anaerobic digestion community. This includes an updated taxonomy and online profiles for the abundant organisms of full-scale anaerobic digester systems. In addition, the most abundant influent wastewater organisms are also incorporated, allowing for an assessment of the influence of migration on described phylotypes, which is critical to understanding their ecology. Coverage of the influent wastewater, activated sludge and anaerobic digester communities gives a holistic view of the microbial ecology of wastewater treatment facilities.

Curation and expansion of the MiDAS taxonomy

The MiDAS taxonomy is a curated version of the SILVA taxonomy (7), with MiDAS release 2.1 based on SILVA release 1.23. As with the previous MiDAS release (4), the annotation of novel sequence clades was guided by the position of representative OTU sequences, added to the SILVA base tree with the 'ARB parsimony insertion tool,' and their closest full-length sequences (percentage sequence similarity). These amplicons represent the abundant organisms from large-scale surveys of influent wastewater, activated sludge and anaerobic digesters fed with primary sludge and surplus activated sludge (summarized in Table 1). There are further plans to incorporate anaerobic digester systems treating other organics, such as food and industrial wastes. While the current MiDAS release is based solely on lists from Danish treatment systems, it will be updated to be relevant to plants globally in the near future. Preliminary data for activated sludge shows that the same abundant organisms appear to be found in systems globally, making the current database version already relevant for use with systems outside Denmark (Nierychlo, M., Nielsen, P.H. and others, unpublished).

The datasets used for the analyses presented in Table 2 and Figures 1–3 include influent wastewater (14 WWTPs, 3 months), activated sludge (24 WWTPs, 5 years), mesophilic ADs (26 reactors at 14 WWTPs, 6 years) and thermophilic ADs (7 reactors at 5 WWTPs, 6 years), and are taken from a recent 16S rRNA gene amplicon sequencing survey study of Danish WWTPs—the reader is referred to this article for further details (19). The benefit of applying the MiDAS taxonomy is evident in Table 2, which shows that a substantial proportion of the sequences present in the analysed systems are classified to novel MiDAS genus-level taxa. Several of the most abundant genera for the analysed sample types include novel MiDAS taxa (Figures 1–3)—such as T78, KD1-22 and *Candidatus* Fermentibacteria for the mesophilic ADs (Figure 2A). Importantly, these novel taxa would currently not be identified with other publically available taxonomies. The improvement for the different sample types varies, but on average it is > 25% for the Bacteria of anaerobic digesters. Limited improvement is shown for the influent genera,

Table 1. The sequence data sets used to guide the curation of the MiDAS v. 2.1 taxonomy

System	Number of WWTPs	Sampling period	Sampling frequency (per year)	16S amplicon region	Kingdom	No. OTUs used	Reference
Wastewater influent	14	3 months	6	V1-3	Bacteria	50	Kirkegaard *et al.* (19)
Activated sludge	13	2 years	2–6	V4	Bacteria	104[a]	Saunders *et al.* (12)
	20	8 years	4	V1-3	Bacteria	100	McIlroy *et al.* (4)
AD—mesophilic	14 (26 ADs)	6 years	≤4	V1-3	Bacteria	40	Kirkegaard *et al.* (19)
				V3-5	Archaea	20	Kirkegaard *et al*, (19)
AD—thermophilic	5 (7 ADs)	6 years	≤4	V1-3	Bacteria	40	Kirkegaard *et al.* (19)
				V3-5	Archaea	20	Kirkegaard *et al.* (19)
					Total	374	

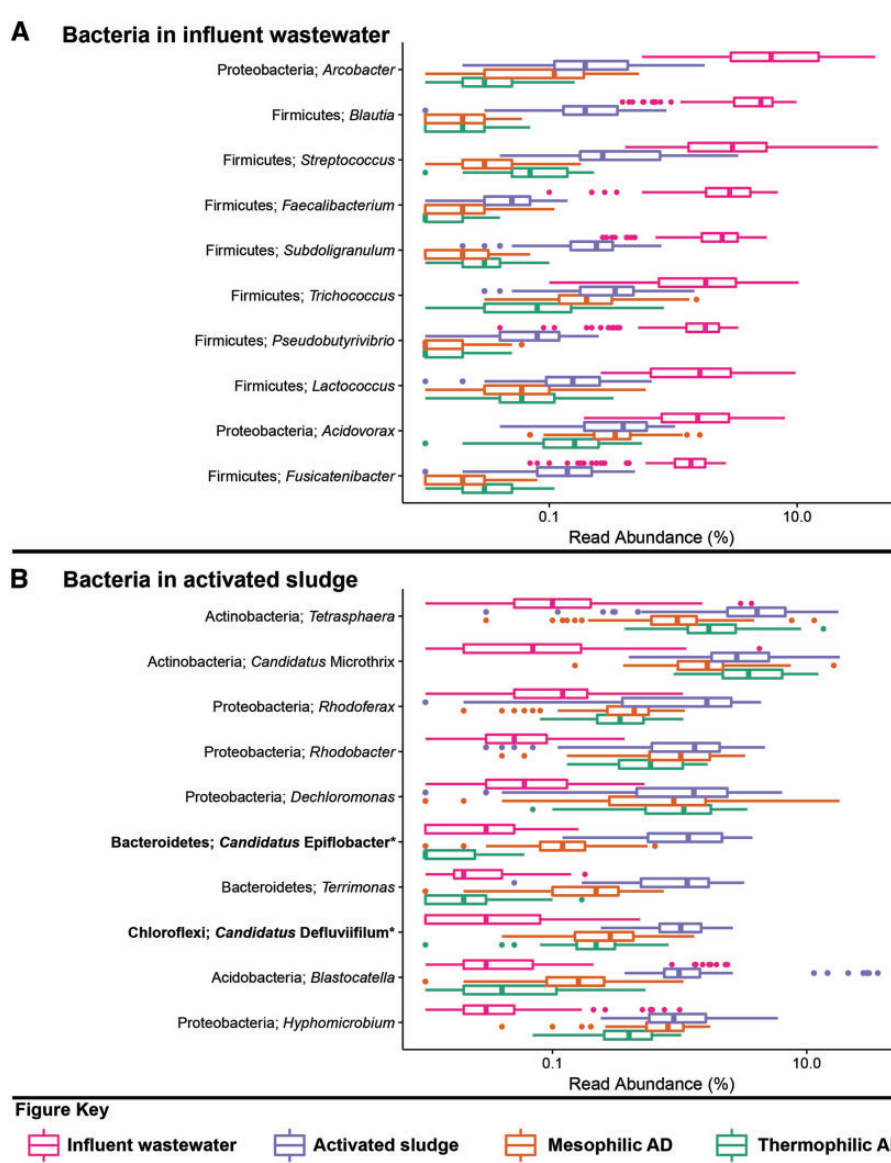[a]OTUs 94% sequence similarity cut-off value. All others use a cut-off of 97%.

Figure 1. Box plot for the distribution of the 10 most abundant bacterial genus-level-taxa by median abundance in **A**. influent wastewater and **B**. activated sludge. Amplicon abundance values (V1-3 region) are given as a percentage of total bacterial reads for the influent wastewater, activated sludge and anaerobic digester communities (19). *Novel MiDAS genus level taxa are given in bold.
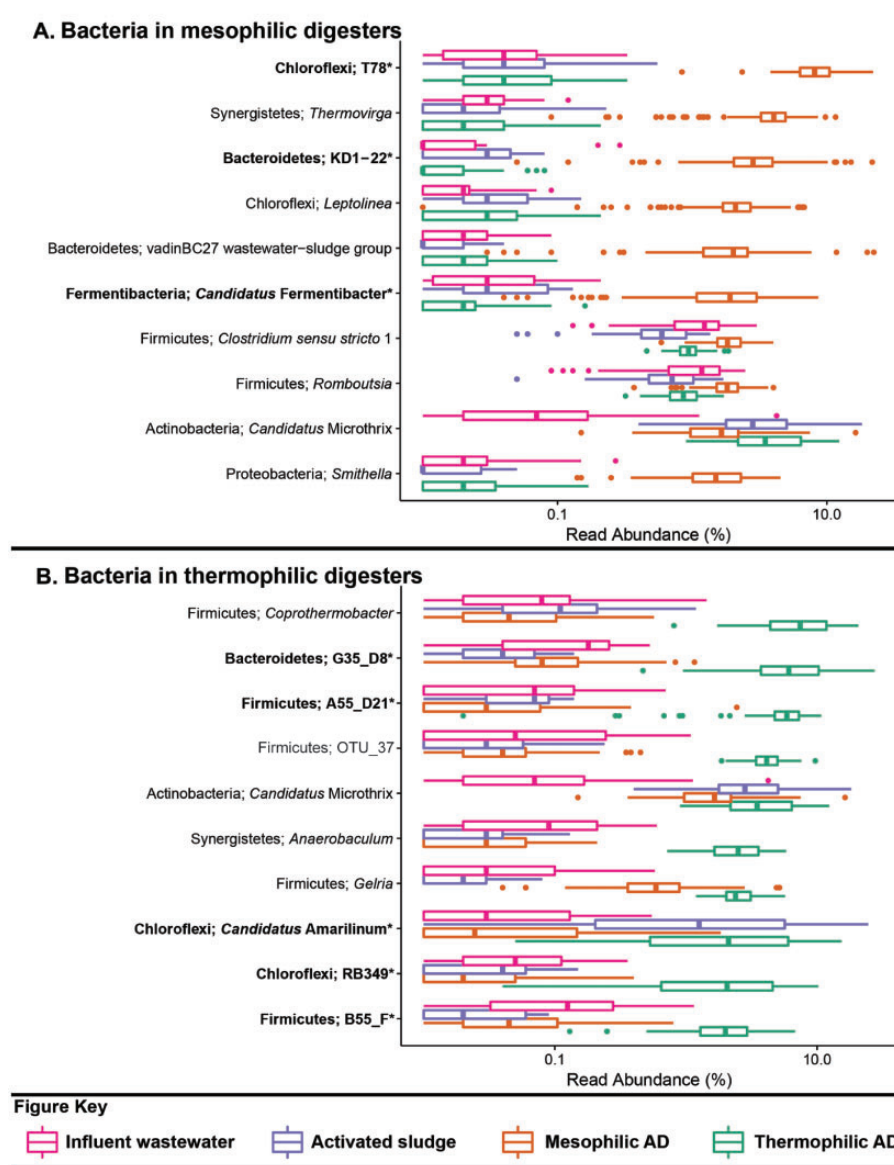
Figure 2. Box plot for the distribution of the 10 most abundant bacterial genus-level-taxa by median abundance in **A**. mesophilic and **B**. thermophilic anaerobic digesters. Amplicon abundance values (V1-3 region) are given as a percentage of total bacterial reads for the influent wastewater, activated sludge and anaerobic digester communities (19). *Novel MiDAS genus level taxa are given in bold.

given most are associated with well-characterized faecal organism groups (Figure 1A; Table 2).

Despite manual targeted annotation, not all of the abundant OTUs are assigned to a genus (Figures 2B and 3; Table 2). A lack of genus level annotation is partly due to the absence of closely related reference sequences in the database and the inadequate resolution of the short amplicon sequence for some closely related genera. This is particularly evident for placement of the archaeal sequences (see Figure 3), likely due to their relatively short length (variable region V3-5; 275 bps). The recently described method for high-throughput sequencing of the full rRNA SSU gene will improve the under-populated and environmentally skewed databases, resulting in improved amplicon classification (20).

The ability to generate comprehensive datasets of full-length sequences, for a given environment, allows for the possibility for ecosystem-specific reference databases – which may also permit a higher resolution for amplicon classification.

Linking taxonomy with function

The searchable online MiDAS database allows users to link genus level names, obtained with the use of the MiDAS taxonomy, to available information on their distribution and likely function in wastewater systems. Online MiDAS profiles are provided for the abundant genera of the influent wastewater, activated sludge and associated anaerobic digesters, based on extensive in-house surveys (Table 1; Figure 4). The
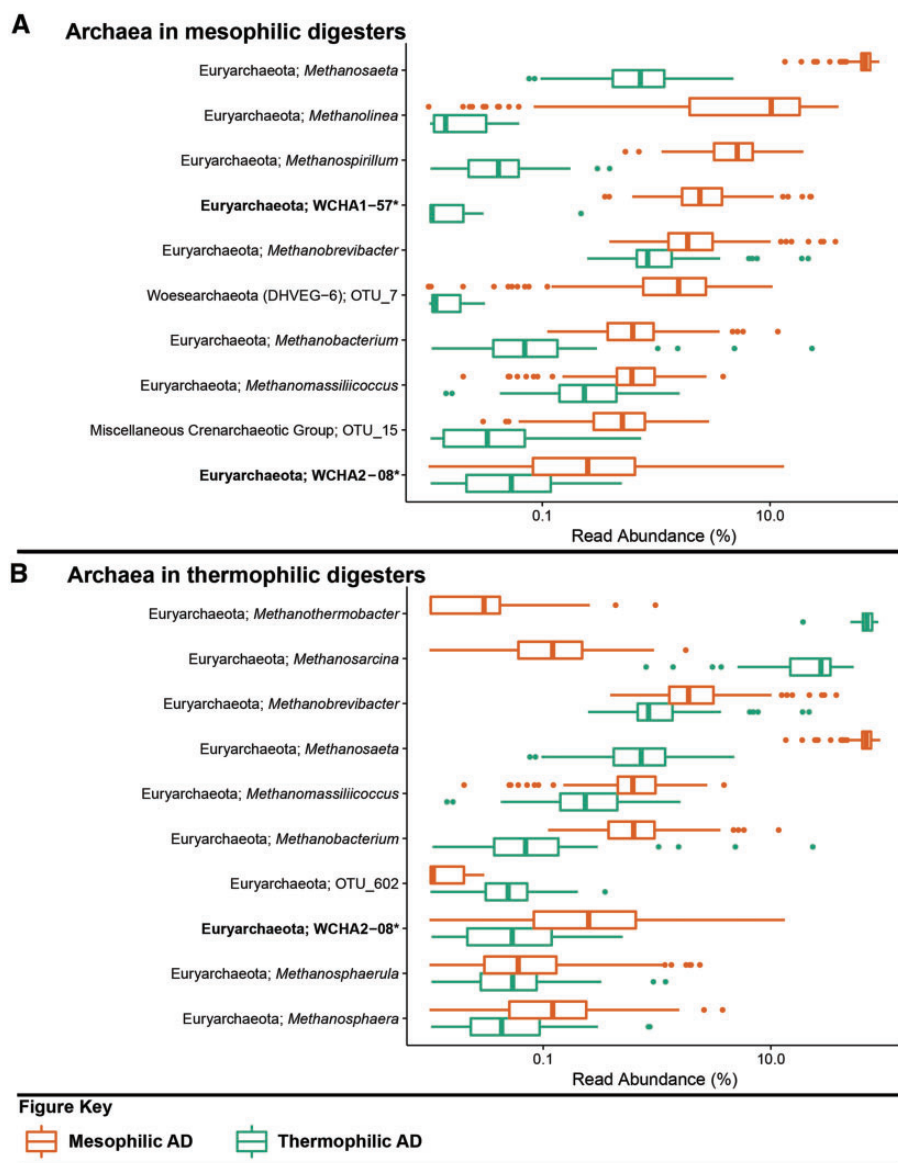
Figure 3. Box plot for the distribution of the 10 most abundant archaeal genus-level-taxa by median abundance in **A**. mesophilic and **B**. thermophilic anaerobic digesters. Amplicon abundance values (V3-5 region) are given as a percentage of total archaeal reads for each phylotype (19). *Novel MiDAS genus level taxa are given in bold.

Table 2. Classification of wastewater treatment related systems with the MiDAS taxonomy

Environment	Kingdom	Amplicons classified to genus level[a]			
		Novel MiDAS taxa[b]		Total classified	
		% top 50 OTUs	% total reads	% top 50 OTUs	% total reads
Influent wastewater	Bacteria	0.0	1.2	94.0	89.0
Activated sludge	Bacteria	30.0	24.4	88.0	79.3
Mesophilic AD	Bacteria	38.0	26.6	84.0	79.6
	Archaea	12.0	4.3	88.0	87.0
Thermophilic AD	Bacteria	32.0	28.6	86.0	80.9
	Archaea	4.0	< 1	84.0	99.8

Sequences for all datasets (19) are classified with the MiDAS taxonomy v. 2.1.

[a]Sequences from all plants.

[b]Genus level taxa not included in the base SILVA 123 taxonomy.

Candidatus Fermentibacter

Genus name	Candidatus Fermentibacter
Source	
Alternative names	B-1AC (MiDAS 1.21)

Classification

Domain	Bacteria
Phylum	Fermentibacteria
Class	Fermentibacteria
Order	Fermentibacterales
Family	Fermentibacteriaceae
Genus	Candidatus Fermentibacter

16S gene copy number	1
# Genomes	Yes[1]
Source	Published

	In situ	Other
Filamentous	●[1]	○
Hydrophobic Cell Surface	○	○
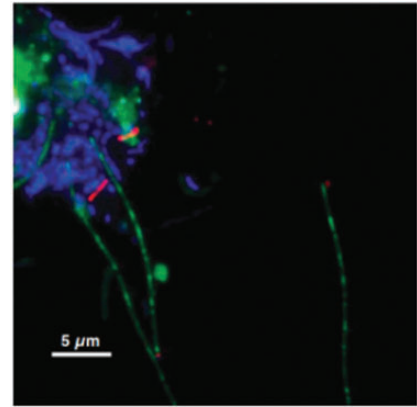
Composite FISH micrographs of the "Ca. Fermentibacter spp." in the Randers anaerobic digester sludge."Ca. Fermentibacter spp." cells appear red (Hyd24-12_659, Cy3), other bacterial cells appear blue (EUBmix probe set, Cy5) and archaeal cells appear green (ARCH915, FLUOS). - Source:[1]

● Fermentation
● Sugars

●POS ●NEG ●Variable ●Not assessed

Description

Based on annotation of representative genomes, derived from full-scale mesophilic anaerobic digesters, it appears that members have a primarily fermentative metabolism. They appear to be strict anaerobes with no TCA cycle and possibly utilise elemental sulfur as an electron acceptor. Cells are small rods typically distributed throughout the biomass[1].

FISH probes

Hyd24-12_658; Hyd24-12_468[1]

Metabolism

	In situ	Other
Autotroph/mixotroph	○	●[1]
AOB	○	●[1]
NOB	○	●[1]
Anammox	○	●[1]
Aerobic Heterotroph	○	●[1]
PAO	○	●[1]
GAO	○	●[1]
Nitrite Reduction	○	●[1]
Sulphate Reduction	○	●[1]
Fermentation	○	●[1]
Acetogen	○	●[1]
Methanogen	○	●[1]
Fatty Acids	○	●[1]
Sugars	○	●[1]
Proteins/amino Acids	○	○

Abundance Information

	10 % percentile	Median	90 % percentile
Influent	0	0	0
Activated Sludge	0	0	0
Digester-Mesophilic	0	1.1	3.5
Digester-Thermophilic	0	0	0

References

[1] Kirkegaard, Dueholm, McIlroy, Nierychlo, Karst, Albertsen, et al. - Genomic insights into members of the candidate phylum Hyd24-12 common in mesophilic anaerobic digesters. - ISME J 10(10): 2352-64. doi:10.1038/ismej.2016.43

Figure 4. An example profile for the online MiDAS database. For a detailed explanation of each field see Supplementary Table S1.
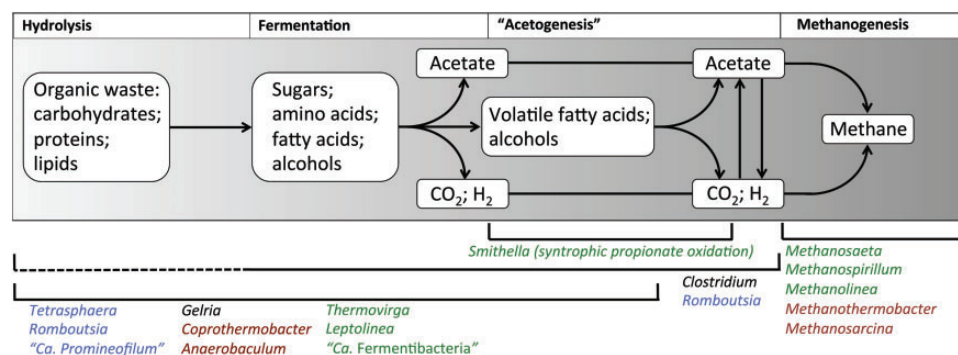
Figure 5. Basic outline of the carbon flow for the anaerobic digestion process. A selection of organisms found to be abundant in Danish systems that are associated with these process steps are listed for each (indicated by brackets, broken lines indicate uncertainty about the pathway in these organisms). Font colour indicates their niche system (green = mesophilic ADs; burgundy = thermophilic ADs; black = both mesophilic and thermophilic ADs; blue = abundant in both AS and recipient ADs).

current MiDAS release includes detailed profiles for 217 bacterial and 12 archaeal genera. These profiles are routinely updated and users are encouraged to contact us regarding relevant new information (mail@midasfieldguide.org).

Several fields relevant for the activated sludge system have been retained in the updated online database and key functional guilds for the anaerobic digester environment have been added. A full description of these fields is given in Supplementary Table S1. The key microbial mediated processes in the conversion of complex organics into methane include hydrolysis, fermentation, acetogenesis and methanogenesis (Figure 5). All except hydrolysis are covered with dedicated fields. Special roles in hydrolysis are noted in the organism *Description* fields; otherwise it is assumed that the fermentative organisms mediate the breakdown of complex organics. The *Acetogenesis* field in MiDAS uses the microbiological definition of the trait. Acetogenesis is routinely applied to broadly describe the synthesis of acetate. However, this definition is somewhat misleading as several metabolic strategies result in acetate production (21). The microbiological definition of acetogenesis is the conversion of CO_2 and H_2 to acetyl-CoA, which is utilized for biomass production or converted to acetate, via the reductive acetyl-CoA pathway (21).

Fields identifying problematic groups for the anaerobic digestion process are also included. Sulphate reducers are identified due to their unfavourable association with bio-corrosion of metal surfaces, odour production and competition with methanogens, resulting in reduced methane yield (22). In addition, hydrophobicity assessment has been included to identify organisms with a possible role in the stabilisation of foams. Foam formation is a common operational problem in both activated sludge systems and anaerobic digesters (23,24). Other pertinent information to the ecology and role of the MiDAS organisms is covered in the *Description* text field.

In addition to accessing individual organism profiles, database searches can also be filtered by each of the profile fields, e.g. users can search for all genera with fermentative

or filamentous members. Organism lists can also be sorted by these fields, e.g. by abundance in a particular system, and can also be exported in tabular format (Figure 6).

Distribution and the importance of migration

An important expansion of MiDAS is the inclusion of abundance values of selected organisms for influent wastewater, activated sludge and anaerobic digesters. Viewing the abundance values in isolation ignores the important influence of migration (12). The population composition of the activated sludge and the anaerobic digesters is substantially influenced by the high migration with the primary sludge from the influent and secondary sludge [surplus activated sludge; Figures 1 and 2 (12,19,25)]. Focussing on the abundant members of the mesophilic sludge, the T78, *Thermovirga*, KD1-22, vadinBC27, *Leptolinea*, *Smithella* and 'Candidatus Fermentibacter' genera are all clearly selected for, while *Clostridium*, *Romboutsia* and 'Candidatus Microthrix' make up a substantial portion of the activated sludge fed into the system (Figure 2) and efforts to determine the activity of the latter are required. Migrating species may also be associated with operational problems, irrespective of metabolic activity, i.e. potential for stabilisation of foams for *Gordonia* spp. and 'Candidatus Microthrix spp.' (23).

Concluding remarks

Recent advances in sequencing technology for the first time allow rapid and high throughput analyses of the composition of the microbial communities of wastewater treatment systems. However, there are a number of impediments to our progress in achieving a holistic understanding of the ecology of wastewater systems, which the MiDAS initiative endeavours to address. These include the inconsistent use of taxonomy, DNA extraction and primer selection, all of which
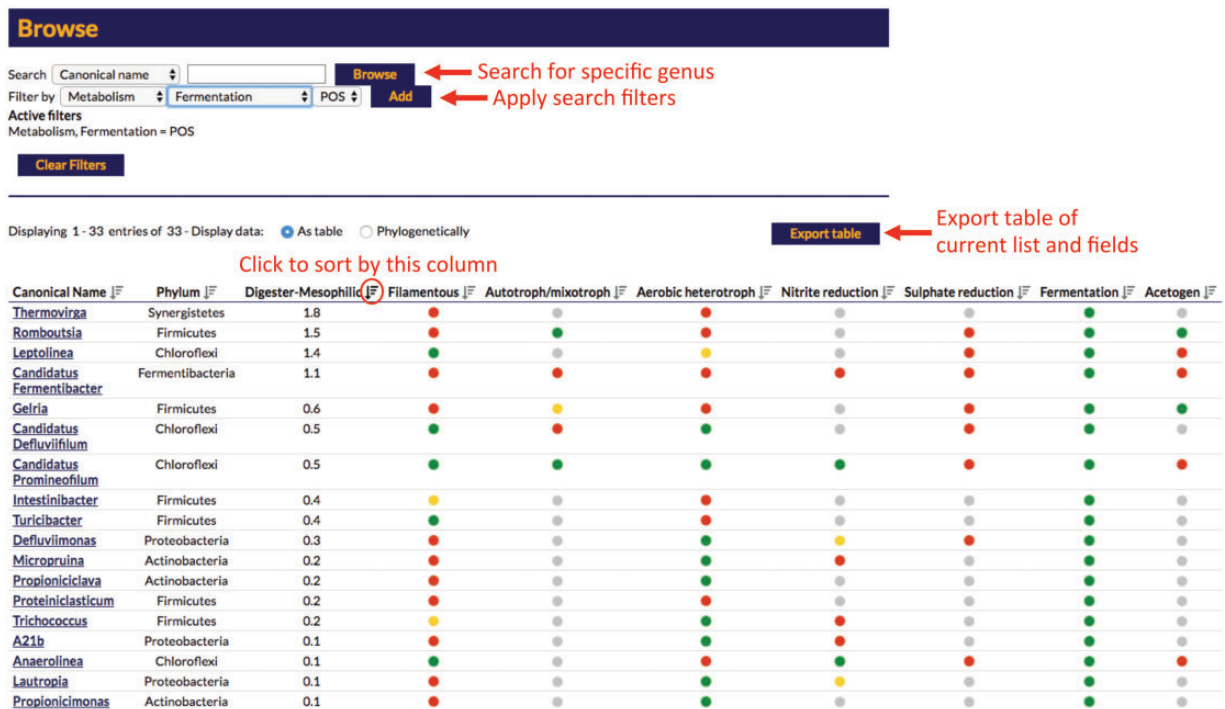
Figure 6. Screenshot of the result displayed from a search of the online MiDAS database. The displayed example is applying a search filter for fermentative organisms sorted by median read % abundance in mesophilic anaerobic digesters. Select features are highlighted in red.

have a substantial impact on the resulting analyses and make cross study comparison difficult (14,26,27). The MiDAS initiative consequently encourages the use of more consistent workflows by providing a public resource, which includes protocols and a taxonomy that are tailored to the wastewater treatment environment. An important subsequent point is the value of consistent genus-level-identifiers. These names enable the collation of relevant information for important phylotypes, which is also facilitated by the online MiDAS field guide. The ambition of MiDAS is to become a collaborative resource for those working in research and to facilitate the accumulation of knowledge made accessible to all with an interest in the biotechnological field of wastewater treatment and bioenergy.

Supplementary data

Supplementary data are available at *Database* Online.

Funding

References

1. McCarty,P.L., Bae,J. and Kim,J. (2011) Domestic wastewater treatment as a net energy producer–can this be achieved? *Environ. Sci. Technol.*, 45, 7100–7106.
2. Read,S., Marzorati,M., Guimarães,B.C.M. *et al.* (2011) Microbial Resource Management revisited: successful parameters and new concepts. *Appl. Microbiol. Biotechnol.*, 90, 861–871.
3. Vanwonterghem,I., Jensen,P.D., Ho,D.P. *et al.* (2014) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.*, 27, 55–64.
4. McIlroy,S.J., Saunders,A.M., Albertsen,M. *et al.* (2015) MiDAS: the field guide to the microbes of activated sludge. *Database*, 2015, bav062.
5. Martiny,A.C., Treseder,K. and Pusch,G. (2013) Phylogenetic conservatism of functional traits in microorganisms. *Isme J.*, 7, 830–838.
6. McDonald,D., Price,M.N., Goodrich,J. *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme J.*, 6, 610–618.
7. Quast,C., Pruesse,E., Yilmaz,P. *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, 41, D590–D596.
8. Cole,J.R., Wang,Q., Fish,J.A. *et al.* (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.*, 42, D633–D642.
9. Riviere,D., Desvignes,V., Pelletier,E. *et al.* (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *Isme J.*, 3, 700–714.

10. Nelson,M.C., Morrison,M. and Yu,Z. (2011) A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresour. Technol.*, 102, 3730–3739.

11. Zhang,T., Shao,M.F. and Ye,L. (2012) 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *Isme J.*, 6, 1137–1147.

12. Saunders,A.M., Albertsen,M., Vollertsen,J. *et al.* (2016) The activated sludge ecosystem contains a core community of abundant organisms. *Isme J.*, 10, 11–20.

13. Kirkegaard,R.H., Dueholm,M.S., McIlroy,S.J. *et al.* (2016) Genomic insights into members of the Candidate phylum Hyd24-12 common in mesophilic anaerobic digesters. *Isme J.*, 10, 2352–2364.

14. Werner,J.J., Koren,O., Hugenholtz,P. *et al.* (2012) Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *Isme J.*, 6, 94–103.

15. Chen,T., Yu,W.H., Izard,J. *et al.* (2010) The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database*, 2010, baq013.

16. Dewhirst,F.E., Klein,E.A., Thompson,E.C. *et al.* (2012) The canine oral microbiome. *PLoS One*, 7, e36067.

17. Dewhirst,F.E., Klein,E.A., Bennett,M.L. *et al.* (2015) The feline oral microbiome: a provisional 16S rRNA gene based taxonomy with full-length reference sequences. *Vet. Microbiol.*, 175, 294–303.

18. Ritari,J., Salojärvi,J., Lahti,L. *et al.* (2015) Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics*, 16, 1056.

19. Kirkegaard,R.H., McIlroy,S.J., Kristensen,J.M. *et al.* (2017) Identifying the abundant and active microorganisms common to full-scale anaerobic digesters. *bioRxiv*, doi: https://doi.org/10.1101/104620.

20. Karst,S.M., Dueholm,M.S., McIlroy,S.J. *et al.* (2016) Thousands of primer-free, high-quality, full-length SSU rRNA sequences from all domains of life. *bioRxiv*, doi: 10.1101/070771.

21. Drake,H.L., Küsel,K. and Matthies,C. (2013) Acetogenic prokaryotes. In: Rosenberg, E., De Long, E.F., Lory,S., Stackebrandt, E., Thomson,F. (ed). *The Prokaryotes: Prokaryotic Physiology and Biochemistry*. Springer-Verlag, Berlin, pp. 3–60.

22. Muyzer,G. and Stams,A.J.M. (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nat. Rev. Microbiol.*, 6, 441.

23. Ganidi,N., Tyrrel,S. and Cartmell,E. (2009) Anaerobic digestion foaming causes: a review. *Bioresour. Technol.*, 100, 5546–5554.

24. Jenkins,D., Richard,M.G. and Daigger,G.T. (2004) *Manual on the Causes and Control of Activated Sludge Bulking, Foaming and Other Solids Separation Problems*, 3rd. edn. CRC Press LLC, London, England, p. 236.

25. Mei,R., Narihiro,T., Nobu,M.K. *et al.* (2016) Evaluating digestion efficiency in full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial activity. *Sci. Rep.*, 6, 34090.

26. Albertsen,M., Karst,S.M., Ziegler,A.S. *et al.* (2015) Back to basics: the influence of DNA extraction and primer choice on phylogenetic analysis in activated sludge communities. *Plos One*, 10, e0132783.

27. Yilmaz,P., Parfrey,L.W., Yarza,P. *et al.* (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.*, 42, D643–D648.