ORIGINAL ARTICLE

Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge

Delphine Rivière[1,2,3,4], Virginie Desvignes[1], Eric Pelletier[2,3], Sébastien Chaussonnerie[2,3], Sonda Guermazi[2,3,4], Jean Weissenbach[2,3,4], Tianlun Li[1], Patricia Camacho[1] and Abdelghani Sghir[2,3,4]

[1]*CIRSEE Suez Environment, 38 rue du Président Wilson, Le Pecq, France;* [2]*CNRS-UMR 8030, 2 rue Gaston Crémieux, Evry, France;* [3]*CEA Genoscope, 2 rue Gaston Crémieux, Evry, France and* [4]*Université d'Evry Val d'Essonne, Boulevard François Mitterrand, Evry Cedex, France*

The microbial consortium involved in anaerobic digestion has not yet been precisely characterized and this process remains a 'black box' with limited efficiency. In this study, seven anaerobic sludge digesters were selected based on technology, type of sludge, process and water quality. The prokaryotic community of these digesters was examined by constructing and analysing a total of 9890 16S rRNA gene clones. Libraries were constructed using primers specific for the *Bacteria* and *Archaea* domains for each digester, respectively. After phylogenetic affiliation, the libraries were compared using statistical tools to determine the similarities or differences among the seven digesters. Results show that the prokaryotic community of an anaerobic digester is composed of phylotypes commonly found in all anaerobic digesters sampled and also of specific phylotypes. The *Archaea* community is represented by an equilibrium among a restricted number of operational taxonomic units (OTUs). These OTUs are affiliated with *Methanosarcinales*, *Methanomicrobiales* and Arc I phylogenetic groups. Statistical analysis revealed that the *Bacteria* community can be described as a three component model: one-third making up a core group of phylotypes common to most of the digesters, one-third are phylotypes shared among a few digesters and another one-third are specific phylotypes. The core group is composed of only six OTUs affiliated with *Chloroflexi*, *Betaproteobacteria*, *Bacteroidetes* and *Synergistetes*. Its role in anaerobic degradation appears critical to investigate. This comparison of anaerobic digester populations is a first step towards a future understanding of the relationship among biodiversity, operating conditions and digester efficiency.
The ISME Journal (2009) **3**, 700–714; doi:10.1038/ismej.2009.2; published online 26 February 2009
Subject Category: microbial engineering
Keywords: anaerobic digestion; *Archaea*; *Bacteria*; 16S rRNA gene libraries; statistical comparison

Introduction

The anaerobic digestion process has been used for years in wastewater sludge treatment and represents an attractive technique for sludge reduction along with energy production via biogas. However, this process remains a 'black box' and has limited performances. Our knowledge about the microbial consortia involved in this process is limited because of a lack of phylogenetic and metabolic data on these predominantly uncultivated microorganisms. As an alternative to culture techniques, several molecular

inventories, based on the study of the 16S rRNA gene, were carried out on anaerobic environments and have shown the extent of the diversity in these complex ecosystems (Godon *et al.*, 1997a, b; Sekiguchi *et al.*, 1998; Leclerc *et al.*, 2004; Chouari *et al.*, 2005a). These studies have limitations in that they often focused on one type of process, one sample, and often only lab-scale units. In our exploratory study, we analysed seven full-scale anaerobic digesters. The chemical components entering the plant were hardly ever known and were not controlled as in lab-scale experiments and microbial communities were adapted to a complex input. In some cases, toxic components from industrial effluents may have entered the wastewater treatment plant and disturbed the established communities.

Our anaerobic digesters were selected to represent conventional reactors working under standard

operating conditions, degrading volatile solids and producing biogas. All are located in municipal wastewater treatment plants which receive mostly domestic effluent and for some of them a small fraction of industrial effluents. In these reactors, organic matter is degraded with an efficiency varying between 40% and 55% of volatile solids removed.

In this study, we analysed the microbial diversity to see if differences in the microbial communities could explain the variability in efficiency. A previous analysis revealed that the same dominant phylogenetic groups were found in all the digesters but in variable proportions (unpublished data). These preliminary results led to several questions:

What is the structure of the community in an anaerobic digester? Are there populations common to all the digesters? Are the microorganisms the same? Are they specifically adapted to the treated effluent?

To answer these questions, we conducted an extensive molecular analysis of the prokaryotic diversity of seven digesters. Large 16S rRNA gene libraries were constructed targeting *Bacteria* and *Archaea* domains. A statistical comparison of the members of these populations was performed to compare the structure and the organization of the communities.

Within the past 5 years, several tools such as AMOVA (Excoffier *et al.*, 1992), UNIFRAC (Ley *et al.*, 2005), ∫-Libshuff (Schloss *et al.*, 2004), SONS (Schloss and Handelsman, 2006a) or Treeclimber (Schloss and Handelsman, 2006b) were developed and used to compare different types of microbial populations (Sakamoto *et al.*, 2000; Stach *et al.*, 2003; Ley *et al.*, 2005). Here, several statistical tools were selected from the statistic toolbox available to compare clone libraries from seven anaerobic digesters and to determine the similarities and differences among these communities.

Materials and methods

Description of digesters
Seven full-scale anaerobic digesters were selected in France, Germany and Chile to represent the conventional process used in Europe with efficiencies varying between 40% and 55% of volatile solids removed. All are cylindrical mesophilic reactors receiving municipal sewage sludge and mixed by biogas reinjection. The characteristics of the seven digesters are listed in Table 1. The industrial effluent represents <30% of the entering effluent which is a classical ratio for plants receiving municipal wastewater. Digester F receives effluents high in sugar whereas digester G receives slaughterhouse sludge. Three clone libraries (G, H and I) were constructed in parallel from the digester of Cholet. The original sample was divided in three aliquots that were treated as three different samples, from the DNA extraction step to the sequencing, to validate the reproducibility of the methods.

Sludge sampling
One litre of sludge was sampled during December 2006 from all digesters. Sludge was taken from the recirculation loop to obtain a representative sample, and samples were then aliquoted and frozen at $-20\,^{\circ}C$ for further analysis.

Table 1 Characteristics of the seven full-scale mesophilic anaerobic digesters treating municipal sewage sludge

	Location	Clone prefix in libraries	Waterline process		Temperature of the digester ($^{\circ}C$)	Sludge retention time (days)	Industrial effluent (%)	Efficiency (% volatile solids removed)
			Primary treatment	Secondary treatment				
A	Evry, France	*Bacteria*: QEDN *Archaea*: QEEL	Settling	Activated sludge	29	34	<10	43
B	Asnières, France	*Bacteria*: QEDP *Archaea*: QEEC	Settling	Activated sludge	37	30	<10	55
C	Clos de Hilde, France	*Bacteria*: QEDT *Archaea*: QEEG	Lamella and physico-chemical settling	Biological filter	36	22	0	55
D	El Trebal, Chile	*Bacteria*: QEDV *Archaea*: QEEH	Settling	Activated sludge	35	21	0	49
E	La Farfana, Chile	*Bacteria*: QEEA *Archaea*: QEEI	Settling	Activated sludge	35	21	—	54
F	Rostock, Germany	*Bacteria*: QEEB *Archaea*: QEEK	Settling	Activated sludge	37	17	28	47
G-H-I	Cholet, France	*Bacteria*: QEDQ, QEDR, QEDS *Archaea*: QEED, QEEE, QEEF	Lamella settling	Activated sludge	35	37	15[a]	40

[a]Of sludge entering the digester.

DNA extraction, PCR Amplification of 16S rRNA gene, cloning and sequencing
Genomic DNA was extracted from 200–250 mg aliquots as described previously (Chouari *et al.*, 2003). PCR amplification and cloning were carried out as already described with specific primers for *Bacteria* (8F—AGAGTTTGATCCTGGCTCAG; Hicks *et al.*, 1992) and *Archaea* (21F—TTCCGGTTGATCC YGCCGGA; DeLong, 1992) domains and a reverse universal primer (1390R—GACGGGCGGTGTGTAC AA; Zheng *et al.*, 1996). The nucleotide sequence of plasmid inserts was determined by classical auto-mated Sanger sequencing. For each sample, 1000 clones were sequenced for *Bacteria* clone libraries and 384 for *Archaea* (3730XL sequencer; Applied Biosystems, Darmstadt, Germany).

Sequence analysis
Assembly. The 16S rRNA gene sequences were assembled by Phrap (http://www.phrap.org). Only good quality 16S rRNA gene sequences (longer than 1200 bp and with Phred qualities of above 15 for each base) were selected for further analysis. Chimera were searched before analysis by using the procedure described by Juretschko *et al.* (1998) followed by an analysis with the Pintail tool which compares the sequences with a set of reference to identify chimeras (Ashelford *et al.*, 2005).

Phylogenetic affiliation. The retrieved sequences were compared by BLAST (Altschul *et al.*, 1990) to three different databases:

- the Greengenes database updated on the 28th of May 2007 (DeSantis *et al.*, 2006);
- the RDP database 9.50. (Maidak *et al.*, 1994);
- an 'in-house' database with sequences obtained from anaerobic digesters (Chouari *et al.*, 2003, 2005a, b; Guermazi *et al.*, 2008).

Sequences were phylogenetically affiliated with a given taxon only if the alignment length was ⩾1000 bp with ⩾90% sequence identity. Otherwise, sequences were considered unclassified if these criteria were not respected or if the results from the different databases were different.

OTU assignment. 16S rRNA gene sequences were aligned with MAFFT (Katoh *et al.*, 2002), distance matrices were calculated using DNAdist software using the F84 model (Felsenstein and Churchill, 1996) and operational taxonomic units (OTUs) were defined using the furthest neighbour clustering algorithm of DOTUR software with a 97% identity threshold (Schloss and Handelsman, 2005). Total richness of the samples (total number of OTUs) was extrapolated from the observed OTUs (total number of OTUs in the libraries) with two non-parametric estimators: ACE (Chao and Yang, 1993) and Chao1 (Chao, 1984) using DOTUR software. Diversity coverage was calculated using Good's formula (Good, 1953).

Phylogenetic tree construction. The tree in Figure 5 was constructed using the ARB program and database package (Ludwig *et al.*, 2004). It was built with the Neighbour Joining algorithm and Felsenstein correction. *Methanomicrobium mobile* was used as the outgroup.

Statistical comparison of the clone libraries
∫-Libshuff. The seven digesters were compared bilaterally with ∫-Libshuff (Schloss *et al.*, 2004) using the distance matrix calculated by DNAdist as input. With a Monte Carlo procedure, the probabilities that the observed difference among the diges-ters is due to chance were calculated. With an experimentwise error rate of 0.05, and taking into account a Bonferroni correction due to the multiple comparisons, the libraries were considered signi-ficantly different if the *P* value was inferior to 0.000712.

Treeclimber. The software Quicktree (Howe *et al.*, 2002) was used to build the general trees for *Bacteria* and *Archaea* domains. Quicktree is based on the Neighbour Joining algorithm. For *Archaea* trees, the *Chloroflexi* sequence CU927657 was used as an outgroup and for *Bacteria* trees we used a *Methanosarcinales* sequence CU917424.
Treeclimber (Schloss and Handelsman, 2006b) was then applied to these phylogenetic trees to determine whether the topology of the tree reflects the real differences among the communities or is due to an accumulation of random variation. Treeclimber implements the parsimony test for calculating scores for user tree and random trees.

SONS. Results obtained with DOTUR at a dis-tance of 0.03 were used to run SONS (Schloss and Handelsman, 2006a) and to compare the seven 16S clone libraries. SONS was used to compare the membership and structures in communities by accounting for OTU abundance and distributions that are either specific to one library or shared by two libraries. By comparing the seven libraries, SONS allowed us to distinguish the shared OTUs among the total number of observed OTUs and the number of sequences in each OTU.

Nucleotide sequence accession numbers
Sequences reported in this study were deposited in the EMBL databases under accession numbers CU915828 to CU917468 for *Archaea* sequences and CU917469 to CU927884 for *Bacteria* sequences. The correspondence among accession numbers, clones names and OTU numbers is given in Supplementary Table 1.

Results

16S rRNA gene clone distribution among different phylogenetic groups

Taxonomic affiliation of the closest relatives of the sequences was used to assign each clone to a phylogenetic group. We adopted Hugenholtz phyla as defined in the Greengenes database except for *Synergistetes, Coprothermobacteria* and WWE1, which were considered as separate phyla. The average proportion of each group, among the seven digesters, is listed in Table 2.

For *Euryarchaeaota*, the three main phylogenetic groups are *Methanosarcinales*, the lineage Arc I (Chouari *et al.*, 2003) and *Methanomicrobiales*. Their proportions vary among the different libraries. Sequences affiliated with *Methanobacteriales* were also found within the archaeal clone libraries but in smaller proportions. Sequences affiliated with *Crenarchaeota* were also found in the libraries in small proportions.

For the *Bacteria* domain, sequence analyses revealed dominant groups in various proportions in the different digesters. Most of the sequences are affiliated with four major groups: *Chloroflexi, Proteobacteria, Bacteroidetes* and *Firmicutes*. Other phylogenetic groups such as Aminanaerobia, WWE1, *Actinobacteria, Synergistetes, Coprothermobacteria, Spirochaetes* can represent a large proportion (up to 12%) of the sequences but they are not necessarily present in the seven digesters. The rare phyla (<1% of the sequences) were grouped into 'minor groups'; they are composed of sequences affiliated with *Chlorobi, Cyanobacteria, Fusobacteria, Lentisphaerae, Planctomycetes, Thermotogae*, CC331, NKB19, OD1, OD2, OP5, OP8, OP9, OP10, OP11, JS1, WS1, WPS-2, TM6, TM7, Marine gp A, BRC1. For the 'unclassified' sequences, no close relative was found in the databases using our BLAST criteria. They are phylogenetically far from the sequences present in the database. These results show that anaerobic digesters are specific ecosystems with a high diversity and that they contain some sequences that have no relatives in other environments and are therefore still hard to affiliate.

As shown in Table 2, at the phylum level, the major groups are the same in the seven digesters, but a more detailed analysis is needed if we want to explore the similarities and the differences among the digesters' populations at a deeper phylogenetic level.

Statistical comparisons of the libraries were performed sequence by sequence and also at the OTU level to distinguish the most abundant microorganisms in the anaerobic degradation of organic matter.

Evaluation of diversity coverage and richness of the clone libraries

To evaluate the level of information contained in the different libraries, several parameters were measured.

First, we compared the observed number of OTUs to the estimated total number of OTUs.

We used two estimators ACE and Chao1 to calculate the total diversity of each digester based on the observed OTUs (Table 3).

For *Archaea*, the observed number of OTUs obtained with about 200 clones (154–221) approaches the total diversity of the sample given by the estimators in most of the cases. This result indicates that our libraries cover almost all the diversity originally present in the sample.

For *Bacteria*, the observed number of OTUs is far from reaching the estimated total number of OTUs. This shows that the diversity was not exhaustively sampled. Moreover, we showed that even with an equal extensive sequencing for the nine libraries, the estimated total richness can vary and is specific to a given sample. For about 1000 clones initially sequenced, the total richness was estimated to be between 1000 and 2000 OTUs for most of the libraries. Library H is the most diversified with 3332 and 3342 estimated OTUs depending on the estimator used. Nevertheless, the confidence intervals of the estimators are larger for this library, which indicates that the estimation is less accurate. The difference is due to the large proportion of OTUs with one sequence in library H, compared to libraries G and I, which increases the value of the

Table 2 Proportion of the main phylogenetic groups among the different digesters

Phylogenetic group	Average proportion of each group in % ± standard deviation	(Min–max)
Bacteria domain		
Chloroflexi	32 ± 9	(15–45)
Proteobacteria	18 ± 5	(11–24)
Bacteroidetes	11 ± 6	(3–25)
Firmicutes	9 ± 6	(5–25)
Aminanaerobia[a]	2 ± 2	(ND–6)
WWE1	2 ± 4	(ND–12)
Actinobacteria	2 ± 1	(0.4–2)
Synergistetes	4 ± 4	(0.4–12)
Coprothermobacteria	1 ± 2	(ND–7)
Spirochaete	1 ± 1	(ND–2)
Minor groups	6 ± 3	(2–9)
Unclassified	12 ± 4	(6–19)
Archaea domain		
Euryarchaeota		
Methanosarcinales	51 ± 30	(12–93)
ArcI	36 ± 34	(ND–77)
Methanomicrobiales	10 ± 2	(2–35)
Methanobacteriales	0.2 ± 0.3	(ND–1)
Crenarchaeota	2 ± 3	(ND–7)

Abbreviation: ND, not detectable in our library.
For each group the average proportion of sequences in each digester was calculated, and the range and standard deviation.
[a]Aminanaerobia as defined by Hugenholtz minus the *Synergistetes* that were considered apart (DeSantis *et al.*, 2006).

Table 3 Comparison of diversity estimators and coverage for *Archaea* and *Bacteria* 16S r RNA gene clone libraries constructed from 9 samples recovered from anaerobic digesters

Digester	Archaea *domain*					Bacteria *domain*				
	No. of OTUs	No. of Sequences	Coverage (%)[a]	Richness estimation[b]		No. of OTUs	No. of Sequences	Coverage (%)[a]	Richness estimation[b]	
				ACE	Chao1				ACE	Chao1
A	8	154	97	16 (9–57)	11 (8–31)	378	1000	71	1555 (1224–2018)	1339 (1037–1779)
B	10	212	98	15 (11–37)	13 (10–33)	299	961	77	1041 (802–1392)	999 (760–1361)
C	11	221	98	20 (13–52)	13 (11–28)	238	931	82	748 (586–985)	683 (517–950)
D	8	206	98	14 (9–49)	11 (8–31)	344	960	73	1346 (1067–1732)	1364 (1014–1896)
E	8	133	97	16 (9–57)	11 (8–31)	262	639	66	1567 (1120–2246)	1399 (958–2118)
F	23	183	93	37 (27–71)	39 (27–87)	381	947	69	1739 (1351–2281)	1561 (1179–2126)
G	14	164	94	100 (28–560)	59 (28–162)	441	846	59	1946 (1548–2486)	1887 (1444–2525)
H	26	203	92	70 (39–171)	71 (39–183)	514	988	56	3332 (2594–4331)	3432 (2488–4801)
I	20	165	92	55 (27–190)	39 (25–93)	468	977	63	2005 (1615–2528)	1874 (1458–2464)
All together	69	1641	98	99 (83–135)	95 (79–132)	2661	8249	72	11641 (10989–12345)	10715 (9588–12025)

Abbreviation: OTU, operational taxonomic unit.
[a]Calculated by Good's formula.
[b]Values in parentheses are the 95% confidence intervals.

estimators. This result shows that richness estimations are sensitive to the number of rare OTUs and they must be used with caution but the confidence interval is a clue to help in the determination of the accuracy of the richness estimation.

Another method for evaluation of the level of information contained in the libraries is calculation of the diversity coverage. It represents the percentage of chance for a clone newly sequenced to belong to an OTU already observed. In this study, coverage values are >92% for *Archaea* libraries and vary between 56% and 82% for *Bacteria* libraries. These values confirm that the most common phylogenetic groups were detected in our libraries.

These results show that for *Bacteria*, even with a large number of clones sequenced, the libraries do not reflect the total diversity of the original sample due to the large diversity encountered in anaerobic environments. This implies that the rare phylotypes cannot be compared because they are not equally sampled in the libraries. Nevertheless, the high coverage values allow us to study the most abundant phylogenetic groups found in anaerobic digesters, which are well represented in our 16S rRNA gene libraries.

Comparisons of the different communities will help determine which proportion of the population of a digester is shared with other digesters and which proportion is specific.

Statistical comparison of the 16S rRNA gene libraries
Comparison of the libraries sequence by sequence. ∫-Libshuff software was used to compare the libraries sequence by sequence for a range of distances taking into account both frequently observed and rare sequences. A pairwise comparison was achieved for the nine libraries

using ∫-Libshuff for *Bacteria* and *Archaea* domains.

In both domains, libraries G, H and I are statistically similar. These libraries were built in parallel from the same sludge sample. Therefore, the results confirm that the entire procedure used for the construction of the libraries is reproducible.

The other digesters are considered as different by ∫-Libshuff as this software does not take into account the frequency of the sequences. As a consequence, several digesters sharing the same major groups but with different rare phylotypes are considered by ∫-Libshuff as significantly different.

In this study, the use of ∫-Libshuff showed that the different libraries are constituted of a majority of sequences specific of one digester, which is an obstacle in the search of sequences shared among all digesters. In conclusion, the results of ∫-Libshuff can underline some similarities among 16S rRNA gene libraries but further analyses with other tools were necessary to explore the similarities among the libraries.

As it was demonstrated here that libraries G, H, and I are identical, they were pooled and considered as a unique sample in the rest of the study considering that the OTUs found only in these libraries as specific of Cholet digester. Our analysis showed that fusion of these data has no impact on the statistical analysis (data unpublished).

Comparison of the libraries based on phylogenetic trees. Phylogenetic trees constructed from several libraries were used to evaluate the similarities among the sequences without using any clustering method. Trees were built for *Bacteria* and *Archaea* domains using the total number of sequences retrieved in the different libraries. First, to confirm

if the topology of the tree is not due to chance, parsimony scores of the trees were compared to scores of 1000 random trees with Treeclimber software. A comparison of the parsimony scores for input tree to the cumulative probability distribution indicates the probability of observing the same or a lower parsimony score by chance.

For the *Archaea* tree, the parsimony scores are distributed between 1290 and 1364 whereas the parsimony score of the input tree is 662. The calculated *P* value is <0.001. For the *Bacteria* tree, the same results are observed with a parsimony score of 3783 whereas the scores of random trees vary between 5802 and 5948 leading to a *P* value <0.001. The parsimony scores of the input trees could not be equalled by any random tree which indicates that the topology of the trees is not due to chance, and reflects the real structure of the communities.

We used the trees to perform a first analysis of the structure of the prokaryotic communities. For an easier reading, sequences were grouped at the OTU level. Considering the digester of origin for each sequence, the trees may give indications for shared OTUs found in several digesters. The *Methanosarcinales* tree and a fraction of the tree for *Betaproteobacteria* are represented in Figure 1 as examples.

On the *Methanosarcinales* tree, the majority of sequences form abundant groups that are shared among all digesters. The rest of sequences are either specific of one digester or belong to smaller OTUs of less than five sequences. This result is consistent with the low diversity of *Archaea*. The community is therefore mainly composed of several recurrent phylotypes found in every digester and a small proportion of rare phylotypes.

On the *Betaproteobacteria* tree, it can be observed that the majority of the branches are composed of a unique sequence, which reflects the high number of OTUs specific of one digester. On the other hand, sequences retrieved within different clone libraries can also be grouped together indicating similarity (62% for *Betaproteobacteria*). As a consequence, the tree shows that the *Bacteria* community is composed of a high proportion of specific sequences and also of shared sequences.

After this first analysis based on the observation of the trees, a further analysis with SONS was necessary to find the group of microorganisms shared by the different digesters and to quantify the fraction of the communities represented by shared or specific phylotypes.

Comparison of libraries based on OTU clustering.
SONS was used to find the shared and specific OTUs among the different digesters. The sequences retrieved from the nine libraries were aligned together and then grouped into OTUs. The 1614 *Archaea* sequences formed 69 OTUs and the 8249 *Bacteria* sequences 2661 OTUs (Table 3). To find the shared microorganisms among the different anaerobic digesters, the number of shared OTUs was plotted versus the number of digesters sharing each OTU (Figure 2).

Archaea domain. Proportion of specific and shared OTUs: For *Archaea*, the majority of the sequences belong to 3 OTUs that are shared among 5–7 digesters representing 1028 sequences (64% of the total number of *Archaea* sequences; Figure 2a). Then, 24 other OTUs are shared among 2–4 digesters and represent a total of 559 sequences (35%). The last 42 OTUs are specific of one digester and account for 59 sequences (3.6% of the total number of sequences). Among these specific OTUs, 30 OTUs are composed of a unique sequence, which represents 0.8% of the total number of sequences. We can deduce that most of the OTUs are represented by several sequences in the libraries confirming a good coverage of the diversity. The small proportion of specific sequences is consistent with the low diversity of *Archaea* and confirms the observations made previously using phylogenetic trees.

The proportion of specific and shared OTUs in each library is subject to variations (Figure 3a). For libraries A, B, D and E, the OTUs shared among 5–7 digesters represents between 70% and 90% of the sequences. But for libraries C, F and G-H-I, the OTUs shared among 2–4 digesters represent between 40% and 80% of the sequences. It appears that the organization of *Archaea* population can follow different profiles with a balance among different OTUs. This means that some OTUs are abundant wheareas others are absent and vice versa. A closer look at the phylogenetic affiliation of these OTUs and their metabolic capacities should help in understanding the type of interaction among these microorganisms.

Characterization of the shared OTUs: As shown in the Figure 2a, most of the sequences are grouped into three major OTUs (V, VI, VII) but, to compare the major OTUs, the contribution of each digester is summarized for the major OTUs (more than 40 sequences) in Table 4. These shared OTUs are affiliated with *Methanosarcinales*, Arc I, *Methanomicrobiales* and *Crenarchaeota*.

First, the most represented group is Arc I, a lineage first described by Chouari *et al.* in 2005 and also called WSA2 by Hugenholtz. It is represented mainly by OTUs IV and V. The OTU V is the most numerous with 553 sequences whereas OTU IV accounts for 88 sequences. OTU V is present in five digesters but two cases could be distinguished. First, this OTU may represent a major portion of the sequences in the libraries A, B and G-H-I. On the contrary, this OTU may also represent a minority in the libraries E and F or is absent in libraries C and D. For library F, the lineage Arc I is represented by OTU IV. The absence of this lineage in the libraries C, D and E seems to indicate that the function of this lineage may also be assumed by other phylogenetic groups.
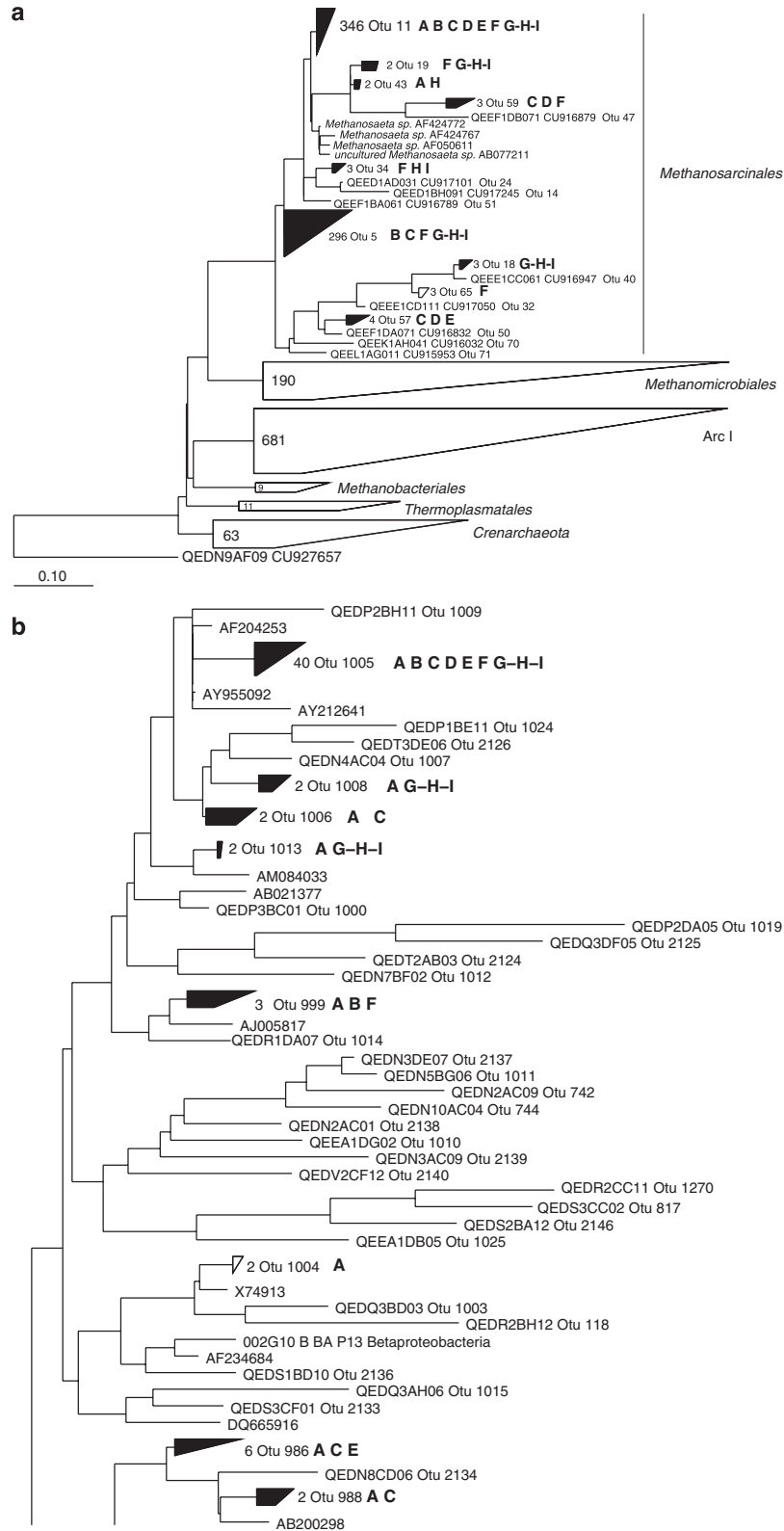
706

Figure 1 Phylogenetic trees for *Methanosarcinales* (**a**) and *Betaproteobacteria* (**b**) constructed using the Neighbour Joining method and the Felsenstein correction. For *Betaproteobacteria,* only a fraction of the entire tree is represented as an example of the structure of *Bacteria* communities. For the *Methanosarcinales* tree, the *Bacteria* sequence CU927657 was used as an outgroup and for the *Betaproteobacteria* tree the *Archaea* sequence CU917424 was the outgroup. Each operational taxonomic unit (OTU) is represented with the number of sequences and the digesters they came from. *Methanosarcinales* are composed mainly of shared OTUs (black triangles) whereas *Betaproteobacteria* exhibit a high proportion of specific OTUs (white triangles).
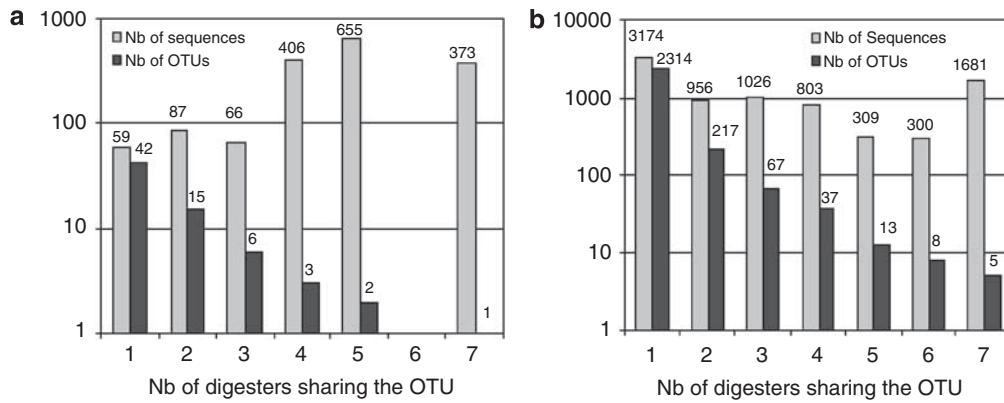
Figure 2 Distribution of the shared operational taxonomic units (OTUs) according to the number of digesters sharing the OTU. (**a**) *Archaea*, (**b**) *Bacteria*.
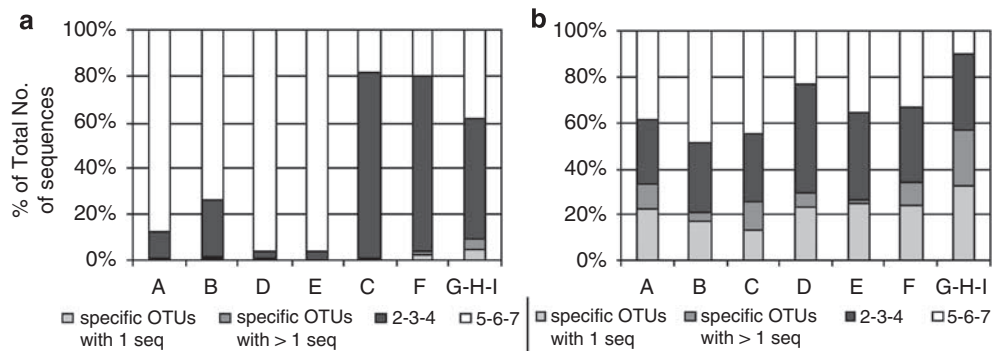

Figure 3 Proportion of specific and shared operational taxonomic units (OTUs) in seven anaerobic digesters. (**a**) *Archaea*, (**b**) *Bacteria*. 2-3-4: OTUs shared among 2–4 digesters. 5-6-7: OTUs shared among 5–7 digesters. G-H-I: the three libraries were combined because they were from the same digester.

Table 4 Affiliation of the major *Archaea* OTUs ($\geqslant$40 seq) and distribution of the total number of sequences in each library (%)

OTU No	Library							Total no. of sequences	Phylogenetic affiliation	Nearest neighbour acc. no.	BLAST score (% of identity)
	A	*B*	*C*	*D*	*E*	*F*	*G-H-I*				
I	0	0	0	0	1	2	8	46	*Crenarchaeotes*	AJ576215	>97
II	0	0	0	0	0	2	8	45	*Methanomicrobiales*	EF198037	>99
III	0	3	3	35	8	1	0	97	*Methanomicrobiales*	AB175345	>97
IV	4	0	1	0	0	41	1	88	Arc I	CT573602	>99
V	49	69	0	0	3	2	61	553	Arc I	CT573729	>99
VI	0	23	77	0	0	22	10	311	*Methanosarcinales*	AF424772	>98
VII	40	1	15	60	84	18	1	373	*Methanosarcinales*	AB248614	>99

Abbreviation: OTU, operational taxonomic unit.

We confirm here the importance of this lineage that is well represented. There is little data available about the metabolic capabilities of the Arc I lineage. It has been demonstrated that this group can grow on formate or H_2/CO_2 (Chouari *et al.*, 2005a), which may indicate a hydrogenotrophic population. However, the metabolism of this lineage has not been completely explored and this group may be able to degrade other substrates such as acetate.

An equilibrium between the OTUs IV-V and the OTUs VI-VII seems to exist. In libraries B and G-H-I, where the OTU V is dominant, the OTUs VI and VII represent a total of 24% and 11% of the sequences, respectively. But in libraries C, D and E, where OTU

708

V is absent or is minor, OTUs VI and VII reach a total of 92%, 60%, 84% and 40% of the sequences, respectively. In library A, where OTU V represents 49% of the sequences, the OTU VII accounts for 40% of the sequences. And finally, in digester F where the OTU IV represents 41% of the sequences, OTU VI and VII account for 40% of the sequences. As these last two OTUs are affiliated with the species *Methanosaeta consilii*, which degrades acetate exclusively, it may reflect a competition for acetate between *Methanosarcinales* and Arc I. This behaviour underlines the possible ability of Arc I to degrade acetate and its role in acetoclastic methanogenesis.

The hydrogenotrophic microorganisms are represented by the less abundant OTUs II and III, which are affiliated with *Methanomicrobiales*. OTU II is present only in libraries F and G-H-I. The OTU III is common to five digesters and represents generally a small fraction of the sequences (0–8%) except for library D where it is particularly abundant, accounting for 35% of the sequences. More precisely, OTU III is affiliated with the genus *Methanospirillaceae*, which is able to degrade formate or H_2-CO_2.

Bacteria domain. Proportion of specific and shared OTUs: For *Bacteria*, 26 OTUs are shared among 5–7 digesters and represent 2290 sequences (28% of the total number of sequences; Figure 2b); 321 other OTUs are shared among 2–4 digesters and represent a total of 1885 sequences (34%). And finally, 2314 OTUs are specific of one digester and accounts for 3174 sequences (38.5%). Therefore, it appears clearly that the bacterial communities are composed mainly of specific sequences.

Of these specific OTUs, 2042 are composed of a unique sequence, which represents 25% of the total sequences. This result indicates that, even after sequencing more than 8000 *Bacteria* clones, 77% of the OTUs contain only one sequence and represent rare phylotypes. Surprisingly, it appeared that these OTUs are mainly affiliated with predominant groups such as *Proteobacteria*, *Chloroflexi*, *Firmicutes* and *Bacteroidetes.* This result indicates that even for phylogenetic groups commonly found in anaerobic digesters, particular populations can develop and be specific of one digester.

Consequently, the *Bacteria* population of a digester seems to be composed of three components:

- A core group of OTUs shared with almost all the digesters;
- Shared OTUs found in some digesters;
- OTUs specific of one digester.

In Figure 3b, the proportions of these three types of populations are represented for each digester.

For most of the digesters, the three types of populations are equally distributed and each represents approximately 30% of the total number of sequences. However, for libraries G-H-I the propor-
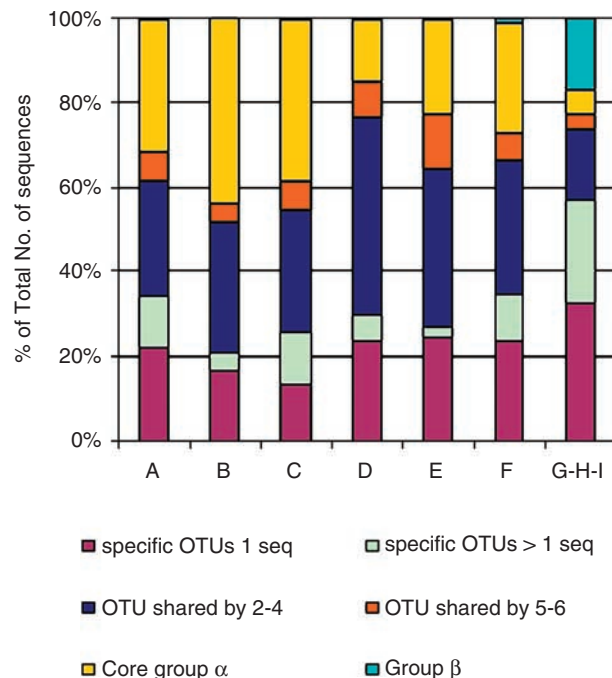

Figure 4 Proportion of the groups α and β in each *Bacteria* library in terms of percentage of the total number of sequences. Specific operational taxonomic unit (OTU) 1 seq: specific OTU containing one sequence. Specific OTU >1 seq: specific OTU with more than one sequence. OTU shared by 2–4: OTUs shared among 2–4 digesters. OTU shared by 5–6: OTUs shared between 5 or 6 digesters. G-H-I: the three libraries were combined because they were from the same digester.

tion of sequences belonging to OTUs specific of this digester reaches almost 60% indicating a particularity of these libraries.

Characterization of the shared OTUs: Based on our observations from Figure 3b, among the shared OTUs two groups were defined:

- Core group α: OTUs shared among most of the digesters;
- Group β: OTUs specific of libraries G-H-I

The proportion of each group is represented in Figure 4.

Core group α: Core group α includes all the OTUs shared among seven digesters and also the major OTUs (>100 sequences) shared among five or six digesters. This core group represents 23% of the total number of sequences (1922 sequences). These sequences are distributed among six OTUs (Table 5). The major one, called OTU VI represents between 5% and 41% of the sequences in each library. OTUs I, III and IV are also shared among the seven digesters but represent a small fraction of the sequences (Table 5). OTUs II and V are shared among six or five digesters and represent up to 2% and 9%, respectively.

The existence of this core group indicates that anaerobic digesters located in different countries,

Table 5 Composition and distribution of groups α and β

No.	A	B	C	D	E	F	G-H-I	Total no. of sequences	Phylogenetic affiliation	Nearest neighbour acc. no.	BLAST score (% of identity)
Core group α											
I	0.1	0.6	0.1	0.3	0.2	0.2	0.1	16	*Betaproteobacteria*	AJ012071	>99
II	1.9	0.2	1.0	0.3	0.2	0.2	0.1	40	*Betaproteobacteria*	AY212635	>99
III	2.0	0.5	0.3	1.0	0.8	1.4	0.2	62	*Chloroflexi*	CR933321	>99
IV	0.6	1.4	6.8	0.8	0.8	4.2	0.5	149	*Synergistetes*	AF280863	>99
V	9.4	0.4	0.3	0.0	0.0	1.5	1.7	162	*Bacteroidetes*	CT573982	>99
VI	17.4	40.7	29.6	12.4	20.8	18.6	5.2	1414	*Chloroflexi*	EF174275	>99
Total	31.4	43.8	38.1	14.9	22.7	26.1	7.8	1843			
Group β											
VII	0.0	0.0	0.0	0.0	0.0	0.1	0.8	24	*Chloroflexi*	AJ249100	>99
VIII	0.0	0.0	0.0	0.0	0.0	0.1	0.9	25	*Bacteroidetes*	AY218558	94–96
IX	0.0	0.0	0.0	0.0	0.2	0.2	1.0	30	*Bacteroidetes*	AY953168	97–98
X	0.0	0.0	0.0	0.0	0.2	0.1	2.1	62	*Chloroflexi*	EF515576	>99
XI	0.3	0.0	0.0	0.0	0.0	0.6	5.6	167	*Chloroflexi*	AF323767	94–95
XII	0.0	0.0	0.5	0.2	0.0	0.0	6.4	186	*Betaproteobacteria*	DQ836253	100
Total	0.3	0.0	0.5	0.2	0.3	1.2	16.8	494			

Group α is a core group composed of OTUs shared between 6 or 7 digesters and also the OTUs shared between 5 digesters with more than 100 sequences. Group β is composed of OTUs specifically abundant in libraries G-H-I. The contribution of each digester is expressed in % of the total number of sequences in each library and the closest relatives in the public database are also given.

treating different effluents can share similar populations. This core group α is present in all digesters and represents between 8% and 44% of the sequences in each libraries.

Group β: The only case where the core group α is less represented is in the libraries G-H-I, where it represents only 8% of the sequences. This is consistent with the higher percentage of specific OTUs in this digester as was mentioned in Figure 3b. To elucidate this specificity, we explored the specific OTUs from the libraries G-H-I. Looking at the OTUs affiliated with *Chloroflexi*, these libraries appear to have a particular population. Indeed, libraries G-H-I are the only case where the OTU VI is minor with only 5% of the sequences. In this digester, *Chloroflexi* are grouped into three other OTUs (VII, X and XI in Table 5) that are underrepresented in other libraries. Finally, as libraries G-H-I seemed to have a different profile, all the OTUs specifically abundant in these libraries were grouped together and called 'group β'. Group β represents 17% of the sequences in G-H-I libraries but less than 1% in all the other libraries.

The different OTUs constituting the groups α and β are represented with their closest relatives in the phylogenetic tree, Figure 5.

Affiliation of groups α and β: To study the differences among the OTUs present in the groups α and β, the taxonomic affiliation was pursued to a deeper phylogenetic level.

Three OTUs are affiliated with *Betaproteobacteria*. The members of the OTU I are affiliated with the genus *Thauera* and more precisely the species *T. phenylacetica* and *T. aminaromatica*. The OTU II is affiliated with the order *Rhodocyclales*.

The OTU XII is affiliated with *Brachymonas*. It is difficult to comment on the metabolic differences between these two orders, as few representatives have been completely sequenced to date.

Five OTUs were affiliated with *Chloroflexi* in which OTU VI contains 74% of the total sequences affiliated with this phylum. BLAST results revealed that the closest relatives of the *Chloroflexi* in the group β (OTUs VII, X and XI) are sequences retrieved from anaerobic bioreactors, which treat chlorinated compounds, sludge or benzoate, respectively. The *Chloroflexi* of core group α (OTUs III and VI) seem to be more characteristic of wastewater sludge because their closest relatives were found in activated sludge and in an anaerobic digester treating municipal sludge. Nevertheless, it should be noted that all the *Chloroflexi* sequences are affiliated with the order *Anaerolineales* and more precisely with the groups Eub 4 and SHA-24 that were detected in anaerobic consortia which degrade toluene and 1,2 dichloropropane (Ficker *et al.*, 1999; Schlotelburg *et al.*, 2000).

OTU IV is affiliated with *Synergistetes* and its closest relative was retrieved from an anaerobic digester treating pharmaceutical wastewater.

For *Bacteroidetes,* the three OTUs (V, VIII, IX) are affiliated with the order *Bacteroidales* and with the subgroups vadinBC27 and SB-1. The OTUs VIII and IX of the group β are related to sequences retrieved from Antarctica sediments and an anaerobic swine lagoon whereas OTU V is related to a clone from a previous study on the digester of Evry.

As a matter of fact, group β seems to be composed of microorganisms that are not specific of the environment of anaerobic digesters. On the other
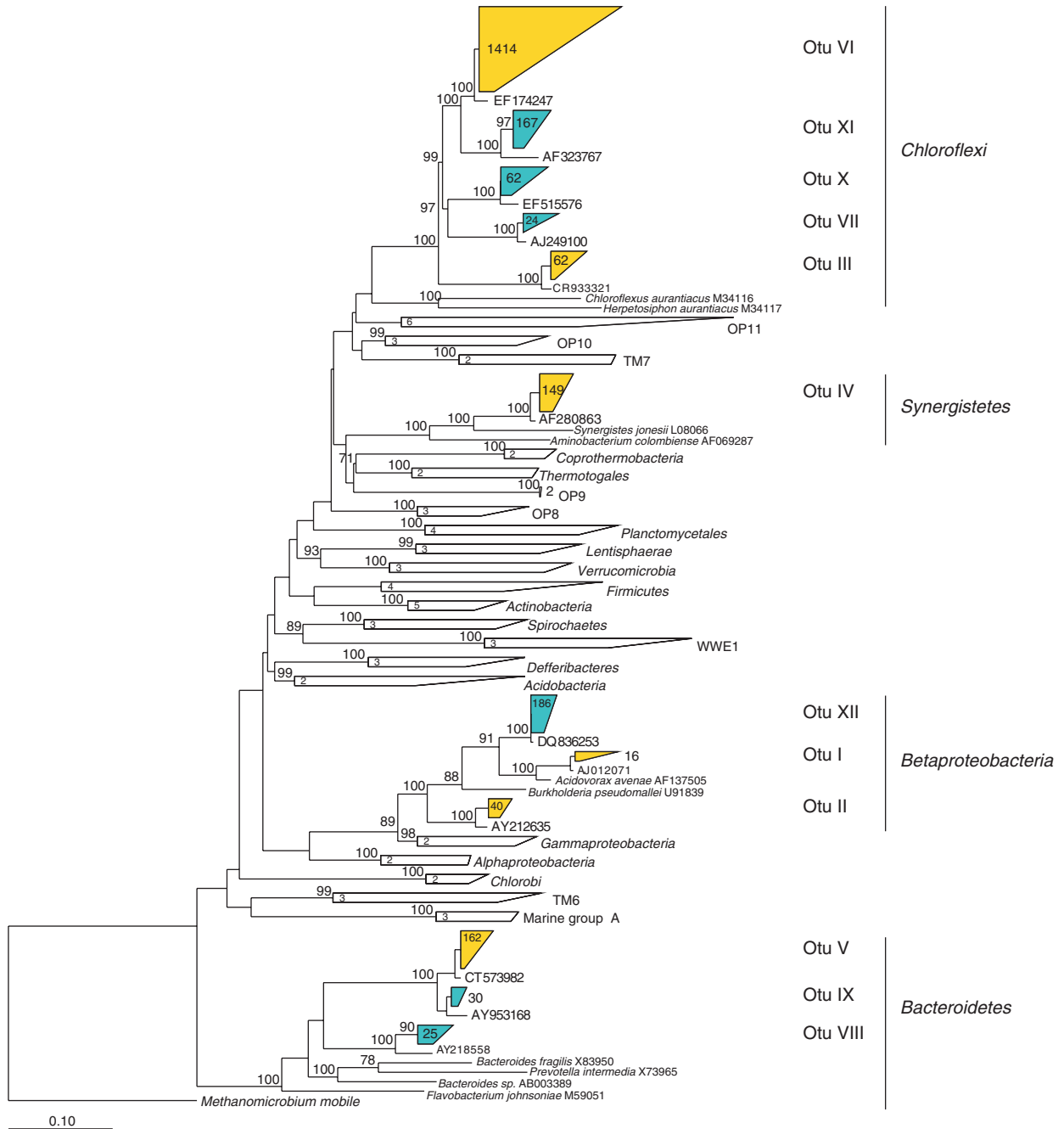
Figure 5 Phylogenetic tree constructed using the Neighbour Joining method and the Felsenstein correction showing the affiliations of the environmental 16S rRNA gene sequences constituting the groups α (in yellow) and β (in blue). The number of sequences in each operational taxonomic unit (OTU) is represented in each triangle. Only the bootstrap values superior to 70% are represented. *Methanomicrobium mobile* was used as outgroup.

hand, the core group α seems to be typical of anaerobic digesters based on the closest relatives of the OTUs, which were all retrieved from anaerobic digesters or reactors treating solid wastes.

The specificity of group β may be the result of a difference in the composition of the sludge. Indeed, samples G, H, I were taken from the digester of Cholet where slaughterhouse sludge is treated. A particular population (for example lipid degraders)

could have developed to treat specific substrates that are not present in conventional digesters. For digester F, which receives also industrial effluents (charged in sugar), there was no particular populations identified. In this study, we were not able to establish a direct link among the populations encountered and the specificity of the effluent based only on 16S rRNA gene library data. However, future analysis of the effluent could help to determine if

this specialization of the population can be correlated with the composition of the sludge treated in the digester.

Surprisingly, several phylogenetic groups that are abundant in each library are not represented in the core group α. Among the *Proteobacteria*, the only group present in the core group α is the *Betaproteobacteria*. This result is linked to the intradivergence of each phylum. *Betaproteobacteria* OTUs are composed of 50% shared OTUs and 50% specific OTUs which explains why they are found in the core group α. On the contrary, OTUs affiliated with *Alphaproteobacteria* are composed of less than 20 sequences indicating a higher diversity of this group. As a consequence, there are no major shared OTUs affiliated with *Alphaproteobacteria*. For *Deltaproteobacteria*, the analysis revealed that most of the OTUs affiliated with this group are specific of one digester (73%) and only 27% of them are shared among 2, 3 or 4 digesters.

For *Bacteroidetes*, only one OTU was found in the core group whereas *Bacteroidetes* represent an average of 12% of the sequences in the libraries. Actually, *Bacteroidetes* OTUs are composed of 50–70% of OTUs specific of one digester. Concerning the *Firmicutes*, they are absent in the core group α whereas they represent an average of 11% of the sequences. A closer look at the OTUs affiliated with *Firmicutes* revealed that there is no OTU common to all digesters. The OTUs affiliated with *Firmicutes* are either specific (40–50%) or shared among 2–4 digesters.

Therefore, the individual variability among the phyla can explain why the core group α is composed of OTUs belonging to the less diversified phyla. The phyla with deep phylogenetic divergence are composed mainly of OTUs specific of one digester and an adaptation to the treated effluent could be emphasized.

These results are not influenced by the threshold of 97% of identity used to cluster the sequences into OTUs. Indeed, the same analysis was performed at 80%, 90%, 95%, 96%, 98% and 99% of identity and the phylogenetic composition of the core group is the same at any threshold. The size of the core group is not subject to huge variations above an identity of 95%. It begins to increase significantly when a threshold of 90% or 80% is chosen with an increase of, respectively, 26% and 70% of the number of sequences included in the core group. For inferior threshold, the number of sequences increases of less than 10% at each step. So, the existence of the core group and the phylogenetic affiliation of its OTUs are legitimate independently of the distance.

Metabolic function: Based on the phylogenetic affiliation of the different OTUs and on the data available in the literature, it is possible to form hypotheses on the metabolic functions of the groups.

Concerning the *Chloroflexi*, the metabolic capacities of these microorganisms are still unclear as few cultivable representatives have been isolated so far. Nevertheless, as this group is abundant in many environments, several studies have attempted to investigate their metabolic function. Several studies showed their potential role in the degradation of carbohydrates (Sekiguchi *et al.*, 2001; Kindaichi *et al.*, 2004; Ariesyady *et al.*, 2007). *Betaproteobacteria* are also microorganisms involved in the first steps of the degradation. Ariesyadi *et al.* showed that they are the main consumers of propionate, butyrate and acetate. *Bacteroidetes* are known to be proteolytic bacteria (Zehnder, 1988; Kindaichi *et al.*, 2004). They first intervene in the degradation of proteins and are able to ferment amino acids to acetate. *Synergistetes* can use amino acids and in turn provide short-chain fatty acids and sulphate for terminal degraders such as the methanogens and sulphate-reducing bacteria (Vartoukian *et al.*, 2007).

Firmicutes are syntrophic bacteria, which can degrade volatile fatty acids such as butyrate and its analogs. This degradation produces H_2, which is then degraded by hydrogenotrophic methanogens.

Based on the current knowledge of the metabolism of anaerobic microorganisms, the core group includes bacteria, which degrade carbohydrates, proteins and amino acids during the hydrolysis and fermentation steps and also acetogenic species that degrade volatile fatty acids to form acetate.

Discussion

In this study, we analysed microbial populations of our seven digesters and we used statistical tools to compare 16S rRNA gene clone libraries.

Various statistical tools exist in the literature to compare libraries. Two major categories can be distinguished: tools based on OTU clustering and tools independent of the choice of a distance. In our study, both were used and gave consistent results. However, some limitations of this statistical approach can be underlined. First, a selection of tools was used in our study but other tools are also available to compare libraries. These approaches were evaluated and can provide complementary information (Schloss, 2008). Second, diversity comparisons require clear OTU definition. Often microbial 'species' are defined by a cut off of percent genetic similarity but there is a lack of consensus and authors use different thresholds to define the OTUs. Here we used several cut-off values and the existence of the core group is relevant independently of the choice of a distance. Finally, this statistical study was carried out on almost full-length 16S rRNA gene sequences. But with the emergence of the new sequencing techniques such as pyrosequencing that generates a large number of short sequences, the application of this type of analysis is questionable. The estimations of total richness can be much more informative with large libraries generated with pyrosequencing as has been done recently on soils (Pelletier *et al.*, 2008).

712

However, in the comparison of ecosystems, the complete sequencing of the 16S rRNA gene allows a more detailed comparison of the sequences at the nucleotide level. Briefly, the statistical tools that we used were efficient in showing the shared phylotypes among the different libraries. It can be hypothesized that this core group of shared OTUs constitutes the basic populations for an anaerobic digester. Then, the specific OTUs may be populations especially adapted to the treated effluent and could be responsible for the differences of efficiency. To clarify this point, future analysis should be performed to follow in parallel the microbial diversity and other physicochemical parameters such as effluent composition, gas production and composition.

Concerning *Archaea* populations, in 2004, Leclerc *et al.* defined 23 'molecular species' occupying the ecological niche of methanogenesis in digesters. In our study, we found a total of 69 different OTUs based on 97% identity on almost full-length 16S rRNA genes. Among them, seven OTUs were found to be dominant. It would be reductive to speak of a core group of OTUs that would be present in each digester and in constant proportions. Depending on the availability and the concentration of the substrates, some OTUs of the seven may become more abundant.

The acetoclastic pathway is mainly represented by *Methanosaeta* genus. We also highlighted the abundance of the lineage Arc I which is probably in competition with *Methanosaeta* species. This implies that Arc I could be an acetate consumer and therefore a major actor in acetoclastic methanogenesis. Previously it was demonstrated that this lineage can grow on formate and H_2/CO_2 (Chouari *et al.*, 2005a). New cultivation experiments are being conducted to explore the ability of Arc I to grow on acetate. Our results are consistent with the previous studies of *Archaea* diversity in anaerobic digesters. A review of the principal studies on *Archaea* populations (Sekiguchi and Kamagata, 2004) shows that *Methanosarcinales* constitute more than 29% of the sequences in all the studies. Sequences affiliated with *Methanosaeta* sp. are most frequently retrieved from methanogenic processes.

The hydrogenotrophic pathway is also represented by *Methanomicrobiales* but in smaller proportions. This is consistent with the review of Sekiguchi *et al.*, which gives proportions for *Methanomicrobiales* in a range of less than 1–29%.

Similar to other studies, the sequences retrieved were affiliated with cultivated species. The only exception is lineage Arc I, which seems to be a lineage detected in abundance only in some anaerobic digesters. The conditions necessary for their development are still unclear.

For *Bacteria* populations, the phyla found in the core group are the less diversified and are affiliated with *Betaproteobacteria*, *Chloroflexi*, *Bacteroidetes* and *Synergistetes*. Other phyla such as *Firmicutes*,

Alpha and *Deltaproteobacteria* are composed of a majority of sequences specific of one digester. These results are consistent with the previous studies carried out on anaerobic digesters (Godon *et al.*, 1997a, b; Chouari *et al.*, 2005a) and provide complementary information about the structure of the community.

We defined the core group α as OTUs commonly found in anaerobic digesters. To determine if our core group had been found in other studies, these OTUs were compared by BLAST to sequences present in public databases and retrieved from anaerobic digesters. Several relatives were found for each OTU of the core group with a percentage of identity above 99% in the sequences published by Chouari *et al.* Relatives with more than 98% of identity were also found only for the *Synergistetes* OTU IV in a study on the diversity of *Synergistetes* in different ecosystems (Godon *et al.*, 2005). So it seems that our core group is not represented in other studies except in the digester of Evry (Chouari *et al.*, 2005a). This result is not really surprising and is directly due to the kind of studies performed so far on anaerobic digesters. First, only a small number of sequences in the databases were exploitable for the BLAST comparison. Most of the sequences were shorter than 500 bp, which is not enough for an accurate comparison with our sequences. Then, most of the sequences come from lab-scale reactors treating synthetic wastewaters or studies on the degradation of organic solid wastes. To our knowledge, since the publications of Chouari *et al.*, no other 16S rRNA gene libraries have been published on full-scale anaerobic digesters treating municipal wastewater sludge, which limits the possible comparison with other data. Nevertheless, the OTUs of the core group were found in smaller libraries built from a large variety of sludge from anaerobic digesters (unpublished data). Out of the 31 samples analysed, OTU I, II, III, IV, V and VI were found in 8, 8, 20, 29, 17 and 26 samples, respectively. These results seem to indicate that the core group α is an important component of the populations involved in anaerobic digestion in full-scale digesters treating wastewater sludge. In other types of processes or effluents, the populations encountered can be close to each other but remain different at the OTU level.

In this study, it was chosen to privilege the sequencing of a large number of clones for each sample collected to increase our knowledge on the diversity of this ecosystem. These data could be useful to design DNA phylochips that could replace the laborious and time-consuming construction of gene libraries. This method will be suited for continuous following of the dynamic variation of the diversity of anaerobic digesters.

Now that the diversity of the anaerobic digester ecosystem has been explored more intensively, the next step is to determine the metabolic functions of these organisms. The development of techniques to

follow the incorporation of varieties of substrates such as stable isotope probing (Radajewski *et al.*, 2000) or MAR-FISH (Kindaichi *et al.*, 2004) are promising and should allow a better comprehension of the metabolic capacities of these organisms.

Another perspective for this work is to quantify the populations encountered. The large database of 16S rDNA sequences that we constituted is a starting point for future analysis. Until then, few sequences were available from anaerobic digesters. To our knowledge, the largest libraries produced to date contained only a maximum of 500 sequences. In our study, we enriched the databases with more than 8000 *Bacteria* sequences and 1500 *Archaea* sequences. From these data, specific probes for major phylogenetic groups were designed to target the anaerobic digester populations specifically. Indeed, the probes commonly found in the literature are far from covering the large diversity of anaerobic digesters as noticed by Ariesyady *et al.* Using a set of probes targeting the 16S and the 23S rDNA, only 67% of the bacterial populations were covered by FISH probes. With our new set of probes, quantitative analysis by Dot-Blot hybridization on 16S rRNA are in progress on the seven digesters. The results will help in the determination of the phylogenetic groups active in the anaerobic degradation of organic matter.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Ariesyady HD, Ito T, Okabe S. (2007). Functional bacterial and archaeal community structures of major trophic groups in a full-scale anaerobic sludge digester. *Water Res* **41**: 1554–1568.

Asshelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.

Chao A. (1984). Non-parametric estimation of the number of classes in a population. *Scand J Stat* **11**: 265–270.

Chao A, Yang MCK. (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80**: 193–201.

Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A. (2003). Molecular evidence for novel planctomycete diversity in a municipal wastewater treatment plant. *Appl Environ Microbiol* **69**: 7354–7363.

Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A. (2005a). Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol* **7**: 1104–1115.

Chouari R, Le Paslier D, Dauga C, Daegelen P, Weissenbach J, Sghir A. (2005b). Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl Environ Microbiol* **71**: 2145–2153.

DeLong EF. (1992). Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89**: 5685–5689.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Excoffier L, Smouse PE, Quattro JM. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Felsenstein J, Churchill GA. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol Biol Evol* **13**: 93–104.

Ficker M, Krastel K, Orlicky S, Edwards E. (1999). Molecular characterization of a toluene-degrading methanogenic consortium. *Appl Environ Microbiol* **65**: 5576–5585.

Godon JJ, Moriniere J, Moletta M, Gaillac M, Bru V, Delgenes JP. (2005). Rarity associated with specific ecological niches in the bacterial world: the 'Synergistes' example. *Environ Microbiol* **7**: 213–224.

Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. (1997a). Microbial 16S rDNA diversity in an anaerobic digester. *Water Sci Technol* **36**: 49–55.

Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. (1997b). Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol* **63**: 2802–2813.

Good IJ. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.

Guermazi S, Daegelen P, Dauga C, Riviere D, Bouchez T, Godon JJ *et al.* (2008). Discovery and characterization of a new bacterial candidate division by an anaerobic sludge digester metagenomic approach. *Environ Microbiol* **10**: 2111–2123.

Hicks RE, Amann RI, Stahl DA. (1992). Dual staining of natural bacterioplankton with 4′,6-diamidino-2-phenylindole and fluorescent oligonucleotide probes targeting kingdom-level 16S rRNA sequences. *Appl Environ Microbiol* **58**: 2158–2163.

Howe K, Bateman A, Durbin R. (2002). QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* **18**: 1546–1547.

Juretschko S, Timmermann G, Schmid M, Schleifer KH, Pommerening-Roser A, Koops HP *et al.* (1998). Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. *Appl Environ Microbiol* **64**: 3042–3051.

714

Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Kindaichi T, Ito T, Okabe S. (2004). Ecophysiological interaction between nitrifying bacteria and heterotrophic bacteria in autotrophic nitrifying biofilms as determined by microautoradiography-fluorescence *in situ* hybridization. *Appl Environ Microbiol* **70**: 1641–1650.

Leclerc M, Delgènes J-P, Godon J-J. (2004). Diversity of the archael community in 44 anaerobic digesters as determined by single strand conformation polymorphism analysis and 16S rDNA sequencing. *Environ Microbiol* **6**: 809–819.

Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA* **102**: 11070–11075.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Kumar Y *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Maidak BL, Larsen N, McCaughey MJ, Overbeek R, Olsen GJ, Fogel K *et al.* (1994). The Ribosomal Database Project. *Nucleic Acids Res* **22**: 3485–3487.

Pelletier E, Kreimeyer A, Bocs S, Rouy Z, Gyapay G, Chouari R *et al.* (2008). 'Candidatus cloacamonas acidaminovorans': genome sequence reconstruction provides a first glimpse of a new bacterial division. *J Bacteriol* **190**: 2572–2579.

Radajewski S, Ineson P, Parekh NR, Murrell JC. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.

Sakamoto M, Umeda M, Ishikawa I, Benno Y. (2000). Comparison of the oral bacterial flora in saliva from a healthy subject and two periodontitis patients by sequence analysis of 16S rDNA libraries. *Microbiol Immunol* **44**: 643–652.

Schloss PD. (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME J* **2**: 265–275.

Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.

Schloss PD, Handelsman J. (2006a). Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol* **72**: 6773–6779.

Schloss PD, Handelsman J. (2006b). Introducing Tree-Climber, a test to compare microbial community structures. *Appl Environ Microbiol* **72**: 2379–2384.

Schloss PD, Larget BR, Handelsman J. (2004). Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* **70**: 5485–5492.

Schlotelburg C, von Wintzingerode F, Hauck R, Hegemann W, Gobel UB. (2000). Bacteria of an anaerobic 1,2-dichloropropane-dechlorinating mixed culture are phylogenetically related to those of other anaerobic dechlorinating consortia. *Int J Syst Evol Microbiol* **50**(Pt 4): 1505–1511.

Sekiguchi Y, Kamagata Y. (2004). Strict and facultative anaerobes: medical and environmental aspects. In: Nakano MM, Zuber P (eds). Microbial Community Structure and Functions in Methane Fermentation Technology for Wastewater Treatment. Horizon Bioscience: UK, pp 361–384.

Sekiguchi Y, Kamagata Y, Syutsubo K, Ohashi A, Harada H, Nakamura K. (1998). Phylogenetic diversity of mesophilic and thermophilic granular sludges determined by 16S rRNA gene analysis. *Microbiology* **144**: 2655–2665.

Sekiguchi Y, Takahashi H, Kamagata Y, Ohashi A, Harada H. (2001). *In situ* detection, isolation, and physiological properties of a thin filamentous microorganism abundant in methanogenic granular sludges: a novel isolate affiliated with a clone cluster, the green non-sulfur bacteria, subdivision I. *Appl Environ Microbiol* **67**: 5740–5749.

Stach JEM, Maldonado LA, Masson DG, Ward AC, Goodfellow M, Bull AT. (2003). Statistical Approaches for Estimating Actinobacterial Diversity in Marine Sediments. *Appl Environ Microbiol* **69**: 6189–6200.

Vartoukian SR, Palmer RM, Wade WG. (2007). The division 'Synergistes'. *Anaerobe* **13**: 99–106.

Zehnder AJB. (1988). *Biology of Anaerobic Microorganisms*. John Wiley & Sons, Inc.: New York. 872 p.

Zheng D, Alm EW, Stahl DA, Raskin L. (1996). Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Appl Environ Microbiol* **62**: 4504–4513.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)