

SCIENTIFIC REPORTS



OPEN

Terminal restriction fragment length polymorphism is an “old school” reliable technique for swift microbial community screening in anaerobic digestion

Jo De Vrieze¹, Umer Z. Ijaz², Aaron M. Saunders³ & Susanne Theuerl⁴

The microbial community in anaerobic digestion has been analysed through microbial fingerprinting techniques, such as terminal restriction fragment length polymorphism (TRFLP), for decades. In the last decade, high-throughput 16S rRNA gene amplicon sequencing has replaced these techniques, but the time-consuming and complex nature of high-throughput techniques is a potential bottleneck for full-scale anaerobic digestion application, when monitoring community dynamics. Here, the bacterial and archaeal TRFLP profiles were compared with 16S rRNA gene amplicon profiles (Illumina platform) of 25 full-scale anaerobic digestion plants. The α -diversity analysis revealed a higher richness based on Illumina data, compared with the TRFLP data. This coincided with a clear difference in community organisation, Pareto distribution, and co-occurrence network statistics, *i.e.*, betweenness centrality and normalised degree. The β -diversity analysis showed a similar clustering profile for the Illumina, bacterial TRFLP and archaeal TRFLP data, based on different distance measures and independent of phylogenetic identification, with pH and temperature as the two key operational parameters determining microbial community composition. The combined knowledge of temporal dynamics and projected clustering in the β -diversity profile, based on the TRFLP data, distinctly showed that TRFLP is a reliable technique for swift microbial community dynamics screening in full-scale anaerobic digestion plants.

Anaerobic digestion (AD) can be considered the first microbial technologies for organic waste treatment to reach full-scale application, and it has been implemented extensively treating various waste streams since the 1980s^{1–3}. Its widespread application is the consequence of (1) its ability to combine organic waste treatment with energy recovery, compared to other treatment technologies, and (2) the potential for process monitoring through, if desirable, on-line monitoring of biogas yield and composition, pH, alkalinity and volatile fatty acids (VFA) concentrations^{4,5}. The appearance of new biomass streams, due to the integration of AD into bioeconomic production systems for food, feed, bioenergy and biomaterials, especially due to the need for energy-neutral and renewable production processes, enforces the efforts to maximize AD process efficiency^{6,7}. This requires much finer control of the AD process, due to its complex character, to ensure the central role of AD in the bio-based economy^{8–10}. The development of alternative process control strategies and early-warning indicators for instabilities enables more direct process control^{11–14}. Nonetheless these indicators do not directly reflect microbial community dynamics and functioning.

In the last decades, simple but robust culture-independent techniques, such as denaturing gradient gel electrophoresis^{15–17}, fluorescent *in situ* hybridisation^{18,19} and terminal restriction length polymorphism (TRFLP)^{20–22}

¹Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, B-9000, Ghent, Belgium.

²Infrastructure and Environment Research Division, School of Engineering, University of Glasgow, Glasgow, UK.

³Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Sohngardsholmsvej 49, 9000, Aalborg, Denmark. ⁴Leibniz Institute for Agricultural Engineering and Bioeconomy e.V. (ATB), Department Bioengineering, Max-Eyth-Allee 100, D-14469, Potsdam, Germany. Correspondence and requests for materials should be addressed to J.D.V. (email: Jo.DeVrieze@UGent.be)

have been used to study the microbial community organisation in AD. Recently, more powerful tools have become available that enable to generate detailed information about the genomic structure and gene expression, which reflects the putative and actual microbial metabolism, but also allow for analysing complex environments and their harbouring microbiomes^{23–28}. The application of 16S rRNA gene amplicon sequencing to unravel the microbial community “black box” led to numerous discoveries of previously unknown microbial life, *e.g.*, in AD. The clear description of the different AD process stages²⁹, as well as the identification of the key species involved in the different steps of the process^{30–32} served as a general basis for process engineering. The overall sensitivity of the methanogenic community and the clear difference in metabolic and physiologic properties among different methanogens denoted methanogenesis as one of the crucial steps with respect to process control^{33,34}. Higher microbial community dynamics were observed in response to changing conditions and/or (partial) inhibition of methanogenesis^{35–39}. Such information can be used to relate microbial community dynamics with AD functional stability to predict and anticipate to process failure^{40,41}, and to determine the main factors that drive the microbial community composition and organization⁴².

The application of next generation amplicon sequencing to make an estimation of microbial community stability in full-scale digesters poses issues in terms of (1) the operational (consumables) and capital (hardware) costs, and (2) the complexity and computationally demanding nature of data analysis. Especially the complexity of the data analysis is an important bottleneck with respect to full-scale application, as robustness and simplicity are considered crucial for frequent and fast application⁸. For fast and simple screening of the microbial community, the extra expense of amplicon sequencing, compared with TRFLP, may be unnecessary^{23,43,44}. The key issue is that simple fingerprinting methods, such as TRFLP, may still capture the important population changes that are critical for process control. The key drawback of TRFLP analysis, *i.e.*, the limited phylogenetic identification depth, due to the short read lengths following the restriction reaction⁴³, does not prevent the determination of microbial community diversity and organisation through various methods⁴⁵. Comparison of TRFLP with next generation amplicon sequencing based on the 16S rRNA gene to screen the microbial community showed the potential of TRFLP for “fast” community screening of the nasopharyngeal microbial community⁴⁶. Comparable results between TRFLP and next generation sequencing methods (*via* the 454 and Illumina and Ion Torrent platforms) were also obtained for the microbial community in the rumen of sheep^{47,48}, the gut of broiler chickens⁴⁹, polar soils⁵⁰, and aquifers⁵¹. The combination of next generation amplicon sequencing and TRFLP analysis in AD ecosystems, thus far, remained restricted to lab-scale experiments^{52,53} or only a limited number of full-scale plants⁵⁴. Hence, even though these studies yielded interesting results, the observed differences between the TRFLP and next generation amplicon sequencing microbial community profile in AD requires further investigation. This will allow to determine the different levels of community profiling for which TRFLP can serve as a *proxy* of next generation amplicon sequencing in AD.

The objective of this research was to investigate the microbial community in various full-scale AD plants using both TRFLP analysis and amplicon sequencing (Illumina platform) of the 16S rRNA gene to determine the potential of TRFLP for a reliable capture of a useful ecological picture of the AD microbiome. A comparison was made based on (1) alpha-diversity, (2) beta-diversity, (3) community organization, (4) the impact of operational parameters on the community profile, and (5) co-occurrence network formation to estimate the potential of TRFLP analysis on different levels.

Material and Methods

Sample collection and storage. A total of 33 unique anaerobic digester content samples were collected from 25 full-scale agricultural and industrial AD plants. Samples from the same plant were given the same name, but a different number, which represents different sampling time points. Sample collection and storage was carried out as described by De Vrieze, *et al.*⁴². The pH was measured directly upon arrival in the lab. Samples for total ammonia nitrogen (NH₄⁺ and NH₃), conductivity, volatile solids and total solids measurement were stored at 4 °C. Samples for VFA content analysis and DNA extraction were stored at –20 °C until further analyses. Information concerning the digester capacity, biogas production rate, organic loading rate, sludge retention time, feedstock composition and temperature was obtained directly from the plant operators.

Microbial community analysis. *DNA extraction.* The DNA extraction was carried out with the FastDNA[®] SPIN Kit for Soil (MP Biomedicals, Solon, OH, USA), starting from 200 mg of digester content, in accordance with the instructions of the manufacturer. The quality of the DNA was validated with 1% agarose gel electrophoresis and PCR analysis to make sure that there were no PCR inhibiting components present in the DNA extracts. The control PCR amplification was performed according to the protocol of Boon, *et al.*⁵⁵, using the bacterial primers P338F and P518R⁵⁶ targeting the 16S rRNA gene. The PCR product quality, *i.e.*, fragment length and the absence of non-desirable DNA fragments, confirmed that there was no PCR inhibition.

Amplicon sequencing and data processing. The DNA extracts were subjected to amplicon sequencing of the variable region 4 (V4) using the primers 515 F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806 R (5′-GGACTACHVGGGTWTCTAAT-3′) on the Illumina HiSeq platform with the protocol of Caporaso, *et al.*⁵⁷. The PCR reaction was carried out in triplicate for each sample, and the PCR mixture and conditions, as well as PCR product purification and quantification were carried out as described by De Vrieze, *et al.*⁴².

The amplicon sequences were trimmed and filtered with Sickle v1.200⁵⁸, using a sliding window approach. Reads with an average quality score below 20 were removed. Error correction was performed with the BayesHammer error correction tool⁵⁹, coupled with the Spades v2.5.0 assembler. The paired-end reads were assembled with PANDAseq v2.4⁶⁰ with a minimum overlap of 20. This three-step approach was used to obtain up to 90% error reduction⁶¹. Taxonomic unit construction was carried out with the UPARSE (v7.0.1001) pipeline⁶². The reads were dereplicated, and sorted by decreasing abundance. The absolute singletons were discarded,

followed by simultaneous chimera filtering and OTU (operational taxonomic unit) clustering, based on 97% similarity. An additional reference-based chimera filtering step was included, using the 'gold' database (<http://drive5.com/uchime/gold.fa>), derived from the ChimeraSlayer reference database in the Broad Microbiome Utilities (<http://microbiomeutil.sourceforge.net>). The standalone RDP Classifier v2.6⁶³ was used to classify representative sequences from the OTUs against the Ribosomal Database Project (RDP) with a confidence threshold of 80%. A table with the abundance of the different OTUs and their taxonomic assignments, and containing each sample was generated, containing re-sequenced samples as individual sample libraries. Samples were sent out for re-sequencing when sampling depth was insufficient (<10,000 reads).

TRFLP analysis. The TRFLP analyses were carried out following the protocols proposed by Rademacher, *et al.*²⁰ and Klang, *et al.*²¹. Briefly, the bacterial and archaeal 16S rRNA genes were amplified (three replicates per crude DNA extract) using the primer pairs 27 F/926MRr (Bacteria) and Ar109f/Ar912r (Archaea), whereby the forward primers were fluorescently labelled with Cy5. After purification of the PCR products, 150–200 ng was digested with *MspI* and *Hin6I* for the bacterial and with *AluI* for the archaeal assay. The digestion fragments were separated using a GenomeLab™ GeXP Genetic Analysis System (AB SCIEX Germany GmbH, Darmstadt, Germany). The obtained data were pre-analysed using the GeXP analysis software (version 10.2), whereby only profiles were considered for further analyses whose internal standard had a standard deviation of 0.39 nucleotides (nt) or less²⁰. A detailed analysis was performed using the software package BioNumerics 7.6 (Applied Maths, Kortrijk, Belgium), according to Klang, *et al.*²¹.

Statistical analyses. The statistical analyses were carried out in R studio, version 3.3.1 (<http://www.r-project.org>)⁶⁴, using the packages *vegan*⁶⁵ and *phyloseq*⁶⁶ for microbial community analysis. Prior to data processing, repeated measures analysis of variance (ANOVA, *avov* function) was used to confirm that relative abundance profiles of re-sequenced samples were not significantly different after which samples were collated as described by Connelly, *et al.*⁶⁷. All samples were rescaled by taking the proportions of each OTU, multiplying it with the minimum sample size, and rounding to the nearest integer⁶⁸. Only OTUs with a relative abundance $\geq 1\%$ were considered both for the Illumina and TRFLP data to allow accurate and direct comparison between the two methods. Significant differences in the order-based Hill's numbers⁶⁹ reflecting richness (number of OTUs or TRFs, H_0), the exponential of the Shannon diversity index (H_1) and the Inverse Simpson index (H_2), between the amplicon and TRFLP data were determined with ANOVA. Spearman's Rank correlations between the Hill numbers obtained from the amplicon and TRFLP data were determined using the *ddply* function (*plyr* package). The non-metric multidimensional scaling (NMDS) plots were constructed based on the Bray-Curtis⁷⁰, Chao⁷¹, Jaccard, Kulczynski⁷², and Mountford⁷³ distance measures. Correlations between the amplicon and TRFLP distance measures were determined using the Spearman's and Kendall Rank correlation. Permutational ANOVA (PERMANOVA) (9999 permutations) with Bonferroni correction was used to determine significant differences in community composition between different clusters, and to evaluate the influence of the operational parameters on the amplicon and TRFLP data profiles using the *adonis* function (*vegan*). A canonical correspondence analysis (CCA) model, using the *envfit* function (*vegan*) was used to evaluate the strength of the correlation of the operational parameters (Table S1). Operational parameters with a significant impact on the microbial community profile were determined through PERMANOVA analysis (9999 permutations), and visualised with canonical correspondence analysis plotting. The community organization (Co) was calculated based on the Lorenz distribution curves^{45,74}, and the Pareto value was calculated as the total relative abundance of the 20% most abundant OTUs or TRFs^{75,76}. Significant differences between the amplicon and TRFLP data and Spearman's Rank correlations were determined, similar as for the Hill numbers. Co-occurrence networks were constructed, based on the Spearman's rank correlation, with the *igraph* (<http://igraph.org>), *sna*⁷⁷ and *network*⁷⁸ packages, and the network statistics betweenness centrality and normalised degree were calculated⁷⁹.

Analytical techniques. Total solids (TS), volatile solids (VS) and total ammonia nitrogen (NH_4^+ and NH_3) were determined via Standard Methods⁸⁰. The total ammonia nitrogen, pH and temperature were used to calculate the free ammonia (NH_3) values⁸¹. The different VFA concentrations were measured through gas chromatography, as described in SI (S1). Conductivity was measured with a C833 conductivity meter (Consort, Turnhout, Belgium), and pH was determined using a C532 pH meter (Consort, Turnhout, Belgium).

Results

Operational data of the full-scale anaerobic digesters. Samples were collected from the 25 different full-scale digesters, with a total volume between 1000–4000 m³, during a period of steady state feedstock supply and stable biogas production (<10% variation) of at least one year. The different digesters had a wide range of feedstock compositions, differences in sludge retention time (18–124 days), organic loading rate (1.5–13.8 kg COD m⁻³ d⁻¹) and biogas production (1.1–7.5 m³ m⁻³ d⁻¹) (Table S1). This corresponded with different pH (7.10–8.52), total ammonia nitrogen (130–6400 mg N L⁻¹), free ammonia nitrogen (1.7–1460 mg N L⁻¹), and conductivity (6.4–62.3 mS cm⁻¹) ranges, and total VFA concentrations of up to 36.8 g COD L⁻¹. Most samples (62%) were collected from mesophilic digesters (33–38 °C), while one sample was collected from a digester operating at 42 °C, and the other samples (35%) were collected from thermophilic digesters (50–55 °C).

Microbial community profiling. Amplicon sequencing of the V4 region of the 16S rRNA gene yielded an average of 40,731 ± 16,647 reads from in total 4,027 different OTUs, with on average of 703 ± 287 OTUs per sample. No significant differences ($P < 0.05$) were observed between samples that were sequenced multiple times. Rarefaction curves indicate that saturation was not reached for all samples (Figure S1), but to allow comparison between the Illumina and TRFLP data, only OTUs with a relative abundance $\geq 1\%$ in at least one sample were

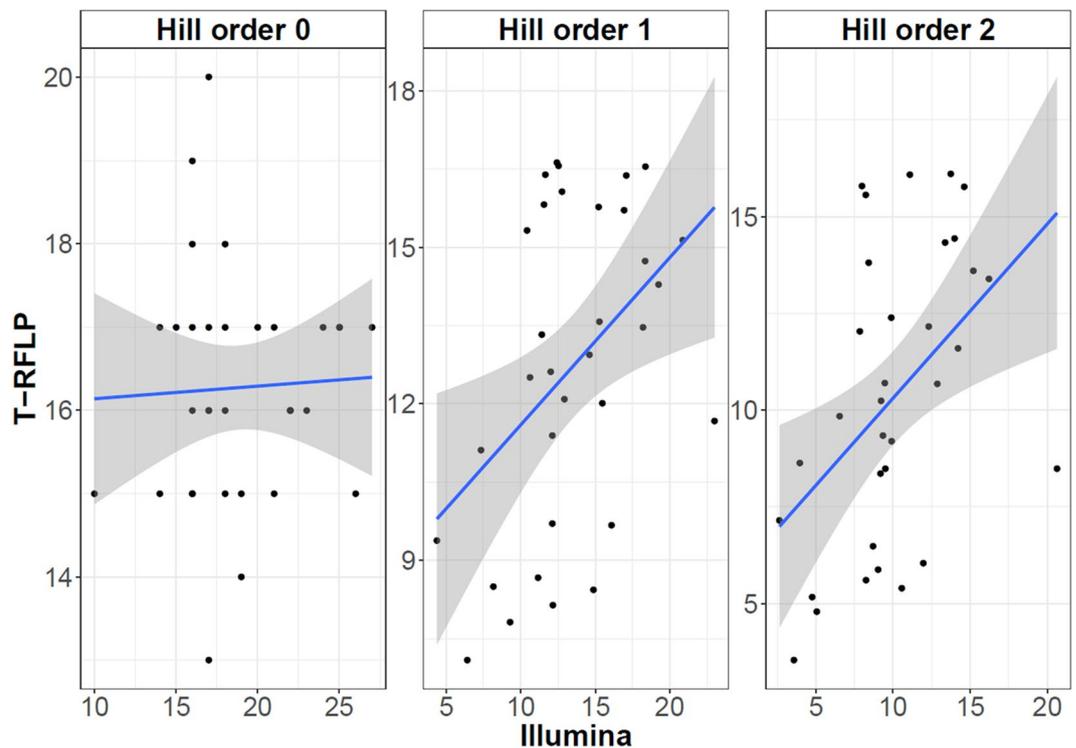


Figure 1. Spearman's Rank correlation for the H_0 , H_1 and H_2 Hill numbers. The richness (H_0), exponential of the Shannon diversity index (H_1) and the Inverse Simpson index (H_2) are included. Correlation between the Illumina and TRFLP data for H_0 ($P=0.87$) was not significant, yet, for H_1 ($\rho=0.37$ & $P=0.033$) and H_2 ($\rho=0.42$ & $P=0.015$) a significant positive correlation was observed. The grey zone represents the 95% confidence interval.

considered. Hence, this did not influence further data processing and interpretation. The 1% cut-off value resulted in total in only 241 different OTUs being retained for further analysis. Using a cut-off value of 1% relative abundance in at least one sample in the Illumina data resulted in an elimination of all archaeal OTUs.

The TRFLP community profiling yielded a total of 142 different bacterial and 12 different archaeal terminal restriction fragments (TRFs), with an average of 16 ± 1 and 4 ± 1 bacterial and archaeal TRFs per sample, respectively. Following the application of the 1% cut-off value, the total number of bacterial TRFs decreased to 120, while no archaeal TRFs were removed. All further analyses were carried out using the datasets on which the 1% cut-off was applied.

Microbial community diversity and organization. Spearman's Rank correlation analysis of the microbial diversity parameters, determined via the Hill's numbers, revealed that there was no significant correlation between the Illumina and bacterial TRFLP data for H_0 ($P=0.87$), but for H_1 ($\rho=0.37$ & $P=0.033$) and H_2 ($\rho=0.42$ & $P=0.015$) a significant positive correlation was observed (Fig. 1). A significantly higher richness (H_0) value ($P=0.0007$) was observed for the Illumina based analysis than for the bacterial TRFLP data (Figure S2). No significant difference was, however, observed between the Illumina and bacterial TRFLP data for the H_1 ($P=0.39$) and H_2 ($P=0.78$) Hill's numbers. The archaeal TRFLP data showed significantly ($P < 0.0001$) lower H_0 , H_1 and H_2 values compared with the bacterial TRFLP, which was a consequence of the fact that, on average, only 4 ± 1 TRFs were observed per sample for the archaeal TRFLP (Figure S2).

A significant positive correlation could be observed between the Illumina and bacterial TRFLP data, both for the community organisation (Co) ($\rho=0.45$ & $P=0.009$) and Pareto ($\rho=0.56$ & $P=0.0009$) values (Fig. 2). Community organisation analysis showed a significantly higher ($P=0.0002$) average Co value of 59.4 ± 12.2 for the Illumina than for the bacterial TRFLP data (46.6 ± 13.6), indicating a more evenly distributed community based on the bacterial TRFLP data (Fig. 2). The community organisation based on the archaeal TRFLP data was significantly higher ($P < 0.0001$) than based on the bacterial TRFLP data (Figure S3). A similar significant difference ($P=0.038$) was observed between the Illumina and bacterial TRFLP data for the Pareto values, which confirms the apparent higher evenness based on the bacterial TRFLP data. The average Pareto value was significantly higher ($P < 0.0001$) for that archaeal TRFLP than for the bacterial TRFLP data, but this was a consequence of the skewed archaeal TRFLP Pareto profile, related to the low TRFs number.

Beta-diversity analysis. For both Illumina and bacterial TRFLP data, the NMDS analysis of the Bray-Curtis dissimilarity matrix revealed the presence of four apparent clusters (Fig. 3). Cluster 1 contained samples from sludge digesters (Den, Oss1, Oss2, and Oss3), characterized by the lowest values of the operational parameters (Table S1), while cluster 2 contained samples from upflow anaerobic sludge bed reactors treating potato or paper

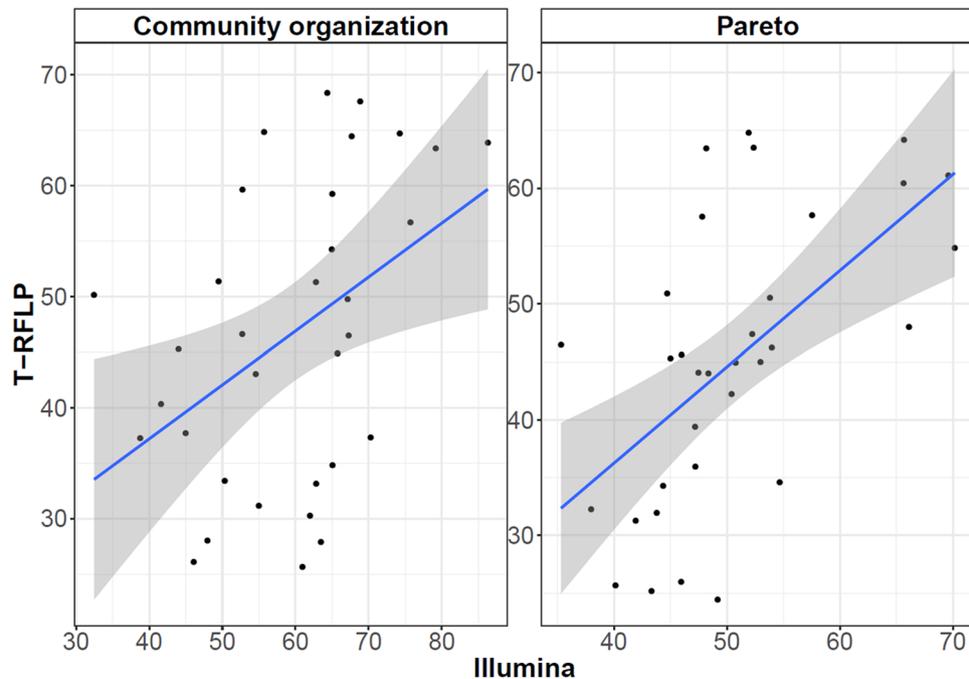


Figure 2. Spearman's Rank correlation for the community organisation (Co) and Pareto values. The Co was calculated based on the Lorenz distribution curves^{45,74}, and the Pareto value was calculated as the total relative abundance of the 20% most abundant OTUs or TRFs^{75,76}. A significant positive correlation between the Illumina and TRFLP data was observed for Co ($\rho = 0.45$ & $P = 0.009$) and Pareto ($\rho = 0.56$ & $P = 0.0009$). The grey zone represents the 95% confidence interval.

mill wastewater (Clar, Myd and VPK). Cluster 3 contained mainly samples from mesophilic digesters treating bio- and agricultural wastes (BCI, SMA, SHA, BAT, BIF, Agri, CAZ, EcP, VCE), and cluster 4 samples from thermophilic digesters (GFT, AGR, DRZ, BBy, BIE), although the mesophilic SHE digester was also included in cluster 4. A significant correlation could be observed between the Bray-Curtis dissimilarity matrix for the Spearman's ($\rho = 0.17$ & $P = 0.0001$) and Kendall ($\tau = 0.12$ & $P = 0.0002$). The PERMANOVA analysis confirmed that clusters 2, 3 & 4 were significantly separated ($P < 0.05$), based on the Illumina and bacterial TRFLP data, but that cluster 1 was not ($P = 0.17$ for Illumina and $P = 0.17$ for bacterial TRFLP data). The same pattern of clustering was detected for data obtained using both the Illumina and bacterial TRFLP methods. Significant differences between the clusters for the archaeal TRFLP data could be observed only between cluster 1 and 3 ($P = 0.006$), cluster 2 and 3 ($P = 0.040$) and cluster 3 and 4 ($P = 0.0018$). A similar observation was made through the NMDS analysis of the Chao, Jaccard, Kulczynski, and Mountford distance measures (Figure S4).

Canonical correspondence analysis demonstrated that the pH and temperature had a significant impact on the Illumina ($P_{\text{pH}} = 0.0001$, $P_{\text{temp}} = 0.0001$), the bacterial TRFLP ($P_{\text{pH}} = 0.001$, $P_{\text{temp}} = 0.002$), as well as the archaeal TRFLP ($P_{\text{pH}} = 0.002$, $P_{\text{temp}} = 0.002$) profiles (Fig. 4, Table S2). The total ammonia concentration, however, had a significant impact for the bacterial TRFLP data ($P = 0.007$), but not for the Illumina ($P = 0.064$) and archaeal TRFLP data ($P = 0.11$). None of the other operational parameters had a significant impact on any of the three community profiles.

Co-occurrence network analysis. Co-occurrence networks were constructed for the Illumina, bacterial TRFLP and archaeal TRFLP profiles to determine to which extent inter-taxa correlations and related network statistics are reflected in a similar manner. No significant ($P < 0.001$) correlations were observed for the archaeal TRFs probably due to the paucity of the data, thus, no additional analyses were carried out. A clear inter-taxa correlation could be observed in the co-occurrence network based on the Illumina data, while this was not the case for the bacterial TRFLP data, which resulted in a much more fragmented profile (Fig. 5). In total 1,494 significant correlations ($P < 0.001$) existed between the 241 OTUs recorded within the Illumina dataset, representing 5.2% of total potential correlations. The 120 bacterial TRFs contained 245 significant correlations, which corresponded with 3.4% of total potential correlations. Only 4 OTUs (1.7%) had no significant correlations, while 20 TRFs (16.7%) had no significant correlations.

Betweenness centrality, which reflects the importance of an OTU or TRF as connector in the network⁷⁹, was significantly higher ($P < 0.0001$) for the Illumina profile, compared with the bacterial TRFLP profile (Fig. 6). Out of 237 OTUs with at least one significant correlation, 86.9% had a betweenness score > 0 , while out of 100 TRFs with at least one significant correlation, only 47.0% had a betweenness score > 0 . The normalised degree, which reflects the proportion of other OTUs/TRFs a certain OTU/TRF interacts with⁷⁹, showed a similar significantly higher ($P < 0.0001$) value for the Illumina network compared with the bacterial TRFLP network.

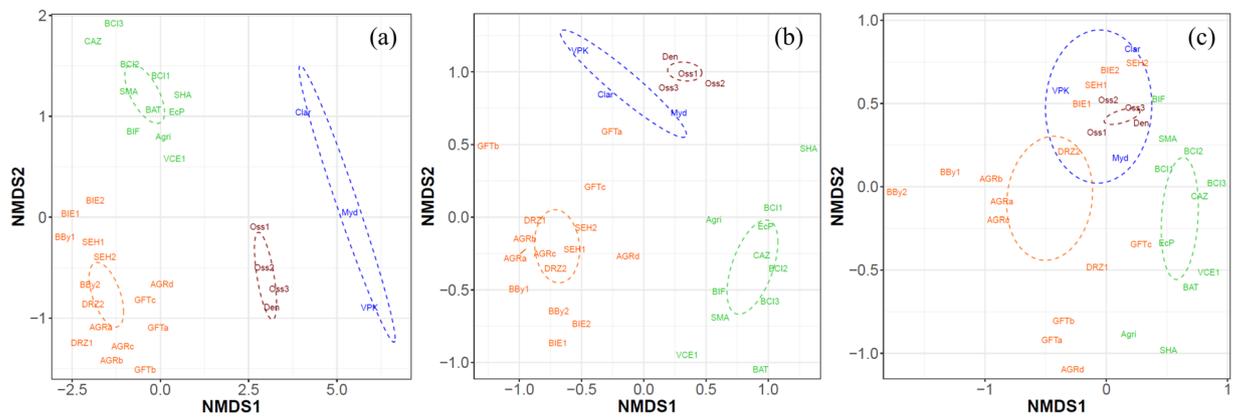


Figure 3. Non-metric distance scaling (NMDS) analysis of the Bray-Curtis dissimilarity distance indices of the (a) Illumina (stress = 0.09), bacterial TRFLP (stress = 0.24), and (c) archaeal TRFLP (stress = 0.17) at OTU/TRF level. The four different clusters Cluster 1 (red), Cluster 2 (blue), Cluster 3 (green), and Cluster 4 (orange) are distinguished, and the ellipses represent the 95% value of the standard error of the average value for each cluster.

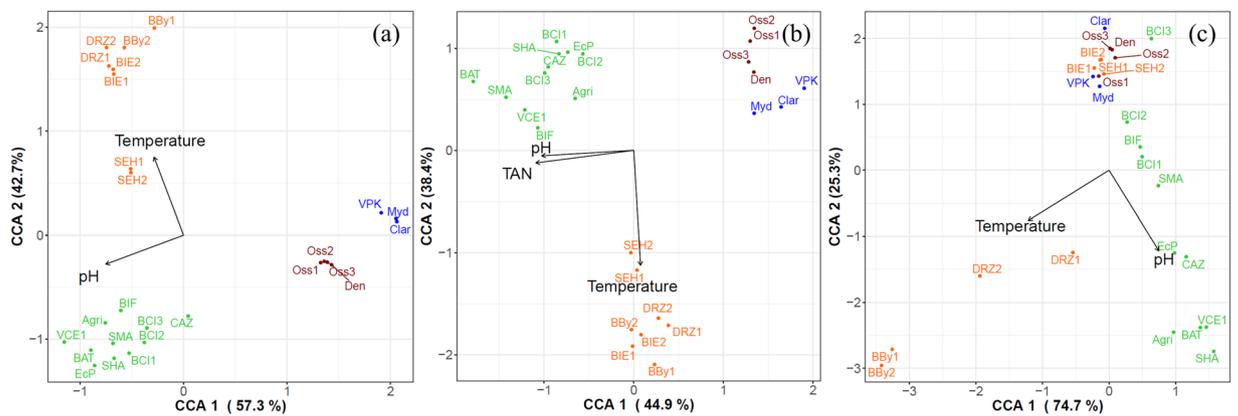


Figure 4. Canonical correspondence analysis of the (a) Illumina, (b) bacterial TRFLP and (c) archaeal TRFLP profile of each sample at OTU/TRF level. The PERMANOVA analysis (9999 permutations) identified the relationship between the diversity and operational parameters on community composition, and significant ($P < 0.05$) correlations are presented by the arrows. The four different clusters Cluster 1 (red), Cluster 2 (blue), Cluster 3 (green), and Cluster 4 (orange) are distinguished. TAN = total ammonia nitrogen.

Discussion

A comparison of the microbiomes in 25 full-scale AD plants was carried out to determine whether TRFLP profiling can be used as a simplified *proxy* of 16S rRNA gene amplicon (Illumina platform) sequencing. While the Illumina and bacterial TRFLP data showed a high degree of similarity at higher-order α -diversity levels, a significant difference in richness was observed. The archaeal TRFLP diversity was significantly lower than the bacterial TRFLP diversity, related to the structured bacteria-archaea succession in stable anaerobic digesters^{82,83}. The β -diversity analysis revealed a similar pattern for the Illumina, bacterial TRFLP and archaeal TRFLP data. This was, however, not reflected in a similar co-occurrence pattern, related to significant differences in betweenness centrality, normalised degree, community organisation and Pareto distribution between the Illumina and bacterial TRFLP data.

The (non)sense of richness in anaerobic digestion. Microbial community richness has been one of the most commonly used parameters to estimate taxonomic diversity and the related metabolic potential, carrying physiological/ecological capacity and overall stability of a highly diverse and complex microbial community in different natural or engineered ecosystems^{45,84–88}. Although a high richness does not necessary or directly reflect a high metabolic potential, it has been suggested to increase the resilience and/or resistance of microbial communities⁸⁹, and both parameters contribute to process stability and efficiency in AD^{90–92}.

Microbial community richness, however, highly depends on the method used to determine it, as observed not only in our study, but also in previously published studies^{93,94}. While based on the Illumina data an average richness of 19 ± 4 OTUs was observed per sample, this was 16 ± 1 TRFs for the bacterial TRFLP data, and only 4 ± 1 TRFs for the archaeal TRFLP data. The low archaeal richness can be explained by the fact that the methanogenic

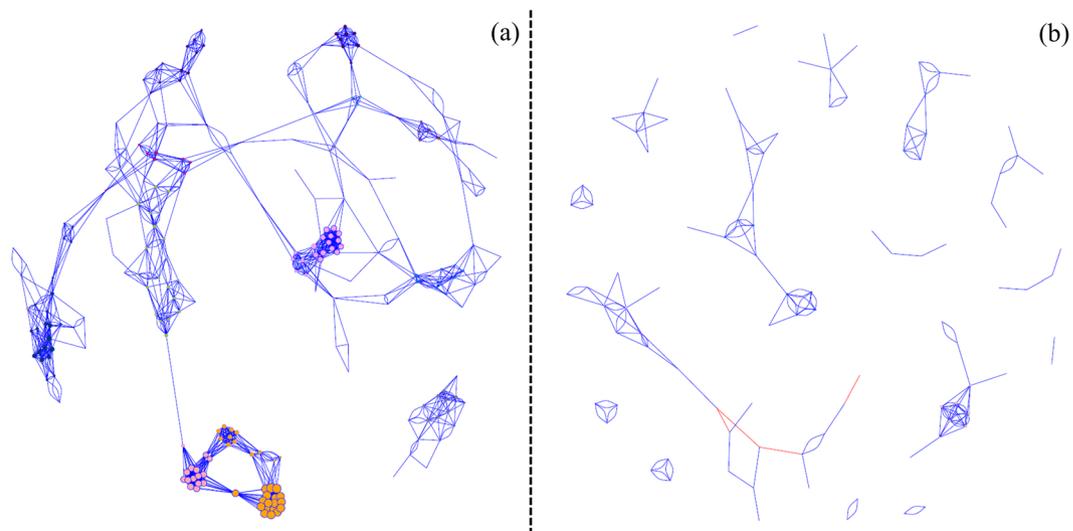


Figure 5. Network of co-occurring (a) OTUs of the Illumina and (b) TRFs of the bacterial data, based on the Spearman correlation analysis. A connection stands for a strong positive (blue) or negative (red) significant (Spearman's $\rho > 0.5$, $P < 0.001$) correlation. The size of each node is proportional to the number of connections (Normalised degree).

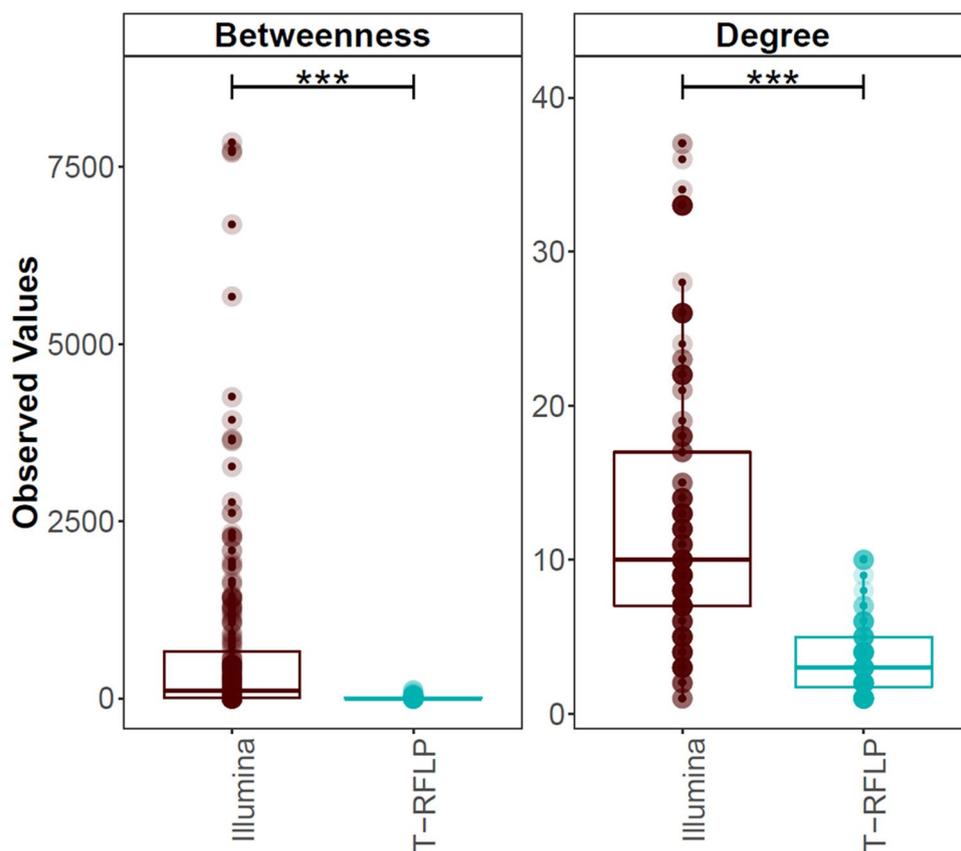


Figure 6. Betweenness centrality and normalised degree, calculated from the co-occurrence network correlations for the Illumina and bacterial TRFLP data. *** $P < 0.001$.

archaea only contribute to the final stage of the AD process, using a narrow substrate spectra, hence, representing a low (relative) abundance, often $< 5\%$ ²⁴. This is also reflected in the fact that using a cut-off value of 1% relative abundance in at least one sample in the Illumina data resulted in an elimination of all archaeal OTUs. Inherently, this coincides with a low archaeal richness.

The significant difference in richness between the Illumina and bacterial TRFLP data is more striking, as both were normalised via the common-scale method⁶⁸ and subjected to a 1% cut-off value. This raises the question what can be considered as the relevant richness, because “total richness” coverage in AD, as determined *via* rarefaction curves, was not obtained in this study, and is rarely met through 16S rRNA gene amplicon sequencing^{31,42}. The importance of the rare or uncommon taxa remains open for debate^{95,96}, but it is clear that deep sequencing is needed to determine the absolute richness⁹³. Even then, method-inherent choices, such as the clustering method and threshold⁹⁷, will strongly influence absolute richness values. Hence, richness as an indicator of microbial stability is too much dependent on the method used to measure it, and may not be useful to define the diverse microbial ecosystem in AD. It raises the question what depth of resolution is useful to identify parameters for process control based on microbial diversity.

Beta-diversity analysis: fingerprinting technique independent. The method-dependence of richness also affected the β -diversity metrics community organisation (Co), Pareto distribution, and co-occurrence networking and related statistics. Their strong dependency on the richness value, especially at low richness values^{79,98,99} explains the significant differences between the Illumina and bacterial TRFLP data, and warrants careful interpretation. In contrast, no significant difference was recorded between the Illumina and bacterial TRFLP regarding the exponential of the Shannon diversity index (H_1) and the Inverse Simpson index (H_2), whereby the Spearman's Rank analysis revealed a significant positive correlation between both data sets. The high similarity of the β -diversity profiles, based on the Illumina, bacterial TRFLP and even archaeal TRFLP data is apparent. This similarity seems to be distance measure independent, as the Bray-Curtis, Chao, Jaccard, Kulczynski, and Mountford distance measures revealed a similar clustering pattern.

The robustness of β -diversity analysis in AD is supported by the fact that canonical correspondence analysis determined pH and temperature as the two operational parameters with a significant impact on the microbiome in the studied full-scale anaerobic digesters, highlighted *via* the Illumina, bacterial TRFLP and archaeal TRFLP community profiles. The strong impact of pH, mainly related to free ammonia toxicity^{19,100}, and temperature on microbial community composition has been observed in several other studies^{32,42,101}. Although total ammonia nitrogen had a significant impact only on the bacterial TRFLP community profile, its insignificance towards the Illumina community profile needs to be considered with care, given the low P-value of 0.064, which only marginally exceeded the threshold value of 0.05 and high R^2 value of 0.856, based on the CCA model. The robustness of β -diversity, hence, also depends on the direct interpretation of P-values, rather than considering a “significant on/off effect”.

The reliability and reproducibility of β -diversity analysis through different fingerprinting methods has been observed in other ecosystems, and can be applied, if allowed by the method, on different phylogenetic levels¹⁰². Although the bacterial TRFLP data in most cases did not allow comparison at deeper phylogenetic levels (*e.g.*, family or genus), they do provide an unambiguous bacterial fingerprint, independent of the taxonomy assignment. Hence, TRFLP can be used as a robust fingerprinting method for β -diversity analysis in AD, as reflected in its similarity with high-throughput amplicon sequencing based β -diversity profiles.

Towards practical applications: microbial community profiling in AD. The robustness of TRFLP analysis for microbial community β -diversity profiling in AD provides a suitable framework for microbial community screening to describe the dominant ecological diversity, related to the 1% cut-off, reflecting the interactions of the dominant microorganisms among each other and with their environment. While high-throughput 16S rRNA gene amplicon sequencing techniques provide a more in-depth microbial community profile, the data evaluation time constraints seriously hamper application on the full-scale level regarding the high flexibility of the microbiomes against changing environmental factors, and, hence, the need for long-term temporal observation/process control. As functional stability in AD does not necessarily reside on the presence of specific species, rather than an overall high functional redundancy^{92,103}, fingerprinting techniques can be used, alternative to molecular techniques that enable identification of the microbial community members, for AD process stability screening. However, this does not exclude the potential importance of certain “abundant core” OTUs in AD^{104–106}. The accurate definition of boundary conditions with respect to microbial community variation over time and/or in response to disturbances is of crucial importance to guarantee the application of microbial community profiling in AD. As the microbial communities in lab- and full-scale plants displays continuous natural dynamics in its taxonomic, functional and ecological diversity over time^{15,21,83,107,108}, it is necessary to distinguish natural variation from disturbance-related unwanted community shifts that lead to process deterioration. The unique community structure in each AD plant, even considering temporal dynamics^{21,91,107}, related to the feedstock characteristics and operational conditions^{32,109}, is a first important consideration. Second, full-scale plants cluster according to feedstock (Cluster 1), digester type (Cluster 2) and operational conditions, *e.g.*, pH, temperature and free/total ammonia concentration, as also observed in other studies^{32,42,104}. When combining the uniqueness of the AD plant, considering its temporal variation at constant and stable conditions, with its expected place in the full-scale plant clustering or AD-typing, β -diversity analysis based on the TRFLP profile can be used for swift microbial community monitoring of full-scale AD plants. Deviations from the natural temporal dynamics during stable conditions would require a corrective action through operational parameter adjustment by applying balanced microbial diversity management strategies, considering the living demands of the occurring microbiome, or alternative biological strategies, such as targeted bioaugmentation. At transient conditions, *e.g.*, related to changes in feedstock composition or deliberate changes in the operational parameters, a shift of the microbial community outside the basic temporal variation, but within its functional redundancy can be expected. In that case, a new temporal dynamics window should be identified, based on (1) the stabilisation of operational parameters, (2) full-scale digesters with similar feedstocks and/or operational parameters, and (3) its expected place in the clustering or AD-typing.

Conclusions

The comparison of the Illumina amplicon, bacterial TRFLP and archaeal TRFLP profiles of 25 full-scale AD plants revealed high degree of similarity in the β -diversity profiles. This suggests that TRFLP may be easier and cheaper, although this strongly depends on the available facilities, constantly evolving techniques and their related costs, alternative to amplicon sequencing for monitoring changes of the overall structure of the microbial communities in full-scale AD to maintain a stable and efficient biogas production process. Even today, in the era of next generation sequencing, the reliability of TRFLP is not diminishing, particularly for the study of microbial community dynamics in relation to the effects of changing environmental, especially process inhibiting, factors on the taxonomic, the functional, as well as the ecological diversity of AD microbiomes. Briefly, TRFLP is not at all obsolete, and will remain of high value in the field of microbial ecology.

Data Availability

The raw fastq files that were used to create the OTU table of the 16S rRNA gene amplicon sequencing data (Illumina HiSeq platform) have been deposited in the National Center for Biotechnology Information (NCBI) database (Accession number SRP132836). The resulting OTU table has been included in the Supplementary Information (Supplementary Table S3), as well as the tables containing the bacterial and archaeal raw TRFs (Supplementary Tables S4 AND S5). All other data generated or analysed during this study are included in this article (and its Supplementary Information files).

References

- Verstraete, W. *et al.* Anaerobic digestion as a core technology in sustainable management of organic matter. *Water Sci. Technol.* **52**, 59–66 (2005).
- Holm-Nielsen, J. B., Al Seadi, T. & Oleskowicz-Popiel, P. The future of anaerobic digestion and biogas utilization. *Bioresour. Technol.* **100**, 5478–5484, <https://doi.org/10.1016/j.biortech.2008.12.046> (2009).
- Appels, L. *et al.* Anaerobic digestion in global bio-energy production: Potential and research challenges. *Renewable & Sustainable Energy Reviews* **15**, 4295–4301, <https://doi.org/10.1016/j.rser.2011.07.121> (2011).
- Bjornsson, L., Murto, M. & Mattiasson, B. Evaluation of parameters for monitoring an anaerobic co-digestion process. *Appl. Microbiol. Biotechnol.* **54**, 844–849 (2000).
- Boe, K., Batstone, D. J., Steyer, J. P. & Angelidaki, I. State indicators for monitoring the anaerobic digestion process. *Water Res.* **44**, 5973–5980, <https://doi.org/10.1016/j.watres.2010.07.043> (2010).
- Ryan, D., Gadd, A., Kavanagh, J. & Barton, G. W. Integrated biorefinery wastewater design. *Chem. Eng. Res. Des.* **87**, 1261–1268, <https://doi.org/10.1016/j.cherd.2009.04.016> (2009).
- Moraes, B. S., Zaiat, M. & Bonomi, A. Anaerobic digestion of vinasse from sugarcane ethanol production in Brazil: Challenges and perspectives. *Renewable & Sustainable Energy Reviews* **44**, 888–903, <https://doi.org/10.1016/j.rser.2015.01.023> (2015).
- Madsen, M., Holm-Nielsen, J. B. & Esbensen, K. H. Monitoring of anaerobic digestion processes: A review perspective. *Renewable & Sustainable Energy Reviews* **15**, 3141–3155, <https://doi.org/10.1016/j.rser.2011.04.026> (2011).
- Maity, S. K. Opportunities, recent trends and challenges of integrated biorefinery: Part I. *Renewable & Sustainable Energy Reviews* **43**, 1427–1445, <https://doi.org/10.1016/j.rser.2014.11.092> (2015).
- Batstone, D. J. & Virdis, B. The role of anaerobic digestion in the emerging energy economy. *Curr. Opin. Biotechnol.* **27**, 142–149, <https://doi.org/10.1016/j.copbio.2014.01.013> (2014).
- Kleyböcker, A. *et al.* Comparison of different procedures to stabilize biogas formation after process failure in a thermophilic waste digestion system: Influence of aggregate formation on process stability. *Waste Manage.* **32**, 1122–1130, <https://doi.org/10.1016/j.wasman.2012.01.015> (2012).
- Kleyböcker, A., Liebrich, M., Verstraete, W., Kraume, M. & Wurdemann, H. Early warning indicators for process failure due to organic overloading by rapeseed oil in one-stage continuously stirred tank reactor, sewage sludge and waste digesters. *Bioresour. Technol.* **123**, 534–541, <https://doi.org/10.1016/j.biortech.2012.07.089> (2012).
- Fdez-Guelfo, L. A., Alvarez-Gallego, C., Sales, D. & Romero, L. I. New indirect parameters for interpreting a destabilization episode in an anaerobic reactor. *Chemical Engineering Journal* **180**, 32–38, <https://doi.org/10.1016/j.cej.2011.10.091> (2012).
- Lv, Z. P. *et al.* Stable isotope composition of biogas allows early warning of complete process failure as a result of ammonia inhibition in anaerobic digesters. *Bioresour. Technol.* **167**, 251–259, <https://doi.org/10.1016/j.biortech.2014.06.029> (2014).
- De Vrieze, J., Verstraete, W. & Boon, N. Repeated pulse feeding induces functional stability in anaerobic digestion. *Microb. Biotechnol.* **6**, 414–424, <https://doi.org/10.1111/1751-7915.12025> (2013).
- Bialek, K. *et al.* Quantitative and qualitative analyses of methanogenic community development in high-rate anaerobic bioreactors. *Water Res.* **45**, 1298–1308, <https://doi.org/10.1016/j.watres.2010.10.010> (2011).
- Hwang, K., Shin, S. G., Kim, J. & Hwang, S. Methanogenic profiles by denaturing gradient gel electrophoresis using order-specific primers in anaerobic sludge digestion. *Appl. Microbiol. Biotechnol.* **80**, 269–276, <https://doi.org/10.1007/s00253-008-1544-9> (2008).
- Angenent, L. T., Sung, S. W. & Raskin, L. Formation of granules and *Methanosaeta* fibres in an anaerobic migrating blanket reactor (AMBR). *Environ. Microbiol.* **6**, 315–322, <https://doi.org/10.1111/j.1462-2920.2004.00597.x> (2004).
- Calli, B., Mertoglu, B., Inanc, B. & Yenigun, O. Effects of high free ammonia concentrations on the performances of anaerobic bioreactors. *Process Biochem.* **40**, 1285–1292, <https://doi.org/10.1016/j.procbio.2004.05.008> (2005).
- Rademacher, A., Nolte, C., Schonberg, M. & Klocke, M. Temperature increases from 55 to 75 A degrees C in a two-phase biogas reactor result in fundamental alterations within the bacterial and archaeal community structure. *Appl. Microbiol. Biotechnol.* **96**, 565–576, <https://doi.org/10.1007/s00253-012-4348-x> (2012).
- Klang, J. *et al.* Dynamic variation of the microbial community structure during the long-time mono-fermentation of maize and sugar beet silage. *Microb. Biotechnol.* **8**, 764–775, <https://doi.org/10.1111/1751-7915.12263> (2015).
- Jimenez, J. *et al.* Methanogenic activity optimization using the response surface methodology, during the anaerobic co-digestion of agriculture and industrial wastes. *Microbial community diversity. Biomass Bioenerg.* **71**, 84–97, <https://doi.org/10.1016/j.biombioe.2014.10.023> (2014).
- Cabezas, A. *et al.* How to use molecular biology tools for the study of the anaerobic digestion process? *Rev. Environ. Sci. Bio-Technol.* **14**, 555–593, <https://doi.org/10.1007/s11157-015-9380-8> (2015).
- Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J. & Tyson, G. W. Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.* **27**, 55–64, <https://doi.org/10.1016/j.copbio.2013.11.004> (2014).
- Alivisatos, A. P. *et al.* A unified initiative to harness Earth's microbiomes. *Science* **350**, 507–508, <https://doi.org/10.1126/science.aac8480> (2015).
- Blaser, M. J. *et al.* Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges. *Mbio* **7**, 16, <https://doi.org/10.1128/mBio.00714-16> (2016).

27. Bouchez, T. *et al.* Molecular microbiology methods for environmental diagnosis. *Environmental Chemistry Letters* **14**, 423–441, <https://doi.org/10.1007/s10311-016-0581-3> (2016).
28. Schnurer, A. In *Anaerobes in Biotechnology* Vol. 156 *Advances in Biochemical Engineering-Biotechnology* (eds R. HattiKaul, G. Mamo, & B. Mattiasson) 195–234 (Springer-Verlag Berlin, 2016).
29. Angenent, L. T., Karim, K., Al-Dahhan, M. H. & Domiguez-Espinosa, R. Production of bioenergy and biochemicals from industrial and agricultural wastewater. *Trends Biotechnol.* **22**, 477–485, <https://doi.org/10.1016/j.tibtech.2004.07.001> (2004).
30. Chouari, R. *et al.* Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ. Microbiol.* **7**, 1104–1115, <https://doi.org/10.1111/j.1462-2920.2005.00795.x> (2005).
31. Nelson, M. C., Morrison, M. & Yu, Z. T. A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresour. Technol.* **102**, 3730–3739, <https://doi.org/10.1016/j.biortech.2010.11.119> (2011).
32. Sundberg, C. *et al.* 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. *FEMS Microbiol. Ecol.* **85**, 612–626, <https://doi.org/10.1111/1574-6941.12148> (2013).
33. De Vrieze, J., Hennebel, T., Boon, N. & Verstraete, W. *Methanosarcina*: The rediscovered methanogen for heavy duty biomethanation. *Bioresour. Technol.* **112**, 1–9, <https://doi.org/10.1016/j.biortech.2012.02.079> (2012).
34. Chen, Y., Cheng, J. J. & Creamer, K. S. Inhibition of anaerobic digestion process: A review. *Bioresour. Technol.* **99**, 4044–4064, <https://doi.org/10.1016/j.biortech.2007.01.057> (2008).
35. Niu, Q. G., Qiao, W., Qiang, H. & Li, Y. Y. Microbial community shifts and biogas conversion computation during steady, inhibited and recovered stages of thermophilic methane fermentation on chicken manure with a wide variation of ammonia. *Bioresour. Technol.* **146**, 223–233, <https://doi.org/10.1016/j.biortech.2013.07.038> (2013).
36. Hao, L. P. *et al.* Self-adaption of methane-producing communities to pH disturbance at different acetate concentrations by shifting pathways and population interaction. *Bioresour. Technol.* **140**, 319–327, <https://doi.org/10.1016/j.biortech.2013.04.113> (2013).
37. Werner, J. J. *et al.* Microbial Community Dynamics and Stability during an Ammonia-Induced Shift to Syntrophic Acetate Oxidation. *Appl. Environ. Microbiol.* **80**, 3375–3383, [https://doi.org/10.1016/0141-4607\(84\)90080-5](https://doi.org/10.1016/0141-4607(84)90080-5) (2014).
38. Nelson, M. C., Morrison, M., Schanbacher, F. & Yu, Z. T. Shifts in microbial community structure of granular and liquid biomass in response to changes in feed and digester design in anaerobic digesters receiving food-processing wastes. *Bioresour. Technol.* **107**, 135–143, <https://doi.org/10.1016/j.biortech.2011.12.070> (2012).
39. Regueiro, L., Veiga, P., Figueroa, M., Lema, J. M. & Carballa, M. Influence of transitional states on the microbial ecology of anaerobic digesters treating solid wastes. *Appl. Microbiol. Biotechnol.* **98**, 2015–2027, <https://doi.org/10.1007/s00253-013-5378-8> (2014).
40. Bocher, B. T. W., Cherukuri, K., Maki, J. S., Johnson, M. & Zitomer, D. H. Relating methanogen community structure and anaerobic digester function. *Water Res.* **70**, 425–435, <https://doi.org/10.1016/j.watres.2014.12.018> (2015).
41. Poirier, S., Bize, A., Bureau, C., Bouchez, T. & Chapleur, O. Community shifts within anaerobic digestion microbiota facing phenol inhibition: Towards early warning microbial indicators? *Water Res.* **100**, 296–305, <https://doi.org/10.1016/j.watres.2016.05.041> (2016).
42. De Vrieze, J. *et al.* Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. *Water Res.* **75**, 312–323, <https://doi.org/10.1016/j.watres.2015.02.025> (2015).
43. Prakash, O., Pandey, P. K., Kulkarni, G. J., Mahale, K. N. & Shouche, Y. S. Technicalities and Glitches of Terminal Restriction Fragment Length Polymorphism (T-RFLP). *Indian Journal of Microbiology* **54**, 255–261, <https://doi.org/10.1007/s12088-014-0461-0> (2014).
44. Liu, W. T., Marsh, T. L., Cheng, H. & Forney, L. J. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **63**, 4516–4522 (1997).
45. Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D. & Verstraete, W. How to get more out of molecular fingerprints: practical tools for microbial ecology. *Environ. Microbiol.* **10**, 1571–1581, <https://doi.org/10.1111/j.1462-2920.2008.01572.x> (2008).
46. Brugger, S. D. *et al.* 16S rRNA Terminal Restriction Fragment Length Polymorphism for the Characterization of the Nasopharyngeal Microbiota. *PLoS One* **7**, 8, <https://doi.org/10.1371/journal.pone.0052241> (2012).
47. Castro-Carrera, T. *et al.* Rumen bacterial community evaluated by 454 pyrosequencing and terminal restriction fragment length polymorphism analyses in dairy sheep fed marine algae. *Journal of Dairy Science* **97**, 1661–1669, <https://doi.org/10.3168/jds.2013-7243> (2014).
48. de la Fuente, G. *et al.* Pros and Cons of Ion-Torrent Next Generation Sequencing versus Terminal Restriction Fragment Length Polymorphism T-RFLP for Studying the Rumen Bacterial Community. *PLoS One* **9**, 13, <https://doi.org/10.1371/journal.pone.0101435> (2014).
49. Witzig, M. *et al.* Spatial Variation of the Gut Microbiota in Broiler Chickens as Affected by Dietary Available Phosphorus and Assessed by T-RFLP Analysis and 454 Pyrosequencing. *PLoS One* **10**, 18, <https://doi.org/10.1371/journal.pone.0143442> (2015).
50. van Dorst, J. *et al.* Community fingerprinting in a sequencing world. *FEMS Microbiol. Ecol.* **89**, 316–330, <https://doi.org/10.1111/1574-6941.12308> (2014).
51. Pilloni, G., Granitsiotis, M. S., Engel, M. & Lueders, T. Testing the Limits of 454 Pyrotag Sequencing: Reproducibility, Quantitative Assessment and Comparison to T-RFLP Fingerprinting of Aquifer Microbes. *PLoS One* **7**, 7, <https://doi.org/10.1371/journal.pone.0040467> (2012).
52. Ziganshina, E. E., Ibragimov, E. M., Vankov, P. Y., Miluykov, V. A. & Ziganshin, A. M. Comparison of anaerobic digestion strategies of nitrogen-rich substrates: Performance of anaerobic reactors and microbial community diversity. *Waste Manage.* **59**, 160–171, <https://doi.org/10.1016/j.wasman.2016.10.038> (2017).
53. Goux, X. *et al.* Microbial community dynamics in replicate anaerobic digesters exposed sequentially to increasing organic loading rate, acidosis, and process recovery. *Biotechnol. Biofuels* **8**, 18, <https://doi.org/10.1186/s13068-015-0309-9> (2015).
54. Lim, J. W., Ge, T. & Tong, Y. W. Monitoring of microbial communities in anaerobic digestion sludge for biogas optimisation. *Waste Manage.* **71**, 334–341, <https://doi.org/10.1016/j.wasman.2017.10.007> (2018).
55. Boon, N., De Windt, W., Verstraete, W. & Top, E. M. Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiol. Ecol.* **39**, 101–112, <https://doi.org/10.1111/j.1574-6941.2002.tb00911.x> (2002).
56. Muyzer, G., Dewaal, E. C. & Uitterlinden, A. G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**, 695–700 (1993).
57. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624, <https://doi.org/10.1038/ismej.2012.8> (2012).
58. Joshi, N. A. & Fass, J. N. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files* (2011).
59. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14**, 11, <https://doi.org/10.1186/1471-2164-14-s1-s7> (2013).
60. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: PAired-eND Assembler for Illumina sequences. *BMC Bioinformatics* **13**, <https://doi.org/10.1186/1471-2105-13-31> (2012).
61. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research* **43**, 16, <https://doi.org/10.1093/nar/gku1341> (2015).

62. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–+, <https://doi.org/10.1038/nmeth.2604> (2013).
63. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267, <https://doi.org/10.1128/aem.00062-07> (2007).
64. R Development Core Team. *R: A Language and Environment for Statistical Computing*. 3.0 ed. Vienna, Austria: R Foundation for Statistical Computing., (2013).
65. Oksanen, J. *et al.* Vegan: Community ecology package. *R package version 2*, 3–4 (2016).
66. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217, <https://doi.org/10.1371/journal.pone.0061217> (2013).
67. Connelly, S. *et al.* Bioreactor Scalability: Laboratory-Scale Bioreactor Design Influences Performance, Ecology, and Community Physiology in Expanded Granular Sludge Bed Bioreactors. *Front. Microbiol.* **8**, 15, <https://doi.org/10.3389/fmicb.2017.00664> (2017).
68. McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, 12, <https://doi.org/10.1371/journal.pcbi.1003531> (2014).
69. Hill, M. O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427–432, <https://doi.org/10.2307/1934352> (1973).
70. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 326–349 (1957).
71. Chao, A. Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics* **11**, 265–270, <https://doi.org/10.2307/4615964> (1984).
72. Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**, 57–68, <https://doi.org/10.1007/bf00038687> (1987).
73. Wolda, H. Similarity indices, sample size and diversity. *Oecologia* **50**, 296–302, <https://doi.org/10.1007/bf00344966> (1981).
74. Lorenz, M. O. Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association* **9**, 209–219, <https://doi.org/10.2307/2276207> (1905).
75. Dejonghe, W., Boon, N., Seghers, D., Top, E. M. & Verstraete, W. Bioaugmentation of soils by increasing microbial richness: missing links. *Environ. Microbiol.* **3**, 649–657, <https://doi.org/10.1046/j.1462-2920.2001.00236.x> (2001).
76. Pareto, V. *Le Cours d'Economie Politique*. (Droz, 1897).
77. Butts, C. T. Social Network Analysis with sna. *Journal of Statistical Software* **24**, 51, <https://doi.org/10.18637/jss.v024.i06> (2008).
78. Butts, C. T. network: A Package for Managing Relational Data in R. *Journal of Statistical Software* **24**, 36, <https://doi.org/10.18637/jss.v024.i02> (2008).
79. Gonzalez, A. M. M., Dalsgaard, B. & Olesen, J. M. Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity* **7**, 36–43, <https://doi.org/10.1016/j.ecocom.2009.03.008> (2010).
80. Greenberg, A. E., Clesceri, L. S. & Eaton, A. D. *Standard Methods for the Examination of Water and Wastewater* 18th edn, (American Public Health Association Publications, 1992).
81. Anthonisen, A. C., Loehr, R. C., Prakasam, T. B. S. & Srinath, E. G. Inhibition of Nitrification by Ammonia and Nitrous Acid. *Journal Water Pollution Control Federation* **48**, 835–852 (1976).
82. Briones, A. & Raskin, L. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr. Opin. Biotechnol.* **14**, 270–276, [https://doi.org/10.1016/s0958-1669\(03\)00065-x](https://doi.org/10.1016/s0958-1669(03)00065-x) (2003).
83. Fernandez, A. *et al.* How stable is stable? Function versus community composition. *Appl. Environ. Microbiol.* **65**, 3697–3704 (1999).
84. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233, <https://doi.org/10.1038/35012234> (2000).
85. Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157–1160, <https://doi.org/10.1038/nature03891> (2005).
86. Cardinale, B. J., Palmer, M. A. & Collins, S. L. Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature* **415**, 426–429, <https://doi.org/10.1038/415426a> (2002).
87. Naeem, S. & Li, S. B. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509, <https://doi.org/10.1038/37348> (1997).
88. Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* **3**, 19, <https://doi.org/10.3389/fmicb.2012.00417> (2012).
89. Allison, S. D. & Martiny, J. B. H. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. USA* **105**, 11512–11519, <https://doi.org/10.1073/pnas.0801925105> (2008).
90. De Vrieze, J. *et al.* Microbial community redundancy in anaerobic digestion drives process recovery after salinity exposure. *Water Res.* **111**, 109–117, <https://doi.org/10.1016/j.watres.2016.12.042> (2017).
91. Werner, J. J. *et al.* Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. USA* **108**, 4158–4163, <https://doi.org/10.1073/pnas.1015676108> (2011).
92. Alsouleman, K. *et al.* Reorganisation of a mesophilic biogas microbiome as response to a stepwise increase of ammonium nitrogen induced by poultry manure supply. *Bioresour. Technol.* **208**, 200–204, <https://doi.org/10.1016/j.biortech.2016.02.104> (2016).
93. Bent, S. J. & Forney, L. J. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J.* **2**, 689–695, <https://doi.org/10.1038/ismej.2008.44> (2008).
94. Gwinn, D. C., Allen, M. S., Bonvechio, K. I., V. Hoyer, M. & Beesley, L. S. Evaluating estimators of species richness: the importance of considering statistical error rates. *Methods in Ecology and Evolution* **7**, 294–302, <https://doi.org/10.1111/2041-210X.12462> (2016).
95. Reeder, J. & Knight, R. The ‘rare biosphere’: a reality check. *Nat. Methods* **6**, 636–637, <https://doi.org/10.1038/nmeth0909-636> (2009).
96. Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* **11**, 853–862, <https://doi.org/10.1038/ismej.2016.174> (2017).
97. Westcott, S. L. & Schloss, P. D. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* **3**, 23, <https://doi.org/10.7717/peerj.1487> (2015).
98. Marzorati, M., Negroni, A., Fava, F., Verstraete, W. & Boon, N. Application of a molecular based approach for the early detection of short term 3-chloroaniline shock loads on activated sludge bacterial community and functionality. *New Biotechnology* **30**, 763–771, <https://doi.org/10.1016/j.nbt.2013.07.004> (2013).
99. Widder, S. *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* <https://doi.org/10.1038/ismej.2016.45> (2016).
100. Li, J. B. *et al.* Substrate Type and Free Ammonia Determine Bacterial Community Structure in Full-Scale Mesophilic Anaerobic Digesters Treating Cattle or Swine Manure. *Front. Microbiol.* **6**, 10, <https://doi.org/10.3389/fmicb.2015.01337> (2015).
101. Luo, G., Fotidis, I. A. & Angelidaki, I. Comparative analysis of taxonomic, functional, and metabolic patterns of microbiomes from 14 full-scale biogas reactors by metagenomic sequencing and radioisotopic analysis. *Biotechnol. Biofuels* **9**, 1–12, <https://doi.org/10.1186/s13068-016-0465-6> (2016).
102. Gobet, A., Boetius, A. & Ramette, A. Ecological coherence of diversity patterns derived from classical fingerprinting and Next Generation Sequencing techniques. *Environ. Microbiol.* **16**, 2672–2681, <https://doi.org/10.1111/1462-2920.12308> (2014).

103. Langer, S. G., Ahmed, S., Einfalt, D., Bengelsdorf, F. R. & Kazda, M. Functionally redundant but dissimilar microbial communities within biogas reactors treating maize silage in co-fermentation with sugar beet silage. *Microb. Biotechnol.* **8**, 828–836, <https://doi.org/10.1111/1751-7915.12308> (2015).
104. Kirkegaard, R. H. *et al.* The impact of immigration on microbial community composition in full-scale anaerobic digesters. *Sci Rep* **7**, 11, <https://doi.org/10.1038/s41598-017-09303-0> (2017).
105. Riviere, D. *et al.* Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *Isme J.* **3**, 700–714, <https://doi.org/10.1038/ismej.2009.2> (2009).
106. Treu, L., Kougias, P. G., Campanaro, S., Bassani, I. & Angelidaki, I. Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. *Bioresour. Technol.* **216**, 260–266, <https://doi.org/10.1016/j.biortech.2016.05.081> (2016).
107. De Vrieze, J. *et al.* The full-scale anaerobic digestion microbiome is represented by specific marker populations. *Water Res.* **104**, 101–110, <https://doi.org/10.1016/j.watres.2016.08.008> (2016).
108. Fernandez, A. S. *et al.* Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. *Appl. Environ. Microbiol.* **66**, 4058–4067 (2000).
109. Zhang, W., Werner, J. J., Agler, M. T. & Angenent, L. T. Substrate type drives variation in reactor microbiomes of anaerobic digesters. *Bioresour. Technol.* **151**, 397–401, <https://doi.org/10.1016/j.biortech.2013.10.004> (2014).

Acknowledgements

Jo De Vrieze is supported as postdoctoral fellow from the Research Foundation Flanders (FWO). The financial support of the COST Action ES1103 ‘Microbial ecology & the earth system: collaborating for insight and success with the new generation of sequencing tools’ is kindly acknowledged. Susanne Theuerl is supported by the German Federal Ministry of Food and Agriculture (BMEL), grant numbers 22403915 (joint research project Biogas Monitoring Program Part III). This study was also supported by NERC NE/L011956/1, awarded to Umer Zeeshan Ijaz.

Author Contributions

Jo De Vrieze wrote the main manuscript text, prepared the figures and tables, and carried out the statistical comparison of the amplicon sequencing and TRFLP data. Umer Z. Ijaz assisted in the statistical processing of the amplicon sequencing and TRFLP data, and Aaron M. Saunders carried out the analysis and interpretation of the amplicon sequencing data. Susanne Theuerl carried out the analysis and interpretation of the TRFLP data. All authors contributed to the writing of the main manuscript to come to the final version.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-34921-7>.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018