**Materials and Methods**

**Communities and fermentation**

The communities used were collected from commercial operations, both from anaerobic digestors (AD plants) and communities present in nature used to seed the AD plants. The source and types of communities used can be found in Table 1. Communities were stored at 4°C prior to use. For all experiments, communities were grown in 500 ml bottles (600ml total volume with headspace; Duran) using the commercially available Automated Methane Potential Test System (AMPTS, Bioprocess Control Sweden AB) to measure CO2-stripped biogas production (referred to as methane in this paper). Samples were fed weekly in a fed-batch mode using a defined medium (see below for media composition). The communities from experiment 1 were equalised in terms of bacterial cells per gram of sample before inoculation using M9 salts to dilute them to the community with the lowest cell density, based on qPCR enumeration of 16S rRNA gene copies. For experiments 2 and 3, starting 16S rRNA copy number was determined (but not equalised between communities) and did not correlate with methane production. The fermenters were inoculated with 275 g of sample and fed with 25 ml of defined medium: meat extract 111.1 gl-1, cellulose 24.9 gl-1, starch 9.8 gl-1 glucose 0.89 gl-1, xylose 3.55 gl-1 (carbon to nitrogen ratio of 15:1) every week, starting with t0. Before the start of the fermentation, 0.3 mL of 1000x Trace Metal stock (1 gl-1 FeCl2 . 4H2O, 0.5 gl-1 MnCl2 . 4H2O, 0.3 gl-1 CoCl2 . 4H2O, 0.2 gl-1 ZnCl2, 0.1 gl-1 NiSO4 . 6H2O, 0.05 gl-1 Na2MoO4 . 4H2O, 0.02 gl-1 H3BO3, 0.008 gl-1 Na2 WO4 . 2H2O, 0.006 gl-1 Na2SeO3 . 5H2O, 0.002 gl-1 CuCl2 . 2H2O) was added to each fermenter. Experiment 1 (shown in Figure 1) ran for 5 weeks, experiment 2 (shown in Figure 2) for 6 weeks and experiment 3 (shown in Figure 4) for 5 weeks.

**Measuring methane content of Biogas**

All resulting lab-scale reactors inoculated with the samples were run at 37°C using the AMPTS. The AMPTS system measures the volume of biogas produced following stripping of CO2 from the produced gas. We have confirmed that the measured biogas is >95% methane using GC-FID.

**DNA extraction, amplicon library construction and sequencing**

DNA for 16S rRNA gene amplicon sequencing was extracted using QIAamp DNA Stool Mini Kit (QIAGEN) or FastDNA™ SPIN Kit for Soil (MP), depending on the experiment. Note that DNA extraction for mixed community P06 from the 10 mix experiment failed. The DNA for qPCR was extracted with the QIAamp DNA Stool Mini Kit (QIAGEN), protocol for pathogen detection with the 95ºC incubation step and the Powerlyzer Powersoil DNA KIT (MOBIO). DNA from *A. baylyi*, *P. fluorescens* SBW25 for Bacteria from *H. salinarum* DSM 669 for Archaea was used as standards. The primers1 used to identify Bacteria were 16S rRNA 338f - ACT CCT ACG GGA GGC AGC AG, 518r - ATT ACC GCG GCT GCT GG for Archaea: 931f - AGG AAT TGG CGG GGG AGC A, m1100r - BGG GTC TCG CTC GTT RCC. The reagents used were: 1x Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix; 150nM 338f and 300nM 518r or 300nM 931f and 300 nM m1100r; ROX 300nM; and BSA 100 ng/µl final concentration. All samples were run in triplicate on a StepOnePlus (Applied Biosystems) qPCR machine using a program with 3’ 95ºC initial denaturation followed by 40 cycles of 5’’ at 95ºC and 10’’ at 60ºC, followed by a melting curve 95ºC for 15’’; 60ºC for 1’ ramping up to 95ºC in steps of 0.3ºC for 15’’ each. The melting curve analysis and the confirmation of the negative controls was done using Stepone Software v.2.3 (life technologies). The Cq values and the efficiencies of the samples and standards was determined as previously using LinRegPCR version 2016.02. The quantities were calculated using the one point calibration method as described earlier3.

# Analyses of sequenced samples

MiSeq amplicon reads were merged using Illumina-utils software4. We quality-filtered only the mismatches in the overlapping region between read pairs using a minimum overlap (--min-overlap-size) of 200 nt and a minimum quality Phred score (--min-qual-score) of Q20. We allowed no more than five mismatches per 100 nt (-P 0.05) over the 200 nt overlapping region.

Reads that fulfilled the quality criteria were analysed using Quantitative Insights Into Microbial Ecology (QIIME v.1.7)5. We removed chimera using the *identify\_chimeric\_seqs.py* script, UCHIME reference 'Gold' database and USEARCH6,7, which we also used to select OTUs. We assigned the taxonomy of our reads with QIIME *pick\_open\_reference\_otus.py* function, using the Greengenes database version v13\_88 as a reference with a minimum cluster size of 2 (i.e., each OTU must contain at least two sequences). We collapsed the technical replicates and filtered out the low abundance OTUs (<0.01% total, *filter\_otus\_from\_otu\_table.py*) and samples rarefied to an even depth of 26702 for both experiments where sequencing data is available. QIIME was used to calculate alpha and beta diversity data and produce NMDS plots.

For the NNLS analysis, following removal of low abundance OTUs and cumulative sum scaling transformation, the resulting biom file was used to create a matrix $A\in Z\_{ \geq 0}^{ m × n}$ (*m* rows of OTUs by *n* sample columns) for the seed bioreactors, and a column vector$ b\in Z\_{ \geq 0}^{ m}$ for each mixed bioreactor; both$ A$ and $b$ hold non-negative integers of OTU abundances. One of the individual samples contained a negligible number of reads and was discarded from the analysis. The contribution, or weight, of each seed sample to the pattern of OTUs observed in a mixed sample is given by the column vector$ x\in R ^{n}$ when solving for a system of linear equations$ Ax=b$.

When modelling count data for environmental samples the fitted parameters of $x$ will also be non-negative and the number of OTUs will usually exceed the number of samples (*m* > *n*). The task is to solve an over-determined system of linear equations where there are more equations than unknowns. It is likely that some of the linear equations will ‘disagree’ and there will be no exact solution. Geometrically, this may be interpreted as $b$ not lying in the column space of$ A$, a (hyper)plane holding the column vectors of $A$, or $Ax-b\ne 0$. A least-squares approach may find the non-negative vector$ \overbar{x}=(A^{T}A)^{-1}A^{T}b$ which is the projection of $b$ back onto the column space of$ A$ that minimises the least-squares error ‘distance’$ \left‖A\overbar{x}-b\right‖$. For our study the non-negative least-squares (NNLS) solution,$ \overbar{x}$, and least-squares errors were computed via the R packages ‘nnls’9 and ‘limSolve’10 for each of the mixed samples.

1. Øvreås, L. & Torsvik, V. Microbial Diversity and Community Structure in Two Different Agricultural Soil Communities. *Microb. Ecol.* **36,** 303–315 (1998).

2. Ruijter, J. M. *et al.* Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* **37,** e45 (2009).

3. Brankatschk, R., Fischer, T., Veste, M. & Zeyer, J. Succession of N cycling processes in biological soil crusts on a Central European inland dune. *FEMS Microbiol. Ecol.* **83,** 149–160 (2013).

4. Eren, A. M. *et al.* Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4,** 1111–1119 (2013).

5. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–336 (2010).

6. Edgar, R. R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

7. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).

8. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6,** 610–8 (2012).

9. Mullen, K. & Van Stokkum, I. The Lawson-Hanson algorithm for non-negative least squares (NNLS). (2007).

10. Soetaert, K., Meersche, K. V. D. & Oevelen, D. V. Package limSolve, solving linear inverse models in R. (2009).