Methanogenesis is the primary biogenic source of methane in the atmosphere and a key contributor to climate change. The long-standing dogma that methanogenesis originated within the Euryarchaeota was recently challenged by the discovery of putative methane-metabolizing genes in members of the Bathyrarchaeota, suggesting that methanogenesis may be more phylogenetically widespread than currently appreciated. Here, we present the discovery of divergent methyl-coenzyme M reductase genes in population genomes recovered from anoxic environments with high methane flux that belong to a new archaeal phylum, the Verstraetearchaeota. These archaea encode the genes required for methylotrophic methanogenesis, and may conserve energy using a mechanism similar to that proposed for the obligate H₂-dependent methylotrophic Methanomassiliicoccales and recently described Candidatus 'Methanofastidiosa'. Our findings indicate that we are only beginning to understand methanogen diversity and support an ancient origin for methane metabolism in the Archaea, which is changing our understanding of the global carbon cycle.

Methanogenesis performed by anaerobic archaea represents the largest biogenic source of methane on Earth. This process is a key component of the global carbon cycle and plays an important role in climate change, owing to the high warming potential of methane. Methanogenic archaea are widespread in nature and have been detected in a diverse range of environments including wetlands, rice agriculture, landfills, hydrothermal vents, ruminants and termites. Traditionally, all methanogens belonged to the archaeal phylum Euryarchaeota and until recently were classified into seven orders (Methanobacterales, Methanococcales, Methanomicrobiales, Methanosarcinales, Methanocellales, Methanopyrales and Methanomassiliicoccales). Substrates for methane production include H₂/CO₂ (hydrogenotrophic), acetate (acetoclastic) and methylated compounds (methylotrophic). Hydrogenotrophic methanogenesis is the most widespread, being found in all methanogenic orders with the exception of Methanomassiliicoccales, and it has been suggested that this is the ancestral form of methane production. Although the prevailing source of atmospheric methane is through direct acetate cleavage, only members of the Methanosarcinales are capable of acetoclastic methanogenesis. Methylotrophic methanogens are found in the orders Methanococcales, Methanobacteriales and Methanomassiliicoccales, and can be classified in two groups based on the presence or absence of cytochromes. Methylotrophs without cytochromes are obligately H₂-dependent, while those that possess cytochromes (that is, members of the Methanosarcinales), can also oxidize methyl groups to CO₂ via a membrane-bound electron transport chain.

Our knowledge of methanogenic substrate utilization and energy conservation is still incomplete but is rapidly expanding through the development of new molecular techniques such as genome-centric metagenomics. This includes the discovery of hydrogenotrophic methanogenesis genes encoded in Methanoanaeta, a genus originally thought to be strictly acetoclastic, the hypothesized mode of energy metabolism in Methanomassiliicoccales, and the recently described sixth class of euryarchaeotal methanogens, Candidatus 'Methanofastidiosa', which is restricted to methanogenesis through methylated thiol reduction. The hypothesis that methane metabolism originated early in the evolution of the Euryarchaeota has recently been challenged following the discovery of putative methane metabolism in the archaeal phylum Bathyrarchaeota (formerly Miscellaneous Crenarchaeota Group) and the development of new molecular techniques such as genome-centric metagenomics. This includes the discovery of hydrogenotrophic Methanomassiliicoccales and recently described Candidatus 'Methanofastidiosa'. Our findings indicate that we are only beginning to understand methanogen diversity and support an ancient origin for methane metabolism in the Archaea, which is changing our understanding of the global carbon cycle.

To enrich for novel methanogenic diversity, triplicate anaerobic digesters were inoculated with a mixture of samples sourced from natural (rumen and lake sediment) and engineered (anaerobic digesters and lagoon) environments, and supplied with alpha cellulose. The reactors showed stable methane production and were sampled over time. Six metagenomes comprising a total of 111 Gb were generated and co-assembled (Supplementary Table 1), and differential coverage binning recovered 68 substantially to near-complete populations, including four containing mcrA genes. The homologues in these genomes are closely related to recognized methanogenic lineages in the Euryarchaeota (94–99% amino acid identity), with four genomes (V1) encoding an mcrA that is divergent from extant sequences (~68% aa identity). This divergent mcrA was used to screen the Sequence Read Archive (SRA) and, closely related mcrA genes (85–100% aa identity, Supplementary Table 2) were identified in three metagenomic data sets obtained from high methane flux habitats: an anaerobic digester treating palm oil mill effluent (Malaysia), an isothermal degrading methanogenic enrichment culture from tailings ponds (Mildred Lake Settling Basin, Canada), and formation waters of coalbed methane wells (CD-8 and PK-28) located within the Surat Basin (Queensland, Australia) (Supplementary Table 1). Subsequent assembly and binning of these data sets led to the recovery of four additional
population genomes (V2–V5, Table 1) containing mcra sequences similar to the one found in V1. All five divergent mcra-containing genomes are near-complete (90.3–99.1%) with low contamination (<1%), with the exception of V5, which is moderately complete (60.2%) with no detectable contamination. The estimated size of these genomes is relatively small (1.2–1.7 Mb), with a tight GC content distribution (55.4–58.1%) and number of predicted genes ranging between 1,248 and 1,758 (Table 1).

Phylogenetic analyses of the mcra and 16S rRNA genes from these population genomes confirmed the novelty of this group (Fig. 1). The mcra genes form a distinct cluster divergent from Euryarchaeota, the recently described Bathyarchaeota, and an environmental clade (recovered from Coal Oil Point seep field, COP)12 (Fig. 1a). The 16S rRNA gene tree shows that these genomes are near-complete (90.3–99.1%) with low contamination. The estimated size of these genomes is relatively small (1.2–1.7 Mb), with a tight GC content distribution (55.4–58.1%) and number of predicted genes ranging between 1,248 and 1,758 (Table 1).

A new archaeal phylum, Verstraetearchaeota
A genome tree comprising the five divergent mcra-containing population genomes and 346 publicly available archaeal genomes was inferred using a concatenated set of 122 archaeal-specific marker genes. This tree confirms the placement of the population genomes in an archaeal phylum that is sister to the Crenarchaeota and Korarchaeota (Fig. 2a and Supplementary Fig. 1), for which we propose the name Candidatus ‘Verstraetearchaeota’. Comparative genomic analysis revealed that these genomes have an average amino acid identity (AAI) of 41 ± 1.5% to other Archaea, supporting their grouping as a separate phylum18.

Table 1 | Summary statistics of the Verstraetearchaeota genomes.

| Genome | Environment | Size (Mb) | Scaffolds (no.) | NS0 (bp) | Genes (no.) | GC (%) | CD (%) | Compl. (%) | Cont. (%) | Strain (%) | aa (no.) |
|--------|-------------|-----------|----------------|----------|-------------|--------|--------|------------|----------|------------|---------|
| V1     | Cellulose degrading anaerobic digester | 1.6     | 88              | 34,490   | 1,578       | 58.1   | 88     | 90.3       | 0.9      | 0          | 18      |
| V2     | Palm oil mill effluent degrading anaerobic digester | 1.7     | 11              | 258,701  | 1,758       | 57.6   | 88.5   | 99.1       | 0.9      | 0          | 18      |
| V3     | Iso-alkane degrading enrichment from tailings ponds | 1.4     | 86              | 24,381   | 1,550       | 55.4   | 87.4   | 99.1       | 0.9      | 0          | 18      |
| V4     | Formation water from coalbed methane well (CD-8) | 1.4     | 84              | 20,373   | 1,423       | 55.8   | 87.3   | 91.6       | 0.0      | 0          | 21      |
| V5     | Formation water from coalbed methane well (PK-2B) | 1.2     | 295             | 2,926    | 1,248       | 55.4   | 83.5   | 60.2       | 0.0      | 0          | 21      |

Genome ID, habitat, estimated size, number of scaffolds, NS0, number of genes, guanine-cytosine (GC) content, coding density (CD), estimated completeness (Compl.), estimated contamination (Cont.), strain heterogeneity (Strain) and estimated number of amino acids (aa) are shown.

Figure 1 | Phylogenetic trees showing the placement of the Verstraetearchaeota mcra and 16S rRNA genes. a, McrA protein tree showing monophyletic clustering of the divergent McrA sequences from V1–V4 (red) outside known euryarchaeotal and bathyarchaeotal methanogenic lineages. b, 16S rRNA gene tree showing the placement of V1–V5 (bolded) with environmental sequences classified as the Terrestrial Miscellaneous Crenarchaeota Group (TMCG), using the Batharchaeota 16S rRNA sequences as the outgroup. Bootstrap values were calculated via non-parametric bootstrapping with 100 replicates, and are represented by circles.
different strains of the same species. We propose the genus names Candidatus 'Methanomethylicus' for V1–V3 and Candidatus 'Methanosurus' for V4–V5.

Methane metabolism

A KEGG orthology (KO) based comparison of the four near-complete Verstraetearchaeota genomes (V1–V4; Table 1) was conducted against 156 selected archaean genomes (Supplementary Fig. 2). At this global metabolic scale, Verstraetearchaeota resemble the Methanomassiliicoccales, C. Methanofastidiosum and Bathyarchaeota, and are distinct from traditional euryarchaeotal methanogens (Supplementary Fig. 2). The V1–V4 genomes share 458 gene orthologues, with an additional 195 shared among V1–V3. The V4 genome contains 147 genes for which no homologues could be identified in the other genomes, consistent with the phylogenetic placement of this genome in a separate genus (Fig. 2a and Supplementary Fig. 3).

Metabolic reconstruction of the near-complete Verstraetearchaeota genomes revealed the presence of key genes associated with methylothrophic methanogenesis (Figs 2b, 3a, and Supplementary Table 3). These include a complete methyl-coenzyme M reductase complex (MCR; mcrABG and ancillary genes mcrCD) with conserved binding sites for coenzyme M, coenzyme B and F430 cofactors (Supplementary Fig. 4), and genes required for methane production from methanol (mtaA) and methanethiol (mtsA). In addition, a variable complement of genes for methylamine utilization (mtbA, mtxB, mtbBC, mtbBC and ramA) was identified across the genomes (Figs 2b, 3a and Supplementary Table 3). However, due to high sequence similarity, it is difficult to infer the substrate specificity of mtaA and mtbA, and it has been suggested that mtaA is capable of functionally replacing mtbA in methylamine utilization2. Subunit H of the tetrahydromethanopterin S-methyltransferase (mtrH) is also present in the Verstraetearchaeota and co-located with a gene encoding a methylamine-specific corrinoid protein, similar to the mtrH gene found in the Bathyarchaeota2,19, suggesting it plays a role in methyamine:coenzyme M methyl transfer activity. The remaining genes for hydrogenotrophic and acetoclastic methanogenesis are missing from the genomes, consistent with obligate H2-dependent methylothrophic methanogenesis. The contigs containing methanogenesis-related genes were independently found in all of the Verstraetearchaeota genomes and have sequence composition characteristics typical of their respective genomes (Supplementary Fig. 5).

The Verstraetearchaeota appear to have three possible mechanisms for the reduction of heterodisulfide (CoM-S-S-CoB), formed during the final step of methanogenesis (Fig. 3a). The first mechanism couples the exergonic H2-dependent reduction of heterodisulfide to ferredoxin reduction, and involves a heterodisulfide reductase (HdrABC) and a putative cytoplasmic F420-non-reducing hydrogenase (MvhADG and MvhB)3 (Fig. 3a). Divergent homologues of the genes encoding this electron-bifurcating complex are found co-located in the Verstraetearchaeota (Supplementary Table 3). In addition, homologues of the membrane-bound NADH-ubiquinone oxidoreductase (Nuo), which show high sequence similarity to the F420-methanophenazine oxidoreductase (Fpo) found in Methanocarcininales20, are present and form an Fpo-like complex capable of re-oxidizing reduced ferredoxin, with concomitant translocation of protons or sodium ions across the cytoplasmic membrane (Fig. 3a). Similar to Methanomassiliicoccales5,9 and Methanoseta thermophila21, the subunits required for binding and oxidation of NADH (NuoEG) or F420 (FpoFO) are missing from the Verstraetearchaeota, supporting the use of ferredoxin as an electron donor. Reduced ferredoxin can also be re-oxidized by the energy-conserving hydrogenase B (Ehb) found in all genomes, as suggested for Methanosphaera stadtmanae3 and the recently described C. Methanofastidiosum methylothrophicus19. The proton gradient generated by the Fpo-like or Ehb complex can drive adenosine triphosphate (ATP) synthesis via an archaeal-type ATP synthase, thereby coupling methanogenesis to energy conservation and enabling internal hydrogen cycling (Fig. 3a). The second mechanism for heterodisulfide reduction uses HdrD, which is present in three copies in all Verstraetearchaeota genomes. Although HdrD is conventionally linked to HdrE (ref. 2), the latter is missing in the genomes, an observation that has also been made for the Methanomassiliicoccales9. C. Methanofastidiosum methylothrophicus19 and Bathyarchaeota21. Instead, HdrD may interact directly with the Fpo-like complex and together function as an energy-converting ferredoxin:heterodisulfide oxidoreductase2. Alternatively, the last mechanism could use HdrD, which is co-located with a gene encoding a cytoplasmic flavin adenine dinucleotide-containing dehydrogenase (gkD), similar in sequence to d-lactate dehydrogenase. This mechanism may allow the reduction of heterodisulfide to be coupled to lactate utilization (Fig. 3a), as suggested for the Bathyarchaeota2 and Archaeoglobus fulgidis22.

Fermentative metabolism

In addition to methane metabolism, the Verstraetearchaeota appear to be capable of utilizing sugars as a carbon source and generating acetyl-CoA via the Embden–Meyerhof–Parnas (EMP) pathway and pyruvate-ferredoxin oxidoreductase (Por) (Fig. 3b and Supplementary Table 4). In contrast to V1–V3, V4 contains multiple polysaccharide and sugar transporters, yet lacks phosphoglucose mutase (App) and enolase (Eno), suggesting that V4 may use an alternative or divergent form of the pathway for the conversion of 3-phosphoglycerate to phosphoenolpyruvate. All genomes also encode an intermediate type II/III ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBiSCO) (Supplementary Fig. 6), which may function in a nucleotide salvage pathway generating 3-phosphoglycerate from adenosine monophosphate (AMP) that can enter the EMP pathway to produce acetyl-CoA (refs 23–25) (Fig. 3b). The presence of multiple genes encoding peptidase and amino-acid transporters suggests that the Verstraetearchaeota are capable of importing these compounds, which can then be degraded to keto-acids by various endopeptidases (PepB, PepD, PepP and PepT), glutamate dehydrogenase (Gdh) and aminotransferases (AspB, ArgD, GabT, IleE, GlnS and HisC). The keto-acids could be converted to acetyl-CoA by 2-oxoacid:ferredoxin oxidoreductases (Kor, Ior, Por and Vor) and aldehyde-ferredoxin oxidoreductases (Aor)2,26 (Fig. 3b and Supplementary Table 4). Acetyl-CoA, generated through the various metabolic processes, can be converted to acetate via an archaeal-type adenine diphosphate (ADP)-forming acetate synthetase (Acd)28, which would allow the Verstraetearchaeota to produce energy via substrate-level phosphorylation. Although this is unusual for microorganisms involved in methane metabolism, the ability to carry out complex fermentation has also been detected in the Bathyarchaeota2 and could suggest substrate diversification for these newly discovered methanogens.

Environmental distribution

To assess the environmental distribution of the Verstraetearchaeota, we searched for related 16S rRNA and mcrA genes in publicly available databases. Full-length 16S rRNA gene sequences in Greengenes most closely related to V1–V3 were detected in two habitats, a mesophilic methanogenic granular sludge reactor29 and deep sediment of the freshwater Lake Pavin (France)30. Those most closely related to V4–V5 were found in the Niibori petroleum reservoir31 and Zavarzin thermal spring32 (Fig. 1b). Most other clones related to Crenarchaeota and found in hot spring sediments are basal to the Verstraetearchaeota. Partial-length gene sequences identified in the SRA suggest that the
Figure 2 | Genome tree and distribution of genes involved in methane metabolism. a. Genome tree illustrating the placement of V1-V5 relative to 346 archaeal genomes, using the DPANN superphylum as the outgroup. The tree was inferred using the maximum-likelihood method with a concatenated set of 122 archaeal-specific marker genes, and bootstrap values were calculated using non-parametric bootstrapping with 100 replicates (represented by circles). Methanogenic Euryarchaeota are expanded and colour-coded based on order-level classification.

b. Presence/absence of genes involved in methane metabolism and energy conservation for euryarchaeotal methanogens, the Verstraetenearchaeota (V1-V5) and Bathyarchaeota (BA1-BA2). Annotation was primarily based on classification of genes into KO groups, and a functional complex consisting of multiple subunits was considered present if >20% of the genes comprising the complex were detected. Colour coding on the right-hand side of the figure reflects the genome tree classification (MMi: Methanomicrobiales; MS: Methanosarcinales; MCo: Methanococcales; MMa: Methanomassiliicoccales; MF: Condatus “Methanofastidiosales”; MB: Methanobacteriales; MCo: Methanococcales; MP: Methanopyrales; V: Verstraetenearchaeota; B: Bathyarchaeota).
Figure 3 | Proposed metabolism of the Verstraetearchaeota. a, Pathways for methylotrophic methanogenesis, hydrogen cycling and suggested energy conservation mechanisms. The first mechanism is shown in blue and entails heterodisulfide reduction by HdrABC–MvhABDG coupled to ferredoxin reoxidation by an Fpo-like (dark blue) or Ehb (light blue) complex. The second mechanism involves HdrD coupled directly to the Fpo-like complex (light green), and the third possibility links HdrD to an FAD hydrogenase (dark green). The second mechanism involves HdrD coupled directly to the Fpo-like complex (light green), and the third possibility links HdrD to an FAD hydrogenase (dark green). b, Other metabolic pathways including potential sugar and amino acid utilization, an incomplete TCA cycle and a nucleotide salvage pathway. Black arrows indicate genes that were found in all near-complete Verstraetearchaeota genomes (V1–V4), and grey arrows represent genes that are present in only a subset of the genomes (coloured circles indicate in which subset of genomes they are found). Dashed light grey arrows show parts of a pathway that are missing in all genomes.

Evolution of methanogenesis

To examine the evolutionary history of methane metabolism in the Archaea, protein trees were constructed for all subunits of the MCR complex and ancillary proteins (Fig. 4 and Supplementary Figs 7, 8). In all cases, the Verstraetearchaeota form a robust monophyletic group, as do the seven traditional and one recently described euryarchaeotal methanogenic orders, the Batharchaeota and the COP environmental clade12, which is consistent with the genome tree and rules out recent lateral gene transfer (LGT) of the MCR complex between orders or phyla. Furthermore, higher-level associations corresponding to the traditional ‘Class I’ and ‘Class II’ groupings of euryarchaeotal methanogens are present and support vertical inheritance13, with the exception of an ancient lateral transfer of mrt from Methanococcales to some Methanobacteriales. The Verstraetearchaeota are not robustly associated with any other methanogenic lineage in the gene trees, which precludes our ability to distinguish between vertical inheritance and ancient LGT from a euryarchaeotal donor. However, given that this lineage has all the core genes required for methanogenesis and associated energy conservation, it is unlikely that methanogenesis in the Verstraetearchaeota could be a result of LGT. We therefore predict that methanogenesis pre-dates the origin of the Euryarchaeota13 and that this trait will be found in as yet undiscovered lineages across the Archaea, supporting methane metabolism as one of the most ancient forms of microbial metabolism.

C. ‘Methanomethylucus mesodigestum’ (gen. nov., sp. nov.)

*Methanomethylucus mesodigestum* (L. n. *methanum*, methane; L. adj. *methylucus*, methyl, referring to the methylotrophic methane-producing metabolism; L. pref. *meso*, middle, highlighting the preference for mesophilic habitats; L. v. *digestus*, form of digest, undergo digestion, pertaining to the anaerobic digester environment). This organism is inferred to be capable of methylotrophic methanogenesis, is non-motile, and not cultivated, represented by near-complete population genomes V1 (type strain) and V2 obtained from anaerobic digesters (Table 1).
C. ‘Methanomethylucus oleusabulum’ (gen. nov., sp. nov.)
Methanomethylucus oleusabulum (L. pref. methanum, methane; L. adj. methylucus, methyl, referring to the methylotrophic methane-producing metabolism; L. pref. oleum, oil; L. n. sabulum, sand, representing the oil sands environment). This organism is inferred to be capable of methylotrophic methanogenesis, is non-motile, and not cultivated, represented by a near-complete population genome V3 obtained from an iso-alkane degrading enrichment culture from tailings ponds (Table 1).

C. ‘Methanosuratus petracarbonis’ (gen. nov., sp. nov.)
Methanosuratus petracarbonis (L. pref. methanum, methane, referring to the methylotrophic methane-producing metabolism; suratus, referring to the Surat Basin where the genomes were found; L. n. petra, rock, representing the formation water habitat; L. n. carbo carbonis, coal, carbon, pertaining to the coalbed methane well environment). This organism is inferred to be capable of methylotrophic methanogenesis, is non-motile, and not cultivated, represented by population genomes V4 (near-complete) and V5 (moderately complete) obtained from formation water of coalbed methane wells (Table 1).

Description of Methanomethyliaeae (fam. nov.)
The description is the same as for the genus Methanomethylucus. Suf. -acieae, ending to denote a family. Type genus: Candidatus ‘Methanomethylucus’, gen. nov.

Description of Methanomethyliales (ord. nov.)
The description is the same as for the genus Methanomethylucus. Suf. -iales, ending to denote an order. Type family: Methanomethyliaeae, fam. nov.

Description of Methanomethylia (class nov.)
The description is the same as for the genus Methanomethylucus. Suf. -ia, ending to denote a class. Type order: Methanomethyliales, ord. nov.

Description of Verstraetearchaeota (phyl. nov.)
We propose the name Verstraetearchaeota for this phylum; Verstraete, recognizing the contributions of Professor Willy Verstraete (Centre for Microbial Ecology and Technology, Ghent University, Belgium) to the development and application of engineered microbial ecosystems such as anaerobic digesters; suf. -archaeota, ending to denote an archaeal phylum.

Methods
Anaerobic digester sample collection, DNA extraction and library preparation.
Triplicate cellulose-degrading anaerobic digesters (biological replicates n = 3) were inoculated with a mixture sourced from natural (rumen and lake sediment) and engineered (anaerobic lagoon, granules and several digesters) environments13. The digesters were operated for 362 days with continual monitoring of reactor performance parameters and microbial community dynamics. Samples for DNA extraction and metagenomic sequencing were taken at day 96 and day 362 (total of six samples), centrifuged at 14,000g for 2 min to collect the biomass, and snap-frozen in liquid nitrogen before storing at −80 °C until further processing. DNA was extracted from these samples using the MP-Bio Fast DNA Spin Kit for Soil (MP Biomedicals), following the manufacturer’s instructions, and stored at −20 °C until sequencing. Libraries for metagenomic sequencing were prepared as described previously and sequenced using one-third of an Illumina HiSeq2000 flowcell lane each14.

Metagenome assembly and population genome binning.
The six metagenomes (2 × 150 bp paired-end reads, 111 Gb) generated from the cellulose-degrading anaerobic digesters were quality filtered, trimmed and combined into a large co-assembly using CLC Genomics Workbench v6 (CLC Bio)15. Population genomes were recovered from the co-assembly based on differential coverage profiles, kmer signatures and GC content using GroupM v2.0 (ref. 34), and manually refined using the ‘refine’ function in GroupM, resulting in 101 population genomes14 (PRJNA284316). One of the genomes (referred to as V1) contained an mcrA sequence with low similarity to euryarchaeotal mcrA sequences (~68% aa identity).
Raw reads from a metagenome constructed from an anaerobic digester treating palm oil mill effluent (RE8276848, 18.1 Gb) were processed using Bowtie2 to remove adaptor sequences and merge overlapping reads (https://github.com/jstjohn/SeqPrep.git), and Nesenso for quality trimming with a quality score threshold of 20 (https://github.com/Victorian-Bioinformatics-Consortium/nesenso.git).

Processed reads were assembled using the \textit{de novo} assembly algorithm in CLC Genomics Workbench v6, with a word size of 25 and bubble size of 50. Contigs larger than 2 kbp were binned into population genomes using VizBin (ref. 35). This resulted in a population genome (referred to as V2) that contained an \textit{mcrA} sequence identical to the one found in \textit{V1} (100% aa identity). Metagenomes produced from iso-alkane-degrading enrichment cultures (SRR494224, 2.4 Gb, technical replicates n = 2) and toluene-degrading microbial communities (SRR942927, SRR943311–13 and SRR943315, 8.56 Gb, technical replicates n = 2) from tailings ponds\textsuperscript{17} (Mildred Lake Settling Basin, Canada) were processed using CLC Genomics Workbench v8 for quality trimming and adapter removal. The metagenomes were combined into one large co-assembly using the \textit{de novo} assembly algorithm in CLC v5 with default parameters and minimum contig size of 500 bp. Binning of the contigs into population genomes was performed using MetaBAT (ref. 36), which resulted in 96 genomes, including one (referred to as V3) with an \textit{mcrA} sequence highly similar to the \textit{V1} \textit{mcrA} (95% aa identity).

Metagenomes generated from formation waters of coal bed methane wells (CD-8) and (9G, PK-28 (7.9 Gb)) were processed with SeqPrep and Nesenso for adapter removal, merging and quality trimming. The processed reads for each metagenome were assembled individually using the \textit{de novo} assembly algorithm in CLC Genomics Workbench v6.5, with a word size of 25 and bubble size of 50. Contigs larger than 500 bp from population genomes using DSB v1.3, which groups contigs based on coverage and tetracneulide frequencies (https://github.com/rdp/dparks134/DSB). Each of the data sets produced a population genome that contained an \textit{mcrA} sequence similar to the \textit{V1} \textit{mcrA}, that is, V4 for CD-8 (88% aa identity) and V5 for PK-28 (86% aa identity).

Population genome quality assessment. The five population genomes were manually refined by including unbinned contigs from the metagenomes based on Blastp (ref. 37) v2.2.30\textsuperscript{+} homology to the other genomes, and removing contaminating contigs based on coverage, GC, tetracneulide frequency and taxonomic profiles of the genes using Decon2M (https://github.com/dparks134/refineM). Scaffolding between contigs was performed using the ‘roundup’ function in FinishM (https://github.com/wwood/finishm) and GapCloser (ref. 38).

Completeness, contamination and strain heterogeneity estimates were calculated based on the presence of lineage-specific single copy marker genes using CheckM (ref. 39). The number of \textit{rRNA} genes within each genome was calculated using \textit{rRNA}-scan with default search parameters\textsuperscript{40}.

Screening of SRA using \textit{Verstraetearchaeota} \textit{mcrA} sequences. A collection of environmental public metagenomes were downloaded from the SRA database, based on metadata from SRAdb (ref. 41), and processed using a custom bio-gem bio-sra (https://github.com/wwood/bioruby-sra/). SRA format files were converted to FASTQ using sra-tools (https://ncbi.github.io/sra-tools/) and HMMER (ref. 42) was used with an \textit{mcrA} model to identify reads that probably represent \textit{mcrA} genes. Each read was translated into six reading frames using OrfF (https://github.com/wwood/OrfM) and only reads with an e-value < 1 \textit{e}\textsuperscript{-10} were retained. Homology comparison between \textit{V1} \textit{mcrA} reads and a database consisting of \textit{mcrA} sequences from \textit{Verstraetearchaeota}, \textit{Bathyarchaeota} and \textit{Euryarchaeota} was performed using Blastp v2.2.30\textsuperscript{+} across the IMG database v4.1, Pfam domains\textsuperscript{43}, and \textit{mcrA} genes. The consensus annotation was used for metabolic reconstruction. Additionally, a selection of 247 publically available genomes in the NCBI database belonging to \textit{Euryarchaeota}, \textit{Crenarchaeota}, \textit{Thaumarchaeota}, \textit{Bathyarchaeota} and \textit{Verstraetearchaeota} were annotated by searching against Uniref100 (accessed October 2015) using the ‘Blast’ function in Diamond (https://github.com/bbuchfink/diamond.g) for each gene with an e-value < 1 \textit{e}\textsuperscript{-10} and used as an outgroup, and modified for publication in Adobe Illustrator, and nodes with <75% bootstrap support were collapsed in the McrABG trees.

Screening of SRA using \textit{Verstraetearchaeota} 16S \textit{rRNA} sequences. The collection of metagenomes downloaded from the SRA database was mapped to the GreenGenes database (ref. 43) v2013_08 99% dereplicated database of 16S rRNA sequences. It also formed the basis for the heatmap showing presence of absence of genes with the Vegan\textsuperscript{56} and RColorBrewer\textsuperscript{57} packages in RStudio v3.2.0.

Comparison of MCR complex sequence active site conservation and conformation. The tertiary structures of McrA, McrB and McrG subunits from the \textit{Methanopyrus kandleri} crystal structure (PDB ID: 1E6V). Superimposing the crystal structure of the AD1 MCR onto the \textit{Methanopyrus} \textit{kandleri} MCR crystal structure (PDB ID: 1E6V). Superimposing the crystal structure of the AD1 MCR onto the \textit{Methanopyrus} \textit{kandleri} MCR crystal structure resulted in global structural alignment scores (TM-scores) for McrA, McrB and McrG of 98.1, 98.2 and 98.3, respectively. The root-mean-square deviations (RMSDs) of the TM-aligned residues for McrA, McrB and McrG of 0.98.1, 0.98.2 and 0.98.3, respectively. The root-mean-square deviations (RMSDs) of the TM-aligned residues for McrA, McrB and McrG were 0.63, 0.51 and 0.86 Å, respectively.

Data availability. The sequences have been deposited at the NCBI Sequence Read Archive (Whole Genome Submission) under BioProject ID PRJNA321438 with accession nos. MAG000000000 (V1), MAG000000000 (V2), MAG000000000 (V3), MAG000000000 (V4) and MAG000000000 (V5).

Received 16 May 2016; accepted 12 August 2016; published 3 October 2016

References

1. Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. \textit{Nat. Rev. Microbiol.}, 6, 579–591 (2008).

2. Hedderich, R. & Whitman, W. B. in The Prokaryotes Vol. 4 (eds Rosenberg, E. et al.) 662 (Springer, 2013).
3. Thauer, R. K. et al. Hydrogenases from methanogenic Archaea, nickel, a novel cofactor and H2 exchange: Annu. Rev. Biochem. 79, 507–536 (2010).

4. IPCC Climate Change 2014: Synthesis Report (eds Core Writing Team, Pachauri, R.K. & Meyer, L.A.) (Cambridge Univ. Press, 2014).

5. Lang, K. et al. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of Candidatus Methanoplasma termophilum. Appl. Environ. Microbiol. 81, 1338–1352 (2015).

6. Baptiste, E., Brochier, E. & Boucher, V. Higher-level classification of the archaea: evolution of methanogenesis and methanogens. Archaea 1, 353–363 (2005).

7. Fournier, G. P. & Gogarten, J. P. Evolution of acetoclastic methanogenesis in Methanosarcina via horizontal gene transfer from cellulosytic Clostridia. J. Bacteriol. 190, 1124–1128 (2008).

8. Rotaru, A.-E. et al. A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to Methanoseta for the reduction of carbon dioxide to methane. Energy Environ. Sci. 7, 408–415 (2014).

9. Borrel, G. et al. Comparative genomics highlights the unique biology of Methanomassilicaicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics 15, 1–23 (2014).

10. Nobu, M. K., Naritiro, T., Kuroda, K., Mei, R. & Liu, W.-T. Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME J. 10, 2478–2487 (2016).

11. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. Phil. Trans. R. Soc. Lond. B 361, 1007–1022 (2006).

12. Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyrarchaeota revealed by genome-centric metagenomics. Science 350, 434–438 (2015).

13. Boes, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood–Ljungdahl pathway: an ancient, versatile and fragile association. Genome Biol. Evol. 8, 1706–1716 (2016).

14. Vanwonterghem, I. et al. Methane metabolism in the archaeal phylum Bathyarchaeota and the methanogen archaeon isolated from methanogenic sludge. Int. J. System. Evol. Microbiol. 65, 554–560 (2015).

15. Konstantinidis, K. T. & Tiedje, J. M. Towards a genome-based taxonomy for Archaea and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072 (2006).

16. Hocking, W. P., Stokke, R., Roalkvam, I. & Steen, I. H. Identification of key components in the energy metabolism of the hyperthermophilic sulphate-reducing archaeon Archaeoglobus fulgidus by transcriptome analyses. Front. Microbiol. 5, 95 (2014).

17. Sato, T., Atomi, H. & Imanaka, T. Archaeal type III RuBisCOs function in a redox-driven proton pump closely related to NADH dehydrogenases. J. Mol. Biol. 325, 1797–1810 (2003).

18. Siddique, T. et al. Phylogenetic diversity of microbial communities associated with crude-oil large insoluble particle and formation-water components of the reservoir fluid from a non-flooded high-temperature petroleum reservoir. J. Biotechnol. 173, 204–210 (2012).

19. Konstantinidis, K. T. & Tiedje, J. M. Genomic evidence for distinct carbon substrate preferences and complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29, 222–226 (2014).

20. Yang, J., Roy, A. K. & Zhang, Y. Protein-ligand binding site recognition using a new model for electron flow during anaerobic digestion: direct interspecies electron transfer to Methanoseta for the reduction of carbon dioxide to methane. Energy Environ. Sci. 7, 408–415 (2014).

21. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

22. Novozamsky, K. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

23. Greeneng, A. M. et al. Expansion of genomic representation of the phylum Cyanobacteria. Genome Biol. Evol. 6, 1031–1045 (2014).

24. Nascichov, E. P. Structural RNA Homology Search and Alignment Using Covariance Models PhD thesis, Washington Univ. Saint Louis (2009).

25. Markowitz, V. M. et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40, 115–122 (2012).

26. Seemann, T. & Prokka, rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).

27. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, 222–226 (2014).

28. Finn, R. D. et al. The Pfam protein families database. Nucleic Acids Res. 42, 103–110 (2014).

29. Kanehisa, M. & Goto, S. KEGG Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 38, 142–147 (2010).

30. Markowitz, V. M. et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40, 115–122 (2012).

31. Nascimento, K. et al. An expanded genomic representation of the phylum Cyanobacteria. Genome Biol. Evol. 6, 1031–1045 (2014).

32. Seemann, T. & Prokka, rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).

33. Marchler-Bauer, A. et al. CDD: NCBI’s conserved domain database. Nucleic Acids Res. 43, 222–226 (2014).

34. Finn, R. D. et al. The Pfam protein families database. Nucleic Acids Res. 42, 103–110 (2014).

35. Kanehisa, M. & Goto, S. KEGG Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 38, 27–30 (2000).

36. Nascimento, K. et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, 199–205 (2014).

37. Greeneng, A. M. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

38. Roy, A. K. & Zhang, Y. I-TASSER A unified platform for automated protein structure and function prediction. Nat. Protoc. 9, 722–738 (2010).

39. Yang, J., Roy, A. K. & Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29, 2588–2595 (2013).

40. Ermel, U., Grabarse, W., Shima, S., Goubeaud, M. & Thauer, R. K. Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. Science 278, 1457–1462 (1997).

Acknowledgements

This study was supported by the US Department of Energy’s Office of Biological Environmental Research (award no. DE-SC0010574), and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Flagship Cluster ‘Biotechnological Solutions to Australia’s Transport, energy and greenhouse gas challenges’ I.V. acknowledges support from the University of Queensland International Scholarship, and P.D.J. acknowledges support from the Australian Meat Processor Corporation (2013/0408 Technology Fellowship). P.H. and I.V. acknowledge support from the Australian Research Council Laureate Program entitled “Reconstructing the universal tree and network of life”.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
G.W.T. acknowledges support by the University of Queensland Vice-Chancellor’s Research Focused Fellowship. The authors thank the sequencing facilities at the Australian Centre for Ecogenomics and Institute for Molecular Biosciences (University of Queensland, Australia) for library preparation and sequencing.

Author contributions
I.V. designed and performed the anaerobic digestion study, analysed the data and wrote the paper. P.E. assisted with the metabolic analysis and performed protein modelling. D.P. assisted with the binning of the formation water metagenomes and the mcrA screening of the SRA database. P.D.J. was involved in the design of the anaerobic digestion study and helped run the reactors. B.J.W. performed the binning of the palm oil mill effluent metagenome and assisted with the 16S rRNA gene screening of the SRA database. P.H. helped interpret the phylogenetic analyses and wrote the paper. G.W.T. oversaw the project, was involved in the design of the anaerobic digestion study, helped interpret the data and wrote the paper.

Additional information
Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.W.T.

Competing interests
The authors declare no competing financial interests.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/