FEMS Microbiology Ecology Advance Access published November 26, 2015

Phylogenetic and genomic analysis of *Methanomassiliicoccales* in wetlands and animal intestinal tracts reveals clade-specific habitat preferences.

Andrea Söllinger[1], Clarissa Schwab[2], Thomas Weinmaier[3], Alexander Loy[3], Alexander T. Tveit[4], Christa Schleper[1], Tim Urich[1,5*]

[1] Department of Ecogenomics and Systems Biology, University of Vienna, Austria

[2] Institute of Food, Nutrition and Health, ETH Zürich, Switzerland

[3] Department of Microbiology and Ecosystem Science, University of Vienna, Austria

[4] Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway

[5] Institute for Microbiology, Ernst-Moritz-Arndt University Greifswald, Germany

[*] corresponding author, Institute for Microbiology, Ernst-Moritz-Arndt University Greifswald, Friedrich-Ludwig-Jahn-Str. 15, 17489 Greifswald, Germany, T: +49-3834-864203, tim.urich@uni-greifswald.de

Abstract

Methanogenic *Thermoplasmata* of the novel order *Methanomassiliicoccales* were recently discovered in human and animal gastro-intestinal tracts (GITs). However their distribution in other methanogenic environments has not been addressed systematically. Here we surveyed *Methanomassiliicoccales* presence in wetland soils, a globally important source of methane emissions to the atmosphere, and in the GITs of different animals by PCR targeting their 16S rRNA and methyl:coenzyme M reductase (α-subunit) genes. We detected *Methanomassiliicoccales* in all 16 peat soils investigated, indicating their wide distribution in these habitats. Additionally, we detected their genes in various animal feces. *Methanomassiliicoccales* were subdivided in two broad phylogenetic clades designated 'environmental' and 'GIT' clades based on differential, although non-exclusive, habitat preferences of their members. A well-supported cluster within the environmental clade comprised more than 80 % of all wetland 16S rRNA gene sequences. Metagenome assembly from bovine rumen fluid enrichments resulted in two almost complete genomes of both *Methanomassiliicoccales* clades. Comparative genomics revealed that members of the environmental clade contain larger genomes and a higher number of genes encoding anti-oxidative enzymes than animal GIT clade representatives. This study highlights the wide distribution of *Methanomassiliicoccales* in wetlands, which suggests that they contribute to methane emissions from these climate-relevant ecosystems.

Keywords: methanogens, archaea, peat soil, methylotrophic methanogenesis, enrichments, metagenomics

Introduction

Methanogenic archaea (methanogens) produce the bulk of biologically generated methane (CH_4) on Earth, and are thus the largest global source of this potent greenhouse gas (Ciais *et al*. 2013). Methanogens occur in a wide range of predominantly anoxic environments including wetlands, marine and freshwater environments, landfills, and intestinal tracts of humans and various animals (Ciais *et al*. 2013). Natural wetlands such as peat soils cover less than 9% of the global land surface (Zedler and Kercher 2005), but are the major natural CH_4 source, contributing approximately 20 - 30 % to global CH_4 emissions to the atmosphere (Ciais *et al*. 2013). However, 50 - 65 % of all CH_4 emitted to the atmosphere results from human activities such as animal husbandry. Enteric fermentation by ruminant animals is one major CH_4 source, contributing approximately 30 % to all anthropogenic CH_4 emissions (Ciais *et al*. 2013).

All known methanogens belong to the phylum of *Euryarchaeota* and were until lately classified into six orders, the *Methanomicrobiales*, *Methanobacteriales*, *Methanosarcinales*, *Methanococcales*, *Methanocellales* and *Methanopyrales*. An additional order of methanogens was recently proposed within the class *Thermoplasmata* (Paul *et al.* 2012) and was confirmed by the isolation of *Methanomassiliicoccus luminyensis* (Dridi *et al.* 2012), the first and still only isolate from this 7th methanogen order, the *Methanomassiliicoccales* (Oren and Garrity 2013). Several recent studies showed that *Methanomassiliicoccales* have an energy metabolism distinct from other methanogens. All currently published enrichment cultures and the sole isolate *M. luminyensis* were obtained on methanol and H_2 (Borrel *et al.* 2012; Dridi *et al.* 2012; Paul *et al.* 2012; Borrel *et al.* 2013a; Iino *et al.* 2013). More recently, methylamines were identified as electron acceptors besides methanol (Poulsen *et al*. 2013; Brugere *et al.* 2013; Lang *et al.* 2014). Genomic analyses showed that all *Methanomassiliicoccales* genomes analysed so far encode a truncated methanogenesis pathway (Lang *et al.* 2014) via H_2-

dependent methylotrophic methanogenesis (Borrel *et al.* 2013b; Borrel *et al.* 2014; Lang *et al.* 2014).

Methanomassiliicoccales affiliated 16S rRNA gene sequences have been recovered from natural and man-made anoxic habitats, e.g. in rice paddy fields (Großkopf *et al.* 1998), anaerobic digestors (Godon *et al.* 1997) and in gastro-intestinal tracts (GIT) of various ruminant and non-ruminant animals, such as cattle (Tajima *et al.* 2001), sheep (Wright *et al.* 2004), reindeer (Sundset *et al.* 2009), yak (Huang *et al.* 2012) and wallaby (Evans *et al.* 2009). Paul *et al*. (2012) were the first to note an environment-specific clustering of *Methanomassiliicoccales* 16S rRNA gene sequences, with the observation of an animal-associated 'intestinal cluster' distinct from environmental sequences. Curiously, the type strain *M. luminyensis,* although obtained from human feces, does not belong to the 'intestinal cluster' but is associated with the environmental group, leaving the 'intestinal cluster' poorly characterised due to the lack of isolates. *Methanomassiliicoccales* of the 'intestinal cluster' have been shown to constitute a large proportion of the archaeal population in GITs of ruminant animals (Gu *et al.* 2011; St-Pierre and Wright 2013) as revealed by 16S rRNA gene-based clone library analyses.

The occurrence of *Methanomassiliicoccales* in wetlands, the major natural CH_4 source, has not been systematically assessed yet. This study aimed to reveal the environmental distribution of *Methanomassiliicoccales* in a range of different wetland types, spanning acidic and neutral peatlands from temperate and arctic regions. For this purpose a PCR screening for *Methanomassiliicoccales* 16S rRNA genes and *mcrA*, encoding the α-subunit of the methanogenesis key enzyme methyl:coenzyme M reductase (Mcr), as phylogenetic and functional marker genes was conducted. The screening was complemented with samples from animal intestinal tracts to enable the assessment of environment-specific patterns of *Methanomassiliicoccales* occurrence. We show that distinct *Methanomassiliicoccales* clades

dominate in wetlands compared to animals. Furthermore, we report the enrichment of rumen *Methanomassiliicoccales* to approximately 30 % relative abundance and an accompanying enrichment bias that favours close relatives of *M. luminyensis*. We additionally provide insights into clade-specific genomic adaptations by comparative genomics, including two new draft genomes that were reconstructed from metagenomic data of the rumen enrichment cultures.

Materials and methods

Wetland and animal samples

Origin and characteristics of wetland samples are given in Table 1. The two arctic peat samples were from Svalbard (Tveit *et al.* 2012) and DNA extracts from 14 temperate wetland samples were from Germany, Austria and Italy (Steger *et al.* 2011). Kangaroo, elephant, giant tortoise, tortoise and rat fecal samples were collected on the 15th of February 2010 from the Viennese Zoo (Tiergarten Schönbrunn), flash-frozen in liquid nitrogen and stored at -80°C until DNA extraction. A bovine rumen fluid sample (AUT RF) was obtained from the University of Veterinary Medicine Vienna. After centrifugation (5 min, 95 g) to remove particulate material, microbial cells were harvested by centrifugation (5 min, 16.500 g) and frozen at -20 °C until usage. A rumen fluid sample (DK RF) was received from Aarhus University, Denmark. The rumen fluid was supplemented with 15 mM trimethylamine, incubated anoxically in serum bottles for 24 h at 38 °C and shipped at 4 °C for 2 d. Cells were harvested by centrifugation (5 min, 16.500 g) and frozen at -20 °C. A third rumen fluid sample (cow) derived from a previously described 24 h incubation experiment (Poulsen *et al.* 2013).

DNA extraction, clone library construction and RFLP

DNA from arctic peat and animal feces samples was extracted using a phenol/ chloroform/isoamyl alcohol bead beating protocol (Urich *et al.* 2008). QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA, USA) was used to extract DNA from Austrian and Danish bovine rumen fluid (AUT/DK RF) according to the manufacturer's instructions. While the usage of different DNA extraction methods likely affects the broad resulting bacterial and archaeal community composition, the reported dominant *Methanomassiliicoccales* phylotypes are likely unaffected. Their cell wall structure consisting of two membranes and no murein sacculus or S-layer should be disrupted even with the latter protocol not containing a bead beating step. DNA extracts were quality checked and quantified by agarose gel electrophoresis and NanoDrop® (ND-1000, Peqlab, Erlangen, Germany). Three different primer pairs were employed to amplify *Methanomassiliicoccales* 16S rRNA genes and *mcrA* of *Methanomassiliicoccales* and other methanogens (see Tab. 2 for details). PCR amplification was performed using a T Professional Thermocycler (biometra, Göttingen, Germany). Reaction mixtures (50 µL) contained 10 µL 5x Green GoTaq® Flexi Buffer, 4 µL MgCl (2 mM), 1 µL dNTPs (0.2 mM), 0.25 µL bovine serum albumine (0.2 $mgmL^{-1}$), 1 µL of each primer (0.5 µM), 0.25 µL DreamTaq (0.025 U μL^{-1}) and 1 µL DNA template. PCRs with AS1/AS2 and mlas-mod For/ mcrA-rev primer pairs were performed with 34 cycles and 45 seconds of annealing (60 °C) and elongation (72 °C). PCR using Tp-mcrA-F/Tp-mcrA-R primers were performed with 36 cycles and 60 seconds of annealing (60 °C) and 45 seconds of elongation (72 °C). Amplified products were quality checked via agarose gel electrophoresis and purified (QIAquick PCR Purification Kit, QIAGEN, Valencia, CA, USA). PCR products were cloned using pGEM®-T Vector System I (Promega, Madison, WI, USA) and chemically competent *E. coli* cells (One Shot® TOP10 Chemically Competent *E. coli*, Invitrogen, Carlsbad, CA, USA). Ligation reactions were incubated one hour at room temperature or overnight at 4 °C until transformation. Subsequently, the cells were plated on LB agar plates containing ampicillin (100 mgL^{-1}), IPTG (0.5 mM) and X-Gal (80 $mgmL^{-1}$),

and incubated overnight at 37 °C. Recombinant *E. coli* colonies were identified by blue-white screening. Colony PCR using T7/SP6 primers was done to identify correct insert sizes with 32 cycles of PCR and 45 seconds of annealing (55 °C) and elongation (72 °C),. More than 200 16S rRNA gene clones and nearly 200 *mcrA* clones (127 and 71 obtained with the universal and the specific *mcrA* primer, respectively) were screened (8 to 16 16S rRNA and 8 to 32 *mcrA* clones per wetland and animal sample). Quality check and purification were done as described above. RFLP analysis of 16S rRNA and *mcrA* PCR products was done with the enzyme HpyCH4V (New England BioLabs, Ipswich, MA, USA). The restriction reaction was performed at 37 °C (3 h) and resulting fragments were visualised by electrophoresis on 3 % agarose gels.

Sequencing and sequence analysis

Sanger sequencing of clones that were representative for the different RFLP patterns was performed by LGC genomics (Berlin, Germany). Contaminating vector sequences were removed using NCBI VecScreen (http://www.ncbi.nlm.nih.gov/tools/vecscreen/). 16S rRNA and *mcrA* gene sequences were verified by megablast against NCBI Nucleotide collection (nr/nt) and by blastx against NCBI Non-redundant protein (nr) database with default settings, respectively (Altschul *et al.* 1990; Camacho *et al.* 2009; Sayers *et al.* 2011).

Representative 16S rRNA gene sequences were clustered into operational taxonomic units (OTUs) at ≥98 % sequence identity using cd-hit-est of the CD-HIT Suite (Huang *et al.* 2010). *McrA* gene sequences were translated into amino acid sequences using BioEdit (Hall 1999) and clustered at ≥98 % amino acid identity. 16S rRNA gene sequences were aligned with a set of representative reference sequences of *Thermoplasmata* and *Methanomassiliicoccales* 16S rRNA genes using SINE online (Pruesse *et al*. 2012). The reference sequences were selected form the SILVA Ref NR SSU r123 database (Quast *et al.* 2013); high quality sequences with a sequence length of >1250 nt were clustered at 94.5 % sequence identity using cd-hit-est of

the CD-HIT Suite (Huang *et al.* 2010) to obtain representative sequences of potential *Methanomassiliicoccales* and *Thermoplasmata* genera (Yarza *et al*. 2014). Inferred McrA amino acid sequences were aligned with a set of reference MrcA amino acid sequences of methanogens derived from Poulsen *et al*. (2013) and other studies on *Methanomassiliicoccales* (minimum sequence length of 549 AA) using MAFFT v6.864, L-INS-i (Katoh and Toh 2008). The McrA amino acid sequence alignment was improved by gap removal with Gblocks (Talavera and Castresana 2007) using the least stringent parameters to avoid losing phylogenetic information. Phylogenetic reference trees (16S rRNA gene and McrA) were constructed using PhyML (Guindon and Gascuel 2003), a maximum likelihood method implemented in ARB (Ludwig *et al*. 2004), and bootstrap values were calculated (500 replicates each) to verify branch support. To select the best suited nt and AA substitution model the respective alignments were uploaded to the Model Selection tool of the IQ-TREE web server (Minh *et al*. 2013). The 16S rRNA reference tree was then calculated based on the GTR model and the McrA reference tree was constructed based on the Dayhoff model. Afterwards, the shorter *Methanomassiliicoccales* 16S rRNA gene and McrA protein sequences obtained in this study were placed into the reference trees using the ARB Maximum Parsimony tool (Ludwig *et al*. 2004). FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used for visualizing phylogeny.

Methanomassiliicoccales **enrichments from cow rumen**

Rumen *Methanomassiliicoccales* enrichments were initiated with rumen fluid of a fistulated Braunvieh (Brown Swiss) fed on hay and dairy concentrate. After collection, cow rumen fluid was immediately centrifuged 5 minutes at 1000 rpm to remove large particles. Initially, serum bottles containing modified *Methanobrevibacter* medium (Dridi *et al.* 2012) were inoculated with 20 % rumen fluid and incubated under an atmosphere containing N_2, CO_2 and H_2 (70:20:10) at 37 °C in the dark. Trimethylamine (TMA) was supplemented as electron

acceptor (10 mM). Rumen *Methanomassiliicoccales* enrichments were obtained after successive transfers (weekly; 10 % inocula) on a medium containing (L^{-1}) 0.5 g KH_2PO_4, 0.4 g $MgSO_4 \cdot 7H_2O$, 5 g NaCl, 1 g NH_4Cl, 0.05 g $CaCl_2 .2H_2O$, 1.6 g sodium acetate, 0.5 g cysteine-HCl, 1 mL trace element solution, 2 mL tungstate-selenite solution, 4 g $NaHCO_3$, 0.5 g $Na_2S .9H_2O$, 2 g sodium formate, 1 mL vitamin solution. This adjusted medium was prepared similar to the medium used by Dridi *et al*. (2012), supplemented with 10 - 60 mM TMA and sterile filtered rumen fluid (7.5 % v/v), and incubated as described above. *Methanomassiliicoccales* growth was monitored via quantitative PCR (qPCR) on a Mastercycler ep realplex (Eppendorf, Hamburg, Germany), using the same protocol and standards as employed by Poulsen and co-workers (2013).

IonTorrent sequencing – (meta)genomic analysis

DNA from two enrichment cultures (RumEn_MG1 from 7th transfer culture [Ion 314™ Chip Kit]; RumEn_MG2 from 13th transfer culture [Ion 316™ Chip Kit]) was single-end sequenced using IonTorrent PGM (life technologies, Carlsbad, CA, USA) and 200 bp chemistry. 16S rRNA gene fragments contained in the two metagenomes were taxonomically classified with CREST (Lanzén *et al.* 2012). Assembly of RumEn_MG2 was done with Newbler v2.9 (454 Life Sciences, Branford, CT, USA). Taxonomic binning of RumEn_MG2 contigs was done with PhymmBL (Brady and Salzberg 2009), after training on all available complete RefSeq genomes as of May 2013 and the two available *Methanomassiliicoccales* genomes *Candidatus* Methanomethylophilus alvus Mx1201 and *M. luminyensis* B10. The PhymmBL classification resulted in two *Methanomassiliicoccales* genome bins (referred to as RumEn M1 and RumEn M2 from now on). The partial genomes were functionally annotated with RAST using default settings (Aziz *et al.* 2008). RAST annotations of genes of interest (e.g. methanogenesis, energy and carbon acquisition and environmental adaptation) were verified using the BLAST tool implemented in RAST and by comparison with annotated

genes from other *Methanomassiliicoccales*, consulting different databases, e.g. Pfam (Finn *et al.* 2013), UniProt (UniProt Consortium 2014), and TCDB (Saier *et al.* 2013). tRNAs were predicted using tRNAscan-SE (Lowe and Eddy 1997) and ARAGORN (Laslett and Canback 2004). Genome completeness and contamination of both *Methanomassiliicoccales* genome bins was assessed using checkM with default settings (Parks *et al.* 2014).

Growth experiments with glycine betaine

Methanomassiliicoccales growth on glycine betaine (Sigma-Aldrich, St. Louis, MO, USA) was tested on *M. luminyensis* (DSMZ; DSM no. 25720). *M. luminyensis* was grown in a bicarbonate buffered medium based on Paul *et al.* (2012) under an atmosphere containing 20 % CO_2 and 80 % H_2. Serum bottles containing medium supplemented either with 16 mM glycine betaine, 60 mM glycine betaine or 50 mM methanol (positive control) as electron acceptors were inoculated with 10 % (v/v) of stationary phase *M. luminyensis* cultures that were grown on 50 mM methanol. These cultures and three negative controls (w/o electron acceptor, w/o inoculum but 50 mM methanol, w/o inoculum but 60 mM glycine betaine) were incubated at 37° C in the dark (three replicates each). Growth was regularly monitored during 46 days of incubation by OD measurements (DU 800 Spectrophotometer, Beckman Coulter, Pasadena, CA, USA) at 578 nm and microscopic cell counts.

Data submission

Representative sequences of OTUs were submitted to GenBank and are deposited under the accession number KT225396 - KT225423. The *Methanomassiliicoccales mcrA* gene sequences are available under the accession number KT225424 - KT225458. Annotated draft genomes are available under the accession numbers LJKK00000000 (RumEn M1) and LJKL00000000 (Rumen M2). Raw metagenomics sequencing data were deposited at the NCBI Sequence Read Archive (accession number SRP064292).

Results

Distribution of *Methanomassiliicoccales* in wetlands and animals

A screening of 16 arctic and temperate wetland samples with *Methanomassiliicoccales* -specific PCR detected *Methanomassiliicoccales* 16S rRNA genes in all samples (Fig. 1 & 2). *Methanomassiliicoccales* 16S rRNA genes were additionally detected in faeces from elephant, giant tortoise, tortoise and rat and in bovine rumen fluid, but not in kangaroo feces. Between 8 and 16 clones of each of the 22 clone libraries were subjected to RFLP analysis and clones with representative RFLP patterns were sequenced. The analysis revealed a low diversity of *Methanomassiliicoccales* with e.g. 10 out of 16 peat clone libraries containing only 1 or 2 species-level OTUs (<98 % sequence identity). Since our aim was the identification of dominant *Methanomassiliicoccales* OTUs in many samples and not an exhaustive characterisation of diversity patterns, we did not expand this analysis by using deep sequencing of *Methanomassiliicoccales* 16S rRNA gene amplicons. The 138 representative Sanger sequences that were obtained in total formed 28 distinct OTUs. Twenty-seven OTUs were affiliated with *Methanomassiliicoccales*, while one OTU belonged to *Thermoplasmata* of the marine benthic group D (MBG-D, Fig. 1). Seven OTUs, which were dominated by wetland clones, formed a monophyletic group with *M. luminyensis, Ca*. M. intestinalis, and 16S rRNA gene sequences from different aquatic and terrestrial environments, and is hereafter referred to as 'environmental clade' (Fig. 1). OTUs 1, 3 and 10 formed a narrow cluster that encompassed approximately 80 % of all wetland 16S rRNA gene sequences ('wetland cluster', Fig. 1), thus typifying the *Methanomassiliicoccales* that inhabit a wide range of peat environments. The sequences retrieved from animals were all but one from cow assigned to a second distinct sequence cluster comprising *Ca*. M. alvus, *Ca*. M. caenicola, *Ca*. M. termitum and sequences obtained from other ruminant and non-ruminant animals (Fig. 1).

This group corresponded to the animal associated clusters identified previously by Paul *et al.* (2012) and Borrel *et al.* (2013b) and is here referred to as 'gastro-intestinal tract (GIT)' clade. Only twelve percent (n=15) of the wetland *Methanomassiliicoccales* sequences were members of the GIT clade. The observed preferential, although non-exclusive, environmental distribution of members of the two clades is illustrated in Figure 2 that displays the relative 16S rRNA clone abundance of both clades in each peat and animal sample. The *Methanomassiliicoccales* populations in all peat samples, independent of wetland type, latitude and pH, were dominated by the environmental clade, with the exception of sample AUT peat 7 (Fig. 2). The latter was an atypical peat sample, since it originated from a ferrous biofilm (Steger *et al.* 2011). All animal samples were dominated by GIT clade *Methanomassiliicoccales* and only one animal sample (cow) contained an environmental clade clone (Fig. 2).

A PCR screening for the functional marker gene of methanogens detected *Methanomassiliicoccales* affiliated *mcrA* sequences in both arctic peat soils (Fig. 1), in four temperate peat soils and in five animals (elephant, giant tortoise, tortoise, rat feces and cow rumen fluid). Again, RFLP analysis was applied and the translated 53 representative *Methanomassiliicoccales* McrA amino acid sequences clustered in 24 distinct OTUs (< 98 % amino acid identity). Analogous to the 16S rRNA gene tree, the McrA tree showed a well-supported separation of *Methanomassiliicoccales* McrA sequences into environmental and GIT clades (Fig. 1). These results are in congruence with previous literature (Paul *et al*. 2012, Borrel *et al*. 2013b). Furthermore, a cluster exclusively comprised of wetland McrA sequences was observed containing 9 out of 14 wetland *Methanomassiliicoccales* McrA sequences, indicative of a wetland cluster. However due to the lack of reference sequences bootstrap support for this cluster could not be given. To obtain insights into the relative abundance of *Methanomassiliicoccales* among methanogen communities in wetlands we

analysed metatranscriptome libraries from the arctic wetlands Knudsenheia and Solvatn (Tveit *et al.* 2014). We screened for rRNA transcripts of MCC and methanogens and found a low relative abundance of MCC (between 0 and 4.7 % of methanogen rRNA transcripts; Tab. 3).

Rumen *Methanomassiliicoccales* enrichment cultures

We conducted enrichment trials with bovine rumen fluid and TMA to obtain novel *Methanomassiliicoccales* isolates of the GIT clade. Quantitative PCR showed that bacteria dominated the prokaryotic community in the rumen fluid (>10^9 bacterial 16S rRNA gene copies mL^{-1} rumen fluid), while methanogens were less abundant (approximately 2 x 10^7 16S rRNA gene copies mL^{-1}), see Figure 3A. Approximately 40 % of all 16S rRNA gene copies from methanogens belonged to the *Methanomassiliicoccales*. All *Methanomassiliicoccales* 16S rRNA clones obtained from the initial rumen fluid (AUT RF; 26 clones) were assigned to the GIT clade (Fig. 1; Fig. 3A). Stable enrichments of up to 32 % *Methanomassiliicoccales* were obtained after successive transfers of subcultures (see material and methods section) with maximal relative abundance of 51 % (data not shown), while other rumen methanogens were not enriched (data not shown). Despite the employment of various anaerobic isolation approaches such as serial dilutions, filtrations, antibiotic treatments and 'roll-tube method', no pure culture of rumen *Methanomassiliicoccales* was obtained.

Two metagenomes, RumEn_MG1 (7^{th} transfer) and RumEn_MG2 (13^{th} transfer), were sequenced from enrichments, consisting of 54 Mbp (482,803 sequences of 114 bp mean length) and 239 Mbp (1,871,707 sequences of 128 bp mean length) sequence information, respectively. Taxonomic classification of the metagenomic 16S rRNA gene fragments showed *Methanomassiliicoccales* 16S rRNAs to be 20 % (RumEn_MG1) and 24 % (RumEn_MG2) of all 16S rRNA gene fragments. The classification also revealed differences in bacterial and *Methanomassiliicoccales* community composition between the enrichments (Fig. 3).

Remarkably, the initially prevalent GIT clade decreased in relative abundance to 31 % of *Methanomassiliicoccales* and the environmental clade became dominant in the enrichments (Fig. 3).

Assembly and analysis of *Methanomassiliicoccales* draft genomes

Newbler assembly of RumEn_MG2 resulted in 8,854 contigs (>500bp) of 10.71 Mbp of average contig length 1210 bp with the largest contig being 202,459 bp in length. Binning with PhymmBL resulted in two partial *Methanomassiliicoccales* genome bins, RumEn M1 (182 contigs, 2.21 Mbp) and RumEn M2 (18 contigs, 1.28 Mbp). Table 4 shows some characteristics of RumEn M1 and RumEn M2 and published *Methanomassiliicoccales* genomes. Analysis of their full-length 16S rRNA genes revealed 89 % 16S rRNA gene sequence identity confirming the distant relationship of M1 and M2. M1 belonged to the environmental clade (Fig. 1). The 96 % sequence identity to the 16S rRNA gene of *M. luminyensis* and *Ca*. M. intestinalis suggested that M1 represents a novel member of the family *Methanomassiliicoccaceae*, while M2 represents a novel member of the GIT clade (95 % seq. id. to *Ca*. M. termitum and 91 % seq. id. to *Ca*. M. alvus; Fig. 1). Both were with 87 % and 92 % 16S rRNA gene sequence identity also distantly related to BRNA1, another *Methanomassiliicoccales* genome obtained from a rumen enrichment culture (GenBank acc. no.: CP002916).

Genome completeness estimates with checkM based on presence/absence of single-copy and multi-copy marker genes revealed that both genomes were almost complete (M1: 96.6 % and M2: 94.1 %) and showed no or only low level of contamination (M1: 1.6 %; M2: 0 %). The completeness was also reflected by the number and type of encoded tRNA and tRNA synthetase genes (Tab. 4). No cysteine and tryptophan specific tRNAs were found in M1. Both genomes contained a nearly full set of 21 aminoacyl tRNA synthetase genes

338 (20 common and one pyrrolysine specific), with $tRNA^{Ala}$ synthetase missing in M1 and
339 $tRNA^{Pyl}$ synthetase missing in M2.

340 Both draft genomes possessed all genes for enzymes involved in the last step of
341 methanogenesis, the reduction of methyl-coenzym M (CH_3-S-CoM) to CH_4 by
342 methyl:coenzyme M reductase (*mcrABGCD*). Furthermore, both genomes encoded a soluble
343 heterodisulfide reductase (HdrABC) and the associated methyl viologen-dependent
344 hydrogenase (MvhADG), required for the regeneration of coenzyme M. Like all other
345 *Methanomassiliicoccales*, both genomes encoded HdrD, one subunit of a membrane-bound
346 heterodisulfide reductase (HdrDE), and a 11-subunit Fpo-like complex, homologue to the 11
347 core subunits of a membrane-bound F_{420}-methanophenazine oxidoreductase complex found in
348 other methanogens (Moparthi and Hägerhäll 2011; Lang *et al.* 2014). However, one subunit
349 (FpoA) was missing in M2.

350 M1, contrary to M2, encoded *mtaBCA*, the complete gene set for reducing methanol to
351 methyl-coenzym M. Gene sets necessary for the utilisation of other methylated substrates
352 were entirely missing in the M2 genome and M1 encoded only an incomplete gene set
353 (*mtmBC*) for the reduction of monomethylamine to methyl-coenzym M. Like all methylamine
354 methyltransferases (Krzycki 2004), the monomethylamine methyltransferase of M1 is
355 predicted to contain pyrrolysine. Consistently, M1 contained all genes necessary for the
356 biosynthesis of pyrrolysine (*pylBCD*) as well as *pylS* and *pylT*, which encode the enzyme
357 catalysing the ligation of pyrrolysine to its specific tRNA (PylS) and the specific tRNA
358 ($tRNA^{Pyl}$). The pyrrolysine-associated gene set (*pylBCDST*) was entirely missing in M2.

359 M1 and M2 lacked all genes necessary for the reduction of CO_2 to CH_4 and both encoded an
360 ADP-forming acetyl-CoA synthetase (AscA) homologue, which allows heterotrophic growth
361 on acetate. These are two common features of all *Methanomassiliicoccales* genomes known
362 so far (Borrel *et al.* 2014; Lang *et al.* 2014).

363 RAST analysis revealed that M1 and M2 both contained a gene encoding a glycine betaine
364 ABC transporter, a homologue to the functional glycine betaine ABC transporter of
365 *Methanosarcina mazei*, encoded by *otaABC* (Schmidt *et al.* 2007). We tested growth of *M.*
366 *luminyensis* with glycine betaine as electron acceptor instead of methanol. However, no
367 growth, even with 60 mM glycine betaine, was observed (data not shown). Furthermore, M1
368 and M2 encoded several genes possibly involved in oxidative stress response. M1 encoded
369 four peroxiredoxin and rubrerythrin genes, two rubredoxin and superoxide reductase genes as
370 well as and one superoxide dismutase gene. In contrast, M2 encoded only two genes for
371 peroxiredoxin, rubrerythrin, and superoxide reductase, two catalase genes, and no rubredoxin
372 and superoxide dismutase genes.

373

374 # Discussion

375 ### *Methanomassiliicoccales* are widely distributed in wetlands

376 There has been no systematic study on the distribution of the recently described 7th
377 methanogen order *Methanomassiliicoccales* in wetlands. Our broadly conceived PCR
378 screening for *Methanomassiliicoccales* in various wetland types, including arctic and
379 temperate acidic and neutral peat soils, revealed the presence of *Methanomassiliicoccales* in
380 all tested samples (Fig. 2), which suggests their ubiquitous distribution in wetlands. Thus,
381 *Methanomassiliicoccales* contribute to the high diversity of methanogens in wetlands, that
382 include observations of *Methanosarcinales*, *Methanobacteriales*, *Methanomicrobiales*,
383 *Methanococcales* and *Methanocellales* in temperate wetlands (Nercessian *et al.* 1999; Upton
384 *et al.* 2000; Yavitt *et al.* 2012; Andersen *et al.* 2013), and *Methanosarcinales*,
385 *Methanobacteriales Methanomicrobiales* and *Methanocellales* in arctic wetlands (Galand *et*
386 *al.* 2002; Hoj *et al.* 2005; Tveit *et al.* 2012), respectively.

There is currently few data on the relative abundance of *Methanomassiliicoccales* among methanogens in wetlands. A screening of five metatranscriptomic libraries from the arctic wetlands Knudsenheia and Solvatn (Tveit *et al.* 2014) for rRNA transcripts of *Methanomassiliicoccales* and methanogens revealed a rather low relative abundance of *Methanomassiliicoccales* among methanogens (Tab. 3). Furthermore, the general *mcrA* primer pair used in this study, targeting all methanogens, amplified mostly *mcrA* sequences associated with *Methanomicrobiales*, *Methanocellales* and *Methanosarcinales* (data not shown), suggesting that these methanogens dominated the wetlands while *Methanomassiliicoccales* were less abundant. This would be in line with many published studies on methanogenic archaea from wetlands (Steinberg and Regan 2008; Yrjälä *et al.* 2011; Yavitt *et al.* 2012; Deng *et al.* 2014; Mondav *et al.* 2014; He *et al.* 2015). One determinant for this might be the parallel dependency of *Methanomassiliicoccales* on two substrates (i.e. H_2 and methylated compounds), as recently suggested by (Lang *et al.* 2014). However, further studies on *Methanomassiliicoccales* abundances and activity in wetlands are required to elucidate their contribution to methanogenesis. In contrast, *Methanomassiliicoccales* comprise a large proportion of the methanogen community in ruminant animals (St-Pierre and Wright, 2013). Correspondingly, approximately 40 % of all methanogens detected in the bovine rumen fluid used in this study belonged to the *Methanomassiliicoccales*. The cause for this might be the stable conditions inside the animal host and the constant supply of both H_2 and methylated compounds. Given the approximately similar magnitudes of CH_4 emitted to the atmosphere from wetlands and ruminant life stock, these preliminary data might indicate that wetland *Methanomassiliicoccales* are contributing less to global CH_4 emissions compared to their relatives in animal GITs.

Clade-specific habitat preferences of *Methanomassiliicoccales*

411 The phylogenetic separation between the majority of wetland and animal associated
412 *Methanomassiliicoccales* 16S rRNA gene and McrA sequences (Fig. 1) revealed an
413 environmental clade representing the *Methanomassiliicoccaceae* and a second family-level
414 clade associated with the animal GIT. Our data from abundant clone library members show
415 that in particular one narrow cluster of OTUs within the environmental clade comprises the
416 majority of wetland *Methanomassiliicoccales*. The methanogenic nature of this wetland
417 cluster is supported by *Methanomassiliicoccales* McrA sequences forming a similarly
418 coherent cluster. Our wetland cluster is part of the 'lake pavin cluster' that was previously
419 mentioned by Borrel *et al*. (2013b). The prevalence of the wetland cluster in 15 out of 16
420 investigated samples, irrespective of wetland type, pH and latitude suggests that our screening
421 has identified the most abundant *Methanomassiliicoccales* in wetlands worldwide.

422 *Methanomassiliicoccales* sequences obtained from a great variety of ruminant and non-
423 ruminant animals form a GIT-specific *Methanomassiliicoccales* clade. However, the habitat
424 distribution of environmental and GIT *Methanomassiliicoccales* clades is non-exclusive, as
425 both *Methanomassiliicoccales* clades contained at least one clone obtained from the other
426 habitat. These results point at a co-occurrence of both clades, although environmental and
427 GIT clade *Methanomassiliicoccales* are dominating the respective habitat. This assumption is
428 strengthened by the results from the rumen enrichment trials where an originally low
429 abundant environmental clade *Methanomassiliicoccales* (represented by the genome bin M1)
430 was dominating the rumen enrichments after several passages.

431 **Enrichment bias favours environmental clade**

432 To our surprise, the two genome bins M1 and M2 from rumen fluid metagenomes represented
433 both *Methanomassiliicoccales* clades and the initially not detected environmental clade (M1)
434 became dominant. One reason for this might be an enrichment bias favouring the
435 environmental over the GIT clade. In fact, the composition of the enrichment medium was

very similar to the one used to isolate *M. luminyensis* from human feces, although human feces is reportedly dominated by members of the GIT clade (Mihajlovski *et al.* 2010; Vanderhaeghen *et al.* 2015; our own unpublished data). Since *M. luminyensis* does not belong to the GIT clade but is associated with the environmental clade, it is reasonable to assume that similar enrichment biases against the intestinal clade took place in both studies, enriching for a previously low abundant *Methanomassiliicoccales* of the *Methanomassiliicoccus* genus in our case. One could speculate that the smaller genome size of GIT clade *Methanomassiliicoccales* reflects the specialization to a very particular environment, which causes a lower competitiveness under the artificial enrichment conditions compared to *Methanomassiliicoccus*.

Methanomassiliicoccales **core metabolism**

The presence of genes for methanogenesis key enzymes (e.g. *mcrABG*, *hdrABC* and *mvhADG*) in both genomes confirmed the methanogenic nature of RumEn M1 and RumEn M2. In general our results agree with recent comparative genomic studies that have established a hydrogen-dependent methylotrophic methanogenesis for *Methanomassiliicoccales* (Borrel *et al.* 2013b; Borrel *et al.* 2014; Lang *et al.* 2014). Detailed metabolic schemes can be found in these earlier publications. The reoxidation of ferredoxin and generation of electrochemical membrane potential is possibly catalyzed by a ferredoxin:heterodisulfide oxidoreductase complex (Fpo-like complex + HdrD) in a recently proposed novel mode of energy conversion (Lang *et al*. 2014). The presence of *acsA* gene suggests that both M1 and M2 are likely chemolithoheterotrophs using acetate as carbon source.

However, the substrate spectrum of M1 and M2 for electron acceptors for methanogenesis remains somewhat obscure. In contrast to *M. luminyensis*, *Ca*. M. intestinalis and *Ca*. M. alvus, both partial genomes do not encode any genes necessary for the reduction of TMA to

methyl-coenzyme M, although the enrichments were incubated with TMA. In fact, no homologues of any gene involved in the reduction of methylated substrates to CH_4 were detected in the M2 genome. On the contrary, M1 encoded enzymes of methanol and monomethylamine reduction. However, these analyses are somewhat hampered by the fact that both genomes are not complete.

Genome differences and possible adaptations

The most evident difference between RumEn M1 (belonging to the environmental clade) and RumEn M2 (belonging to the GIT clade) was their divergent genome size. Despite similar genome completeness, M1 had a nearly 1 Mbp larger genome than M2. Available genome data show that the environmental clade genomes are consistently larger than the GIT clade genomes (Tab. 4), possibly indicative of massive genomic streamlining in the GIT clade. It appears that the

GIT clade is better adapted to life in rather nutrient-rich, stable GIT environments at the expense of versatility and competitiveness under variable environmental conditions. In contrast, the larger genome size within the environmental clade offers this versatility for life in highly fluctuating environments such as soils and sediments.

Borrel *et al.* (2014) suggested several specific adaptations of *Methanomassiliicoccales* to soil environments, i.e. diazotrophy, osmoprotection and a greater antioxidative capacity encoded in the genome of *M. luminyensis* compared to genomes derived from animal GITs. The latter show adaptations to the GIT environment, e.g. presence of resistance genes to bile salts in *Ca*. M. alvus genome (Borrel *et al.* 2014). Neither RumEn M1 nor RumEn M2 encoded a choloylglycine hydrolase gene like *Ca*. M. alvus, which is involved in bile salt resistance. Since bile salts are secreted to the small intestine of ruminants and not to the rumen (Bauman and Lock 2006), there might be no selective pressure for rumen *Methanomassiliicoccales* to keep this gene. M1 and M2 lack 4 out of the 6 previously defined core genes for N_2-fixation

conserved among all diazotrophs (Dos Santos *et al.* 2012), i.e. *nifDEKN* are missing, while homologues of *nifH* and *nifB* are present, suggesting that neither M1 nor M2 are able to fix nitrogen.

Both genomes encode a functional transporter for glycine betaine, a common osmoprotectant in plants and a known precursor of TMA in the rumen (Mitchell *et al.* 1979). Very recently, glycine betaine was identified as a direct substrate for methanogenesis in *Methanococcoides* ssp. (Watkins *et al.* 2014), prompting us to test this possibility with *Methanomassiliicoccales*. However, no growth of *M. luminyensis* with glycine betaine as electron acceptor was observed, suggesting that it is solely used as osmoprotectant. The *Methanomassiliicoccaceae*-related, environmental clade genome M1 encodes a higher antioxidative capacity, in terms of higher diversity and redundancy of genes encoding enzymes for detoxification of reactive oxygen species compared to M2. Neither RumEn M1 nor RumEn M2 encoded a choloylglycine hydrolase gene like *Ca*. M. alvus, which is involved in bile salt resistance. Since bile salts are secreted to the small intestine of ruminants and not to the rumen (Bauman and Lock 2006), there might be no selective pressure for rumen *Methanomassiliicoccales* to keep this gene. In summary, our comparative analysis of these two new *Methanomassiliicoccales* genomes could confirm only some of the proposed *Methanomassiliicoccales* adaptations to GIT and soil environments as suggested by Borrel *et al*. (2014).

Conclusions

This study establishes the wide distribution of a wetland-specific *Methanomassiliicoccales* cluster in many peat soils across Europe. Preliminary data suggest that *Methanomassiliicoccales* are present at low relative abundances among wetland methanogens, and therefore are probably less important contributors to atmospheric CH_4 emissions than

their relatives residing in the GIT of animal. Compared to the GIT clade, *Methanomassiliicoccales* of the environmental clade appear physiologically more flexible to cope with variable soil environment with its fluctuating moisture contents, temperatures and oxygen gradients. Furthermore, this study illustrates the well-known shortcomings of isolation attempts, by enriching for *Methanomassiliicoccales* that are not representative of the majority *in situ*. The cause for this might lie in a yet unidentified auxotrophy of the GIT clade, that has until now prevented their isolation in pure culture, despite many attempts. Nevertheless, cultivation attempts to obtain *Methanomassiliicoccales* isolates of the wetland cluster and GIT clade are indispensable to obtain a detailed understanding of their biology and habitat-specific adaptations.

Funding

This work was financially supported by a University of Vienna 'Focus of Excellence' grant to TU, by a scholarship of the University of Vienna for doctoral candidates (uni:docs) to AS and by the Austrian Science Fund (FWF, P23117-B17 to AL and FWF P25369-B22 to ChS).

Acknowledgements

We thank Ricardo Alves (University of Vienna, Austria) for help with phylogenetic tree constructions. We thank Michael Pester (University of Vienna, Austria) for kindly providing DNA extracts from temperate wetland samples. We thank the Viennese Zoo for providing animal fecal samples, and Morten Poulsen (University of Aarhus, Denmark), Kathrin Deckardt and Johannes Kohl (University of Veterinary Medicine, Vienna, Austria) for providing rumen fluid samples.

References

Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403-10.

Andersen R, Chapman SJ, Artz RRE. Microbial communities in natural and disturbed peatlands: A review. *Soil Biol Biochem* 2013; **57**: 979-94.

Aziz RK, Bartels D, Best AA, *et al*. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008; **9**: 75.

Bauman DE, Lock AL (2006) Concepts in lipid digestion and metabolism in dairy cows. *Proc. Tri-State Dairy Nutr. Conf* 2006; 1-14.

Borrel G, Harris HMB, Toney W, *et al*. Genome Sequence of "Candidatus Methanomethylophilus alvus" Mx1201, a Methanogenic Archaeon from the Human Gut Belonging to a Seventh Order of Methanogens. *J Bacteriol* 2012; **194**: 6944-45.

Borrel G, O'Toole PW, Harris HM, *et al*. Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol Evol* 2013a; **5**: 1769-80.

Borrel G, Harris HM, Parisot N, *et al*. Genome sequence of "Candidatus Methanomassiliicoccus intestinalis" Issoire-Mx1, a third Thermoplasmatales-related methanogenic archaeon from human feces. *Genome Announc* 2013b; **1**: e00453-13.

Borrel G, Parisot N, Harris HM, *et al*. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* 2014; **15**: 679.

Brady A, Salzberg SL. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* 2009; **6**: 673-76.

Brugere JF, Borrel G, Gaci N, *et al*. Archaebiotics: Proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* 2013; **5**: 5-10.

Camacho C, Coulouris G, Avagyan V, *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 421.

Ciais P, Sabine C, Bala G, *et al*. Carbon and Other Biogeochemical Cycles. In: IPCC (ed.). *Climate Change 2013: The Physical Science Basis.* Cambridge: Cambridge University Press, 2013.

Deng Y, Cui X, Hernández M, *et al*. Microbial Diversity in Hummock and Hollow Soils of Three Wetlands on the Qinghai-Tibetan Plateau Revealed by 16S rRNA Pyrosequencing. *PLoS One* 2014, DOI: 10.1371/journal.pone.0103115.

Dos Santos PC, Fang Z, Mason SW, *et al*. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 2012; **13**: 162.

Dridi B, Fardeau M-L, Ollivier B, *et al*. Methanomassiliicoccus luminyensis gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int J Syst Evol Microbiol* 2012; **62**: 1902-07.

Dridi B, Fardeau ML, Ollivier B, *et al*. Methanomassiliicoccus luminyensis gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int J Syst Evol Microbiol* 2012; **62**: 1902-7.

576 Evans PN, Hinds LA, Sly LI, *et al*. Community Composition and Density of Methanogens in
577 the Foregut of the Tammar Wallaby (Macropus eugenii). *Appl Environ Microbiol* 2009; **75**:
578 2598-602.

579 Finn RD, Bateman A, Clements J, *et al*. Pfam: the protein families database. *Nucleic Acids*
580 *Res* 2014; **42**: D222-30.

581 Galand PE, Saarnio S, Fritze H, *et al*. Depth related diversity of methanogen Archaea in
582 Finnish oligotrophic fen. *FEMS Microbiol Ecol* 2002; **42**: 441-49.

583 Godon J-J, Zumstein E, Dabert P, *et al*. Molecular microbial diversity of an anaerobic
584 digestor as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol*
585 1997; **63**: 2802-13.

586 Großkopf R, Stubner S, Liesack W. Novel euryarchaeotal lineages detected on rice roots and
587 in the anoxic bulk soil of flooded rice microcosms. *Appl Environ Microbiol* 1998; **64**: 4983-
588 89.

589 Gu MJ, Alam M, Kim SH, *et al*. Analysis of methanogenic archaeal communities of rumen
590 fluid and rumen particles from Korean black goats. *Anim Sci J* 2011; **82**: 663-72.

591 Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies
592 by maximum likelihood. *Syst Biol* 2003; **52**: 696-704.

593 Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program
594 for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999; **41**: 95-98.

595 He S, Malfatti SA, McFarland JW, *et al*. Patterns in Wetland Microbial Community
596 Composition and Functional Gene Repertoire Associated with Methane Emissions. *mBio*
597 2015; **6**: e00066-15.

598 Hoj L, Olsen RA, Torsvik VL. Archaeal communities in High Arctic wetlands at Spitsbergen,
599 Norway (78 degrees N) as characterized by 16S rRNA gene fingerprinting. *FEMS Microbiol*
600 *Ecol* 2005; **53**: 89-101.

601 Huang X, Tan H, Long R, *et al*. Comparison of methanogen diversity of yak (Bos grunniens)
602 and cattle (Bos taurus) from the Qinghai-Tibetan plateau, China. *BMC Microbiol* 2012; **12**:
603 237.

604 Huang Y, Niu B, Gao Y, *et al*. CD-HIT Suite: a web server for clustering and comparing
605 biological sequences. *Bioinformatics* 2010; **26**: 680-82.

606 Iino T, Tamaki H, Tamazawa S, *et al*. Candidatus Methanogranum caenicola: a Novel
607 Methanogen from the Anaerobic Digested Sludge, and Proposal of Methanomassiliicoccaceae
608 fam. nov and Methanomassiliicoccales ord. nov., for a Methanogenic Lineage of the Class
609 Thermoplasmata. *Microbes Environ* 2013; **28**: 244-50.

610 Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program.
611 *Brief Bioinformatics* 2008; **9**: 286-98.

612 Krzycki JA. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine
613 methyltransferases. *Curr Opin Chem Biol* 2004; **8**: 484-91.

614 Lang K, Schuldes J, Klingl A, *et al*. Comparative Genome Analysis of "Candidatus
615 Methanoplasma termitum" Indicates a New Mode of Energy Metabolism in the Seventh
616 Order of Methanogens. *Appl Environ Microbiol* 2014; AEM. 03389-14.

617 Lanzén A, Jørgensen SL, Huson DH, *et al*. CREST – Classification Resources for
618 Environmental Sequence Tags. *PLoS One* 2012, DOI: 10.1371/journal.pone.0049334.

619 Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in
620 nucleotide sequences. *Nucleic Acids Res* 2004; **32**: 11-16.

621 Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes
622 in genomic sequence. *Nucleic Acids Res* 1997; **25**: 0955-964.

623 Ludwig W, Strunk O, Westram R, *et al*. ARB: a software environment for sequence data.
624 *Nucleic Acids Res* 2004; **32**: 1363-1371.

625 Mihajlovski A, Doré J, Levenez F, *et al*. Molecular evaluation of the human gut
626 methanogenic archaeal microbiota reveals an age-associated increase of the diversity. *Environ*
627 *Microbiol Rep* 2010; **2**: 272-80.

628 Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap.
629 *Mol Biol Evol* 2013; **30**: 1188-95.

630 Mitchell AD, Chappell A, Knox K. Metabolism of betaine in the ruminant. *J Anim Sci* 1979;
631 **49**: 764-74.

632 Mondav R, Woodcroft BJ, Kim E-H, *et al*. Discovery of a novel methanogen prevalent in
633 thawing permafrost. *Nat Commun* 2014, DOI: 10.1038/ncomms4212.

634 Moparthi VK, Hägerhäll C. The evolution of respiratory chain complex I from a smaller last
635 common ancestor consisting of 11 protein subunits. *J Mol Evol* 2011; **72**: 484-97.

636 Nercessian D, Upton M, Lloyd D, *et al*. Phylogenetic analysis of peat bog methanogen
637 populations. *FEMS Microbiol Lett* 1999; **173**: 425-29.

638 Oren A, Garrity GM. List of new names and new combinations previously effectively, but not
639 validly, published. *Int J Syst Evol Microbiol* 2013; **63**: 3931-34.

640 Parks DH, Imelfort M, Skennerton CT, *et al*. CheckM: assessing the quality of microbial
641 genomes recovered from isolates, single cells, and metagenomes. *PeerJ PrePrints* 2015, DOI:
642 10.1101/gr.186072.114.

643 Paul K, Nonoh JO, Mikulski L, *et al*. "Methanoplasmatales," Thermoplasmatales-related
644 archaea in termite guts and other environments, are the seventh order of methanogens. *Appl*
645 *Environ Microbiol* 2012; **78**: 8245-53.

646 Petersen SO, Højberg O, Poulsen M, *et al*. (2014). Methanogenic community changes, and
647 emissions of methane and other gases, during storage of acidified and untreated pig slurry. *J*
648 *Appl Microbiol* 2014; **117**: 160-72 .

649 Poulsen M, Schwab C, Jensen BB, *et al*. Methylotrophic methanogenic Thermoplasmata
650 implicated in reduced methane emissions from bovine rumen. *Nat Commun* 2013;
651 10.1038/ncomms2432.

652 Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence
653 alignment of ribosomal RNA genes. *Bioinformatics* 2012; **28**: 1823-1829.

654 Quast C, Pruesse E, Yilmaz P, *at al*. The SILVA ribosomal RNA gene database project:
655 improved data processing and web-based tools. *Nucl. Acids Res* 2013; **41**: D590-D596.

656 Saier MH, Reddy VS, Tamang DG, *et al*. The transporter classification database. *Nucleic*
657 *Acids Res* 2013, DOI: 10.1093/nar/gkt1097.

658 Sayers EW, Barrett T, Benson DA, *et al*. Database resources of the national center for
659 biotechnology information. *Nucleic Acids Res* 2011; **39**: D38-51.

660 Schmidt S, Pfluger K, Kogl S, *et al*. The salt-induced ABC transporter Ota of the
661 methanogenic archaeon Methanosarcina mazei Go1 is a glycine betaine transporter. *FEMS*
662 *Microbiol Lett* 2007; **277**: 44-9.

663 St-Pierre B, Wright A-D. Diversity of gut methanogens in herbivorous animals. *Animal* 2013;
664 **7**: 49-56.

665 Steger D, Wentrup C, Braunegger C, *et al*. Microorganisms with Novel Dissimilatory
666 (Bi)Sulfite Reductase Genes Are Widespread and Part of the Core Microbiota in Low-Sulfate
667 Peatlands. *Appl Environ Microbiol* 2011; **77**: 1231-42.

668 Steinberg LM, Regan JM. Phylogenetic Comparison of the Methanogenic Communities from
669 an Acidic, Oligotrophic Fen and an Anaerobic Digester Treating Municipal Wastewater
670 Sludge. *Appl Environ Microbiol* 2008; **74**: 6663-71.

671 Steinberg LM, Regan JM. mcrA-Targeted Real-Time Quantitative PCR Method To Examine
672 Methanogen Communities. *Appl Environ Microbiol* 2009; **75**: 4435-4442.

673 Sundset MA, Edwards JE, Cheng YF, *et al*. Rumen microbial diversity in Svalbard reindeer,
674 with particular emphasis on methanogenic archaea. *FEMS Microbiol Ecol* 2009; **70**: 553-62.

675 Tajima K, Nagamine T, Matsui H, *et al*. Phylogenetic analysis of archaeal 16S rRNA libraries
676 from the rumen suggests the existence of a novel group of archaea not associated with known
677 methanogens. *FEMS Microbiol Lett* 2001; **200**: 67-72.

678 Talavera G, Castresana J. Improvement of phylogenies after removing divergent and
679 ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007; **56**: 564-77.

680 Tveit A, Urich T, Svenning MM. Metatranscriptomic analysis of Arctic peat soil microbiota.
681 *Appl Environ Microbiol* 2014, DOI: 10.1128/AEM.01030-14.

682 Tveit A, Schwacke R, Svenning MM, *et al*. Organic carbon transformations in high-Arctic
683 peat soils: key functions and microorganisms. *Isme Journal* 2012; **7**: 299-311.

684 UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res* 2014, DOI:
685 10.1093/nar/gku989.

686 Upton M, Hill B, Edwards C, *et al*. Combined molecular ecological and confocal laser
687 scanning microscopic analysis of peat bog methanogen populations. *FEMS Microbiol Lett*
688 2000; **193**: 275-81.

689 Urich T, Lanzén A, Qi J, *et al*. Simultaneous assessment of soil microbial community
690 structure and function through analysis of the meta-transcriptome. *PLoS One* 2008, DOI:
691 10.1371/journal.pone.0002527.

692 Vanderhaeghen S, Lacroix C, Schwab C. Methanogen communities in stool of humans of
693 different age and health status and co-occurrence with bacteria. *FEMS Microbiol Lett* 2015,
694 DOI: 10.1093/femsle/fnv092.

695 Watkins AJ, Roussel EG, Parkes RJ, *et al*. Glycine betaine as a direct substrate for
696 methanogens (Methanococcoides spp.). *Appl Environ Microbiol* 2014; **80**: 289-93.

697 Wright A-DG, Williams AJ, Winder B, *et al*. Molecular diversity of rumen methanogens from
698 sheep in Western Australia. *Appl Environ Microbiol* 2004; **70**: 1263-70.

699 Yarza P, Yilmaz P, Pruesse E, *et al*. Uniting the classification of cultured and uncultured
700 bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014; **12**: 635–645.

Yavitt JB, Yashiro E, Cadillo-Quiroz H, *et al*. Methanogen diversity and community composition in peatlands of the central to northern Appalachian Mountain region, North America. *Biogeochemistry* 2012; **109**: 117-31.

Yrjälä K, Tuomivirta T, Juottonen H, *et al*. CH4 production and oxidation processes in a boreal fen ecosystem after long-term water table drawdown. *Global Change Biol* 2011; **17**: 1311-20.

Zedler JB, Kercher S. Wetland resources: Status, trends, ecosystem services, and restorability. *Annu Rev Environ Resour* 2005; **30**: 39-74

712 **Table 1** Origin and characteristics of arctic and temperate wetland samples as reported by
713 Tveit et al. (2012) and Steger et al. (2011).

	wetland	type	pH	samples/ID
arctic	Knudsenheia, Norway	fen	5.5 ± 0.2	Knudsenheia
	Solvatn, Norway	fen	5.2 ± 0.1	Solvatn
temperate	Roßbrand, Austria	fen	4.1 ± 0.1	AUT 1, 2
	Roßbrand, Austria	fen	4.9 ± 0.2	AUT 3
	Roßbrand, Austria	ferrous biofilm	n.a.	AUT 7
	Große Heide, Austria	bog	5.3 ± 0.1	AUT 4, 5
	Schremser Hochmoor, Austria	bog	4.7 ± 0.4	AUT 6
	Schallhof, Austria	fen	7.3 ± 0.1	AUT 8
	Berndorf, Austria	fen	7.6 ± 0.0	AUT 9, 10
	Krähmoos, Italy	bog	4.0 ± 0.1	IT 1
	Rasner Möser, Italy	fen	4.3 ± 0.1	IT 2, 3
	Schlöppnerbrunnen, Germany	fen	acidic	GER 1

714
715
716
717
718

719 **Table 2** Primers used in this study.

primer	sequence	fragment length	target gene and taxon (reference)
AS1	CAG CAG TCG CGA AAA CTT C	485 nt	16S rRNA of MMC
AS2	AAC AAC TTC TCT CCG GCA		(Mihajlovski et al 2010
Tp-mcrA-F	GAY RAC ATC CTB GAR GAY TA	360 nt	*mcrA* of MMC
Tp-mcrA-R	RTC GWA WCC RTA GAA TCC GAG		(Petersen et al 2014)
mlas-mod For	GGY GGT GTM GGD TTC ACM CAR TA	470 nt	*mcrA* of methanogens
mcrA-rev	CGT TCA TBG CGT AGT TVG GRT AGT		(Steinberg and Regan 2009

720
721
722
723
724

725 **Table 3** Relative abundances (% of total methanogenic community) of methanogens in arctic
726 peat soil from Knudsenheia (Ka, Kb, Kc) and Solvatn (Sa, Sb).

order	Ka	Kb	Kc	Sa	Sb
Methanomassiliicoccales	4.7	0.5	0.7	0.0	1.5
Methanobacteriales	66.3	3.1	3.4	10.0	2.6
Methanosarcinales	20.9	84.7	87.0	70.0	87.9
Methanomicrobiales	8.1	11.7	8.9	20.0	7.9
Methanocellales	2.3	2.7	11.6	0.0	0.0

727
728
729

730

731 **Table 4** Characteristics of the two partial rumen MMC genomes, RumEn M1 (182 contigs)
732 and RumEn M2 (18 contigs) and genomes of other members of MMC[a]. The dashed line
733 separates the environmental clade and the GIT clade MMC.

genome	accession no. (GenBank)	G+C content (mol%)	size [Mbp]	tRNA genes
RumEn M1	LJKK00000000	62.5	~2.21	43
M. luminyensis[b]	CAJE01000001-26	60.5	>2.62	43
Ca. M. intestinalis[b]	CP005934	41.3	1.93	46
RumEn M2	LJKL00000000	54.8	~1.28	44
Ca. M. termitum[c]	CP010070	49.2	1.49	46
Ca. M. alvus	CP004049	55.6	1.67	45
BRNA1[d]	CP002916	58.3	1.46	44

734 [a] Data are derived from the latest MMC comparative genome analysis (Lang et al 2014) and GenBank.
735 [b] Closest relatives of M1; based on a nearly full length 16S rRNA gene sequence (1467 nt), with 96 % 16S
736 rRNA sequence identity to both *Methanomassiliicoccus* species.
737 [c] Closest relative of M2; based on a 16S rRNA gene sequence (1272 nt), with 95 % 16S rRNA sequence
738 identity to *Ca*. M. termitum.
739 [d] BRNA1 was also obtained from bovine rumen but is only distantly related to M1 and M2, with 87 % and
740 92 % 16S rRNA sequence identity, respectively. Number of tRNAs were obtained as for M1 and M2.
741
742
743

FIGURES

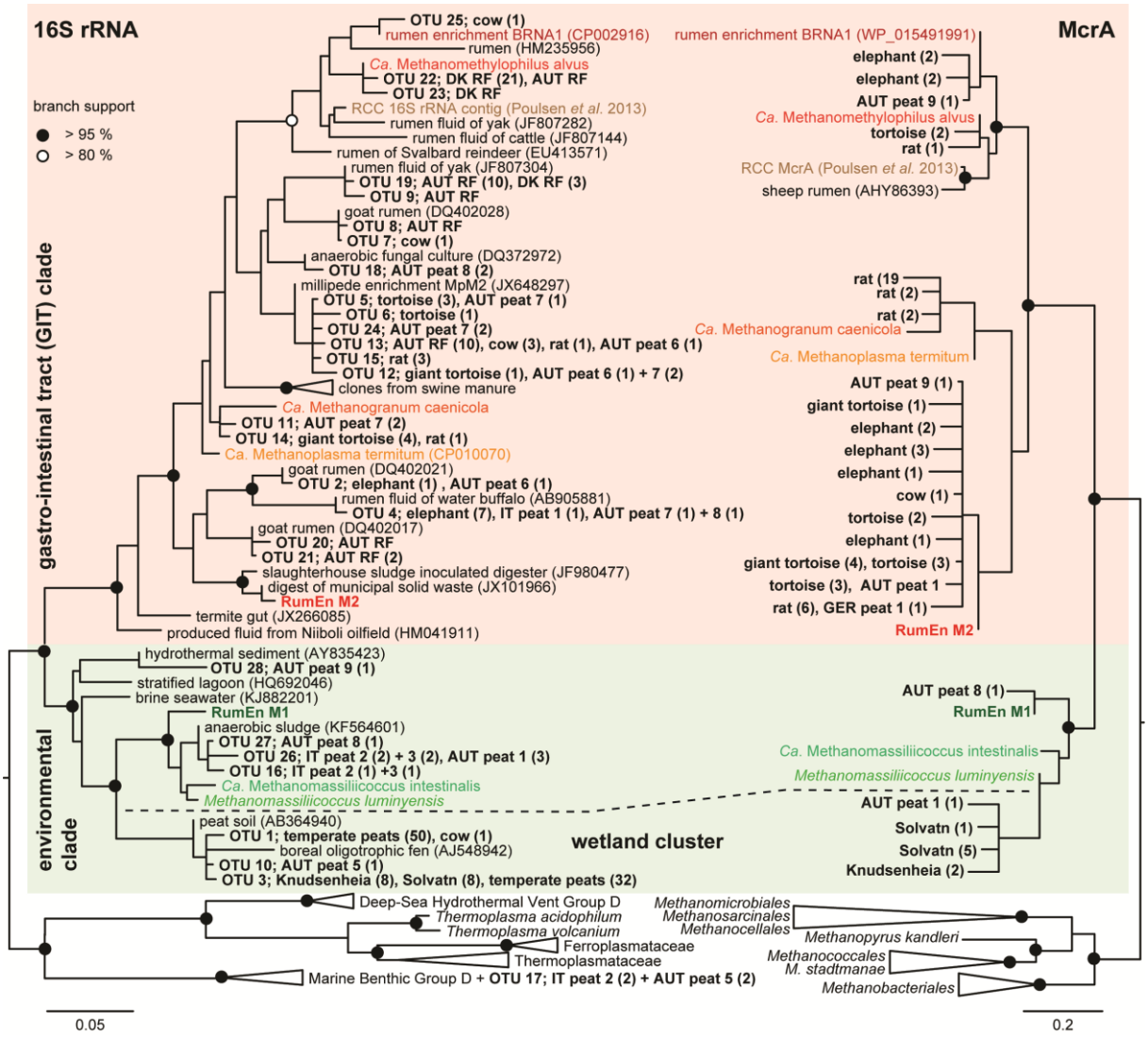
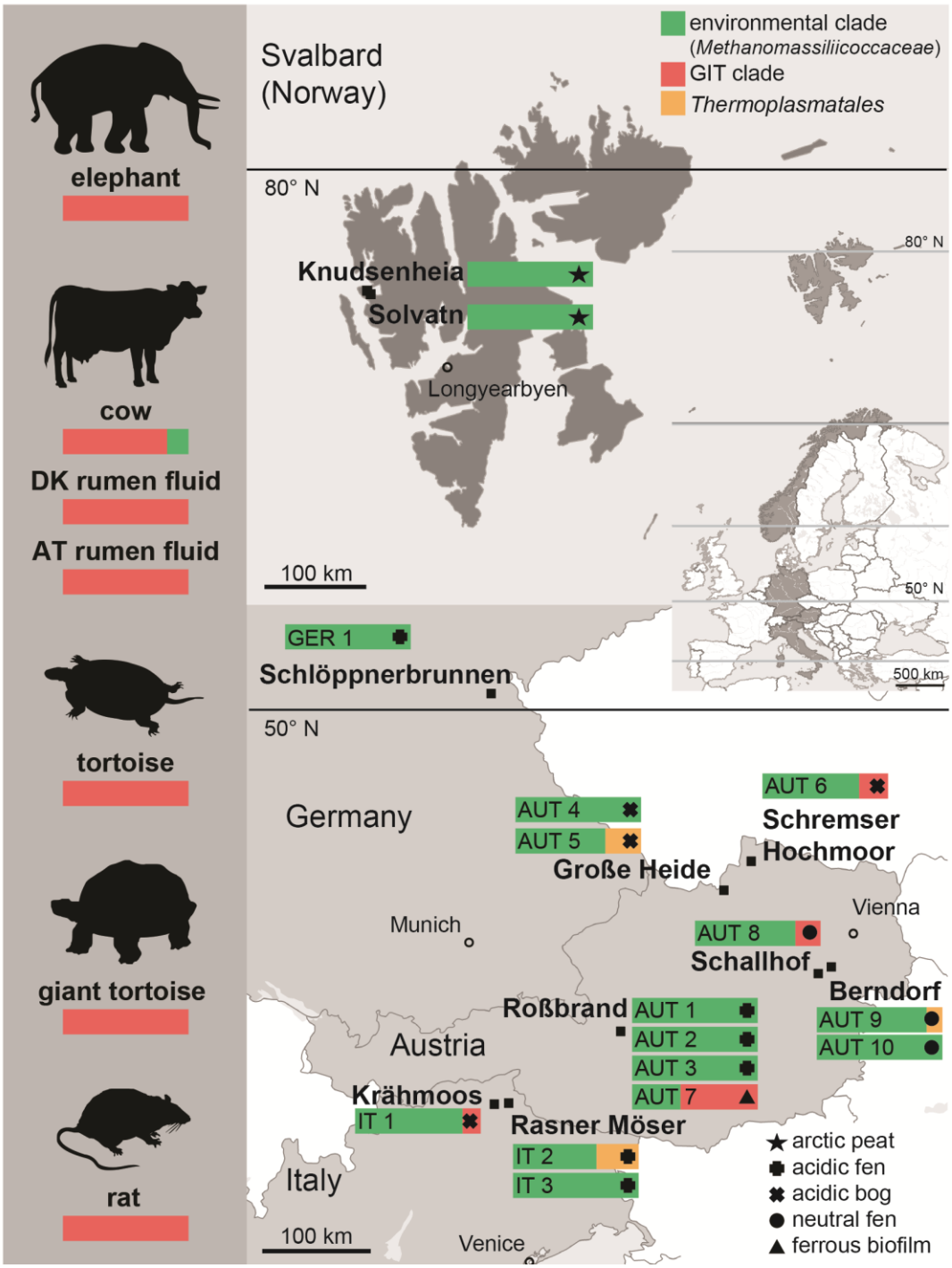


Figure 1. Phylogenetic trees showing the relationship among 16S rRNA gene and McrA protein sequences obtained from animal fecal samples, bovine rumen fluid and wetland soils. Operational taxonomic units (OTUs) with <98 % sequence identity and sequences derived from rumen *Methanomassiliicoccales* genomes (RumEn M1 and RumEn M2) are indicated in bold fonts. The positions of 16S rRNA gene and McrA protein sequences obtained from the same organism or enrichment cultures are shown in the same colour. Maximum likelihood trees of near full-length references sequences were calculated with ARB. The partial *Methanomassiliicoccales* 16S rRNA gene and McrA protein sequences were placed into these

trees using the maximum parsimony tool implemented in ARB. Numbers in brackets refer to clone numbers obtained from the same sample. AUT, Austrian; DK, Danish; RF, rumen fluid; GER, German; IT, Italian. Substitutions per site are indicated with scale bars.



760

761

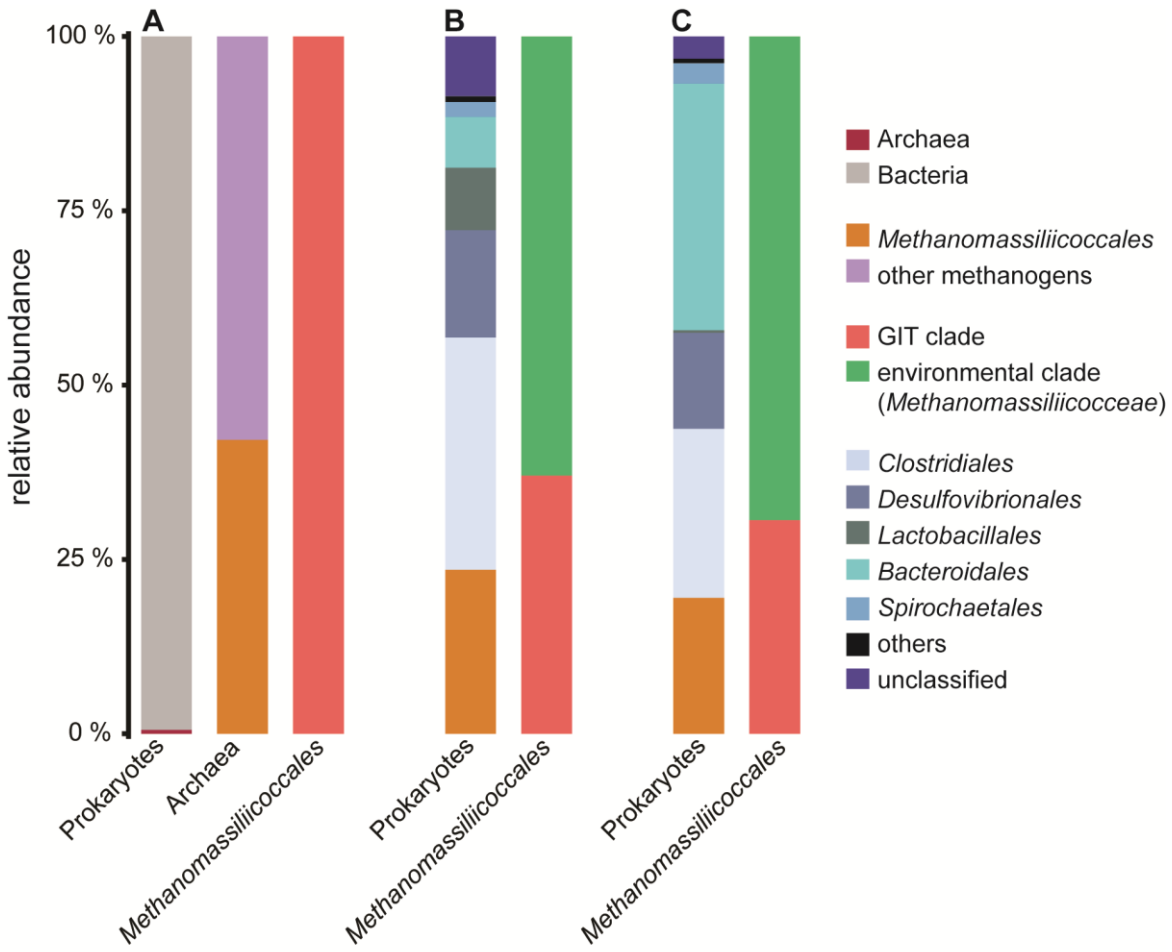
762 **Figure 2**. Distribution of environmental and GIT clade *Methanomassiliicoccales* 16S rRNA

763 gene clones in animal and wetland samples. Colour code indicates the percentages of 16S

764 rRNA gene clones belonging to one of these two *Methanomassiliicoccales* family-level clades

765 or to related *Thermoplasmata*.

766



767

768 **Figure 3**. Enrichment of *Methanomassiliicoccales* clades from bovine rumen fluid. **Panel (A)**
769 community composition of rumen fluid used for initial inoculation of
770 *Methanomassiliicoccales* enrichment cultures. Left column: relative abundance of bacterial
771 and archaeal 16S rRNA gene copy numbers obtained from quantitative PCR assays (see
772 materials and methods for details). Central column: relative abundance of
773 *Methanomassiliicoccales* and other methanogenic archaea obtained from quantitative PCR
774 assays. Right column: community composition of *Methanomassiliicoccales* derived from
775 16S rRNA gene clone library (AUT RF; 26 clones), showing the dominance of the GIT clade.
776 **Panel (B)** community composition in metagenome RumEn_MG1 based on the classification
777 of 234 16S rRNA gene fragments. **Panel (C)** community composition in metagenome
778 RumEn_MG2 based on the classification of 1360 16S rRNA gene fragments. The left

779 columns of (B) and (C) show the 16S rRNA gene fragment abundances of archaeal and
780 bacterial orders. The right columns of (B) and (C) depict the relative abundance of 16S rRNA
781 gene fragments of the *Methanomassiliicoccales* GIT and environmental clades.

782