

ORIGINAL ARTICLE

Phylogenomic analysis of *Candidatus* ‘Izimaplasma’ species: free-living representatives from a *Tenericutes* clade found in methane seeps

Connor T Skennerton¹, Mohamed F Haroon^{2,6}, Ariane Briegel^{3,7}, Jian Shi³, Grant J Jensen^{3,4}, Gene W Tyson^{2,5} and Victoria J Orphan¹

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA;

²Australian Center for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, Queensland, Australia; ³Division of Biology and Biological Engineering, California

Institute of Technology, Pasadena, CA, USA; ⁴Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA, USA and ⁵The Advanced Water Management Center, University of Queensland, St Lucia, Queensland, Australia

Tenericutes are a unique class of bacteria that lack a cell wall and are typically parasites or commensals of eukaryotic hosts. Environmental 16S rDNA surveys have identified a number of tenericute clades in diverse environments, introducing the possibility that these *Tenericutes* may represent non-host-associated, free-living microorganisms. Metagenomic sequencing of deep-sea methane seep sediments resulted in the assembly of two genomes from a *Tenericutes*-affiliated clade currently known as ‘NB1-n’ (SILVA taxonomy) or ‘RF3’ (Greengenes taxonomy). Metabolic reconstruction revealed that, like cultured members of the *Mollicutes*, these ‘NB1-n’ representatives lack a tricarboxylic acid cycle and instead use anaerobic fermentation of simple sugars for substrate level phosphorylation. Notably, the genomes also contained a number of unique metabolic features including hydrogenases and a simplified electron transport chain containing an RNF complex, cytochrome *bd* oxidase and complex I. On the basis of the metabolic potential predicted from the annotated genomes, we devised an anaerobic enrichment media that stimulated the growth of these *Tenericutes* at 10 °C, resulting in a mixed culture where these organisms represented ~60% of the total cells by targeted fluorescence *in situ* hybridization (FISH). Visual identification by FISH confirmed these organisms were not directly associated with Eukaryotes and electron cryomicroscopy of cells in the enrichment culture confirmed an ultrastructure consistent with the defining phenotypic property of *Tenericutes*, with a single membrane and no cell wall. On the basis of their unique gene content, phylogenetic placement and ultrastructure, we propose these organisms represent a novel class within the *Tenericutes*, and suggest the names *Candidatus* ‘Izimaplasma sp. HR1’ and *Candidatus* ‘Izimaplasma sp. HR2’ for the two genome representatives.

The ISME Journal (2016) 10, 2679–2692; doi:10.1038/ismej.2016.55; published online 8 April 2016

Introduction

Mollicutes are bacteria that lack a cell wall and typically form close associations with animal and plant hosts as commensals or parasites (Razin *et al.*, 1998). Taxonomically, the *Mollicutes* are

currently classified as the only class within the phylum *Tenericutes*, however the phylogeny of ribosomal rRNA genes and proteins places the *Mollicutes* as members of the *Firmicutes* and close relatives of the firmicute class *Erysipelotrichia* (Ogawa *et al.*, 2011; Davis *et al.*, 2013; Yutin and Galperin, 2013). Most cultured *Mollicutes* can be divided into four main clades based on 16S rDNA sequences: the hominis group, which contains some members of the genus *Mycoplasma*; the pneumoniae group, which includes some *Mycoplasma* and the *Ureaplasma*; the spiroplasma group, which includes the genera *Spiroplasma*, *Entomoplasma* and *Mesoplasma*; and the anaeroplasmata group, which includes the genera *Anaeroplasmata*, *Acholeplasmata* and *Candidatus* *Phytoplasmata* (Davis *et al.*, 2013). There are also a number of peripheral clades that

Correspondence: CT Skennerton or VJ Orphan, Division of Geological and Planetary Sciences, California Institute of Technology, 1200 East California Blvd, MC 100-23, Pasadena, CA 91125, USA.

E-mail: c.skennerton@gmail.com or vorphan@gps.caltech.edu

⁶Present address: Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.

⁷Present address: Sylvius Laboratorium, Leiden University, Sylviusweg 72, 23333 BE Leiden, The Netherlands.

Received 13 January 2016; revised 29 February 2016; accepted 4 March 2016; published online 8 April 2016

contain cultured organisms such as *Haloplasma contractile* (*H. contractile*), *Turicibacter sanguinis* (*T. sanguinis*) and clades known only through environmental community profiling of 16S rDNA, such as 'RF9' and 'NB1-n' (naming based on the SILVA 119 database).

The most widely studied *Mollicutes*, the *Mycoplasma*, are important as medical and agricultural parasites that have adapted to host-associated environments where nutrients are readily available. Parasitic *Mollicutes* have undergone significant genome reduction that has resulted in the loss of most biosynthetic pathways for nucleotides, amino acids and fatty acids, as well as functions likely present in their firmicute ancestors such as sporulation and cell wall production (Razin *et al.*, 1998). The very smallest *Mollicutes* species contain ~500 kbp genomes that have been of great interest for synthetic biology (Gibson *et al.*, 2010) and as model organisms for determining the minimal genetic content required for life (Glass *et al.*, 2006). However, analysis of other mollicute clades has revealed a broader range of genome sizes; the *Acholeplasma* have an average genome length of ~1.2 Mbp, while closely related organisms, *T. sanguinis* and *H. contractile* have 2.9 and 3.2 Mbp genomes, respectively (Antunes *et al.*, 2011; Cuvř *et al.*, 2011; Lazarev *et al.*, 2011; Kube *et al.*, 2013).

The metabolisms of all previously studied *Mollicutes* exclusively rely on substrate-level phosphorylation for ATP generation. Mollicute genomes sequenced to date lack a complete tricarboxylic acid cycle and cytochromes or quinones for use in oxidative phosphorylation (Razin, 2006). Respiration is achieved through the fermentation of carbohydrates, small organic acids or through the arginine dihydrolase pathway (Razin, 2006). In the *Ureaplasma*, an alternative pathway exists that couples urea hydrolysis with the generation of a membrane potential to generate ATP (Smith *et al.*, 1993).

Recently, *H. contractile*, a wall-less bacteria phylogenetically affiliated with the *Mollicutes* was cultured from the brine-sediment interface of the Red Sea (Antunes *et al.*, 2008). Genome sequencing and phylogenetic analysis of *H. contractile* revealed some features common to *Mollicutes* such as a lack of a cell wall, however its large genome size and lack of known host organism were inconsistent with classification as a member of the *Mollicutes* (Antunes *et al.*, 2008, 2011). *Haloplasma* appears to be a basal lineage of the *Mollicutes* and is most closely related to *T. sanguinis* (a gram positive firmicute) and several other environmental clades. One of these clades, 'NB1-n' contains ~100 sequences from 16S rDNA environmental surveys from diverse habitats such as marine and freshwater sediments, hypersaline microbial mats, rumen and anaerobic digester sludge. Here we present genome sequences of two representatives of clade 'NB1-n', reconstructed from metagenomic sequencing of marine methane seep sediment communities. These

two genomes are more metabolically diverse than known *Mollicutes* and do not appear to be host-associated. However, they branch with members of the *Acholeplasmatales* when using conserved ribosomal and non-ribosomal genes and lack a cell wall, the distinctive phenotypic feature of *Tenericutes*.

Materials and methods

Sample collection and incubation

Methane seep sediment samples, labeled 3730, 5133-5 and 5579, were collected from Hydrate Ridge, off the coast of Oregon on two separate cruises, AT 15-68 in August 2010 for 3730 and AT 18-10 in September 2011 for 5579 and 5133-5. Seep sediment was collected using the HOV *Alvin* (2010) or ROV *Jason II* (2011) with a push core and immediately processed shipboard as described previously (Orphan *et al.*, 2001). Sample 3730 was collected on dive AD4635 within a *Calypptogena* clam bed from Hydrate Ridge south (44° 43.09N; 125° 9.14W; depth 776 m), and 5133-5 and 5579 were collected from a white microbial mat at Hydrate Ridge north, station 7 (44° 40.03N; 125° 6.00W; depth 600 m). Discrete sediment horizons from each of the cores were transferred to glass pyrex bottles (1 L), sparged with methane and stoppered with a butyl rubber stopper. Sediments were maintained at 4 °C in the dark until processing in the lab for large-scale microcosm experiments. For sample 3730, sediment from 0 to 6 cm depth was combined and used in a 30-month microcosm incubation experiment with a methane headspace before DNA extraction and metagenome sequencing. For sample 5133-5, the 0-9-cm sediment horizon was collected and incubated for 15 months before DNA extraction. For sample 5579, 0-12 cm sediment was collected and incubated for 15 months. Sediment microcosm incubations from all samples were mixed 1:3 with anoxic 0.2 µm filtered bottom water to make a slurry and maintained anaerobically, with a 2-atm overpressure of methane at 4 °C. DNA was extracted from the sediment slurry using the MO BIO Powersoil DNA extraction kit (cat # 12888; MO BIO Laboratories Inc., Carlsbad, CA, USA). Extracted DNA was further purified using a CsCl gradient before metagenome library preparation (Saunders and Burke, 1990).

Sequencing and assembly of *Izimaplasma* genomes

Extracted DNA from samples 3730, 5133-5 and 5579 were sequenced using Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA, USA) at the University of Queensland, Australia. Samples 3730 and 5133-5 were further sequenced a second time at Argonne National Laboratory to increase sequencing depth. Raw sequencing reads were quality trimmed (using a threshold of 0.05) and assembled using CLCBio genomics workbench 6.0 (www.clcbio.com).

Sequencing reads from each metagenome were assembled individually and two assemblies were performed by combining sequencing reads from the three data sets sequenced at the University of Queensland and all five data sets sequenced at both Argonne and the University of Queensland. For each assembly, the raw sequencing reads from all data sets were aligned to contigs > 500 bp with BWA 0.7.5a (Li, 2013) using the 'mem' algorithm and converted into bam files using samtools 0.1.19 (Li *et al.*, 2009). These bam files were used as input for GroopM 0.2 (Imelfort *et al.*, 2014), which generated genome bins. The 'refine' command in GroopM was used to improve the genome bins manually by removing contigs with unusual coverage or tetramer frequencies. The completeness and contamination of genome bins was determined using CheckM 0.6.0 (Parks *et al.*, 2014).

The *Izimaplasma* sp. HR1 genome was further improved with additional genome finishing and reassembly. The draft *Izimaplasma* sp. HR1 genome contained four pieces, including a single double-copy transposon. Analysis of paired-end read mapping from *Izimaplasma* sp. HR1 suggested an order of the contigs. The orientation a single contig bordered by the transposon was confirmed by designing PCR primers on either side of the transposon for both of the possible orientations.

Genome annotation and pathway analysis

The two *Izimaplasma* genomes were annotated using Prokka 1.10 (Seemann, 2014) and the KAAS web annotation server (Moriya *et al.*, 2007). Prokka used as components: Prodigal 2.5 (Hyatt *et al.*, 2010), Aragorn 1.2.34 (Laslett and Canback, 2004), SignalP 4.0 (Petersen *et al.*, 2011), hmmer 3.0 (<http://hmmer.org>) and blast+ 2.2.26 (Camacho *et al.*, 2009). As a supplement to this annotation, the KAAS web server was used to determine KEGG orthologies and map genes onto KEGG pathways. The genetic code of both genomes was determined using FACIL (Dutilh *et al.*, 2011). Orthologs were detected between *Izimaplasma* and other mollicute genomes sourced from Integrated Microbial Genomes (IMG) and NCBI genomes databases using proteinortho 5.0 (Lechner *et al.*, 2011). Metabolic pathways from previously sequenced mollicute genomes were compared using the PATRIC web server (Wattam *et al.*, 2014).

16S rRNA gene tree construction

The two *Izimaplasma* 16S rRNA gene sequences were imported into ARB 6.0.2 (Ludwig *et al.*, 2004) and aligned in ARB using a PT server generated from the SILVA 119 release database. A maximum likelihood tree of all sequences in SILVA 119 that were part of the 'NB1-n' clade was generated with RAXML 8.1.7 (Stamatakis, 2014) with 100 replicates of the rapid bootstrap analysis and GTR+GAMMA model of substitution rates (commandline: raxmlHPC-PTHREADS-AVX -f a -k -x 12345 -p 12345 -N 100 -m

GTRGAMMAI). In addition, a second tree was constructed with RAXML independent of the SILVA alignment. Sequences from the *Tenericutes*, *Erysipelotrichia*, and representative *Firmicutes* and *Actinobacteria* were aligned with Infernal v1.1.1 (Nawrocki and Eddy, 2013) using the bacterial covariance model provided by the Ribosomal Database Project (RDP; <https://rdp.cme.msu.edu/misc/resources.jsp#aligns>). A phylogenetic tree was constructed using RAXML with the following parameters: raxmlHPC-PTHREADS-SSE3 -f a -k -x 67842 -p 19881103 -N 100 -T 16 -m GTRGAMMAI.

Genome tree construction

The genome tree was constructed from a concatenated alignment of 38 protein-coding genes that are universally distributed and in a single copy in both archaea and bacteria (Wu *et al.*, 2013). The basis for the tree included both draft and finished genomes from the NCBI and IMG databases that were part of the *Tenericutes* and phylogenetically related members of the *Firmicutes*. The marker genes were identified in each genome using hmmer 3.0 with the 'hmmalign' command (<http://hmmer.org>). A maximum likelihood tree was created using FastTree 2.1.7 (Price *et al.*, 2010) with the default settings.

RpoB and EF-tu tree construction

RpoB or EF-tu sequences were identified in genomes downloaded from IMG genomes using the hidden markov models (HMMs) TIGR02013, TIGR03670 and TIGR00485 from TIGRfams (<http://www.jcvi.org/cgi-bin/tigrfams/index.cgi>). EF-tu sequences were aligned using hmmalign (<http://hmmer.org>), with the TIGRfam model (TIGR00485). RpoB sequences were aligned using muscle 3.8.31 (Edgar, 2004). Phylogenetic trees were constructed using FastTree 2.1.7 (Price *et al.*, 2010).

Cytochrome bd oxidase tree construction

Sequences belonging to PF01654 were downloaded from Pfam (<http://pfam.xfam.org/family/PF01654/alignment/long/gzipped>) and were clustered using cd-hit 4.6.1 (Li and Godzik, 2006) at 90% identity to generate a representative set of sequences from previously published genomes. The same HMM was used to search against open reading frames translated from the contigs of the metagenomes using hmmsearch 3.1. Significant (<1e-100) complete protein sequences were added to the representative set of published sequences, aligned using muscle 3.8.31 and a phylogenetic tree was constructed using FastTree 2.1.7.

Fluorescence in situ hybridization

A FISH oligonucleotide probe, Izzy659 (5'-ATTCCA CTAACCTCTGCC-3'), was designed to target some

members of the 'NB1-n' clade that contained the two genome representatives. Using the SILVA 119 database and the probe design tool in ARB 6.0.2 (Ludwig *et al.*, 2004) we designed a probe that targeted 15 16S rRNA sequences, including the two genome representatives (Supplementary Figure S2). FISH hybridization was performed on sediment #5579—an anaerobic incubation maintained with a 2-atm methane headspace at 4 °C. #5579 samples were fixed with 3% formaldehyde for 1.5 h at 4 °C, washed with 1 × phosphate-buffered saline, and stored in 1:1 1 × phosphate-buffered saline:ethanol. The optimal formamide concentration for the Izzy659 probe was determined empirically through FISH hybridizations at 20%, 30% and 40%. Formamide concentrations of 30% gave the brightest fluorescence signal. FISH hybridizations were performed on glass slides at 46 °C for 2 h using the protocol described in Pernthaler *et al.* (2001) with 5 ng ml⁻¹ of Eub338-FITC and Izzy659-Cy3 and a formamide stringency of 30%. Slides were dried in the dark and then counterstained with 5 mg ml⁻¹ of 4',6'-diamidino-2-phenylindole (DAPI) in CITI-FLUOR (<http://citifluor.com/>) mounting medium. Positively hybridized cells were imaged using an Olympus BX51 epifluorescence microscope at ×1000 magnification using an Olympus UPlanFL N 100X objective fitted with a gray scale camera. Individual color channels were merged using Fiji 2.0.0-rc-19 (<http://fiji.sc>).

Electron cryomicrography

Cells were concentrated by centrifuging 2 × 1 ml aliquots of culture at 10 000 *g* for 5 min and were resuspended in 0.15 ml supernatant. Twenty microliters of a concentrated cell suspension was mixed with BSA-treated colloidal gold and applied to R2/2 copper quantifoil EM grids (Quantifoil Micro Tools, Großlobichau, Germany). Excess liquid was blotted with filter paper and plunged into a liquid ethane/propane mixture (Tivol *et al.*, 2008) either manually or using a Vitrobot (FEI Company, Hillsboro, OR, USA). Images were recorded using an FEI G2 (FEI Company) 300 kV field emission gun electron microscope equipped with a Gatan imaging filter and either a 4 × 4 k UltraCam or a K2 Summit counting direct electron detector camera (Gatan, Pleasanton, CA, USA). Data were collected using digital micrograph and UCSFtom software (Zheng *et al.*, 2007). Cumulative electron doses of ~160 e⁻ Å⁻² or less were used for each tilt series. The images were processed using the IMOD software package (Kremer *et al.*, 1996).

Izimaplasma enrichment media

Izimaplasma were enriched from Hydrate Ridge methane seep sediment sample #5579 using a media composed of artificial seawater amended with glucose (5 g l⁻¹), yeast extract (0.2 g l⁻¹) and the antibiotics

ampicillin and vancomycin (see Supplementary Methods for details of the media recipe). The media was reduced by adding a small amount (~0.2 ml) of Ti (III)-nitritotriacetic acid in a bottle under 80:20 N₂:CO₂ headspace. Ampicillin and Vancomycin were added to the anaerobic, reduced media to a final concentration of 200 and 1 µg ml⁻¹. Enrichment bottles (30 ml media in 70 ml serum bottles) were incubated at 10 °C under 200 kPa of N₂ headspace.

Community profiling

DNA was extracted from the enrichment cultures using the MO BIO PowerSoil DNA isolation kit (MO BIO Laboratories Inc.; cat. no. 12888-50) according to the manufacturer's instructions. Sequencing was performed using the Illumina MiSeq platform according to the earth microbiome project specification; see Mason *et al.* (2015) for a full description of the sequencing and bioinformatics processing of the data.

Data availability

The genome sequences for *Izimaplasma* sp. HR1 and HR2 are available on NCBI as CP009415 and JRF01000000, respectively. The metagenomic data is available on NCBI under the bioproject accession PRJNA290197.

Results

Phylogeny of methane seep *Tenericutes*

Metagenomic assembly and binning resulted in two distinct, highly complete population genome bins, designated HR1 and HR2, that were affiliated with the *Tenericutes* based on phylogenetic analysis of the 16S rRNA gene and a concatenated protein alignment of 38 universally distributed single-copy genes (hereafter referred to as a genome tree; Figure 1; Supplementary Figure S1; Supplementary Table S1). Initial placement of the 16S rRNA gene using the SILVA 119 database and ARB parsimonious insertion tool revealed that the genomes were members of a clade known as 'NB1-n' (clade referred to as RF3 in the greengenes database) that were comprised solely of environmental 16S rRNA sequences (Supplementary Figure S2). 'NB1-n' is defined as an order-level group in the SILVA 119 database that contains sequences derived from mesophilic anaerobic digesters, landfill, animal rumen and saline sediments (Supplementary Figure S2). The genome tree places HR1 and HR2 as a sister group of the *Acholeplasmatales* with *H. contractile* and *T. sanguinis* branching more deeply (Figure 1). Realignment of the 16S rRNA gene using the protocol of the Ribosomal Database Project (RDP) and phylogenetic construction of clade 'NB1-n' with cultured *Tenericutes*, some *Firmicutes* and *Actinobacteria* supported the placement of clade 'NB1-n' as a sister group of the *Acholeplasmatales*; however, in



Figure 1 Maximum likelihood phylogenetic tree of *Tenericutes* and the firmicute class *Eysipelotrichia* constructed from the concatenated alignment of 38 single-copy genes; *Bacillus subtilis* was used as the outgroup. Collapsed wedges represent monophyletic groups of genomes. Nodes with greater than 95% bootstrap support are annotated with a black circle. Scale bar indicates substitutions per site.

this tree *H. contractile* and *T. sanguinis* branch within clade ‘NB1-n’ as defined by SILVA 119 (Figure 2).

The phylum *Tenericutes* was defined in part by using non-ribosomal protein markers that showed that members of the *Mollicutes* were phylogenetically distinct from the *Firmicutes* (Ludwig and Schleifer, 2005). Further phylogenetic analysis of HR1 and HR2 using the beta subunit of the RNA polymerase (*rpoB*) and elongation factor Tu (EF-Tu) placed HR1 and HR2 basal to members of the *Acholeplasma* (Supplementary Figure S3; Supplementary Figure S4). The broader phylogeny of RpoB revealed that the *Mollicutes* are polyphyletic; members of the genus *Mycoplasma* branch deeply, and separate from the *Firmicutes*, in agreement with previous phylogenetic analysis (Ludwig and Schleifer, 2005); while members from the *Acholeplasma*, *Entomoplasma* and *Candidatus Phytoplasma* branch within the *Firmicutes*. The EF-Tu phylogeny places the *Mycoplasma* and *Acholeplasma* in two distinct clades, separated by members of the *Eysipelotrichia* and some bacilli. *H. contractile* and *T. sanguinis* branch together in both the RpoB and EF-Tu trees in support of their close

association in the genome tree and 16S rRNA gene tree (Figure 1; Figure 2; Supplementary Figure S3; Supplementary Figure S4).

Based on the unique phylogenetic position of these two genomes, we propose the names *Candidatus* ‘*Izimaplasma* sp. HR1’ and *Candidatus* ‘*Izimaplasma* sp. HR2’ (*Izima*-, Greek for sediment; -*plasma*, formed or molded, refers to the lack of cell wall) for the two genome sequences presented here.

Genome-based metabolic reconstruction of *Izimaplasma*

The two *Candidatus* ‘*Izimaplasma*’ genomes (referred to as *Izimaplasma* sp. HR1 and *Izimaplasma* sp. HR2 hereafter) are comparatively larger than most other *Mollicutes*, at 1.8 and 2.1 Mbp (Supplementary Table S1; Supplementary Figure S6), however, they are of similar size to some members from the *Acholeplasmatales*, which are phylogenetically the closest recognized *Tenericute* clade. Like the *Acholeplasmatales*, the *Izimaplasma* genomes use the bacterial genetic code with UGA as a stop codon rather than encoding for tryptophan (Supplementary Figure S7). Both genomes contain hallmarks of genome streamlining, including a

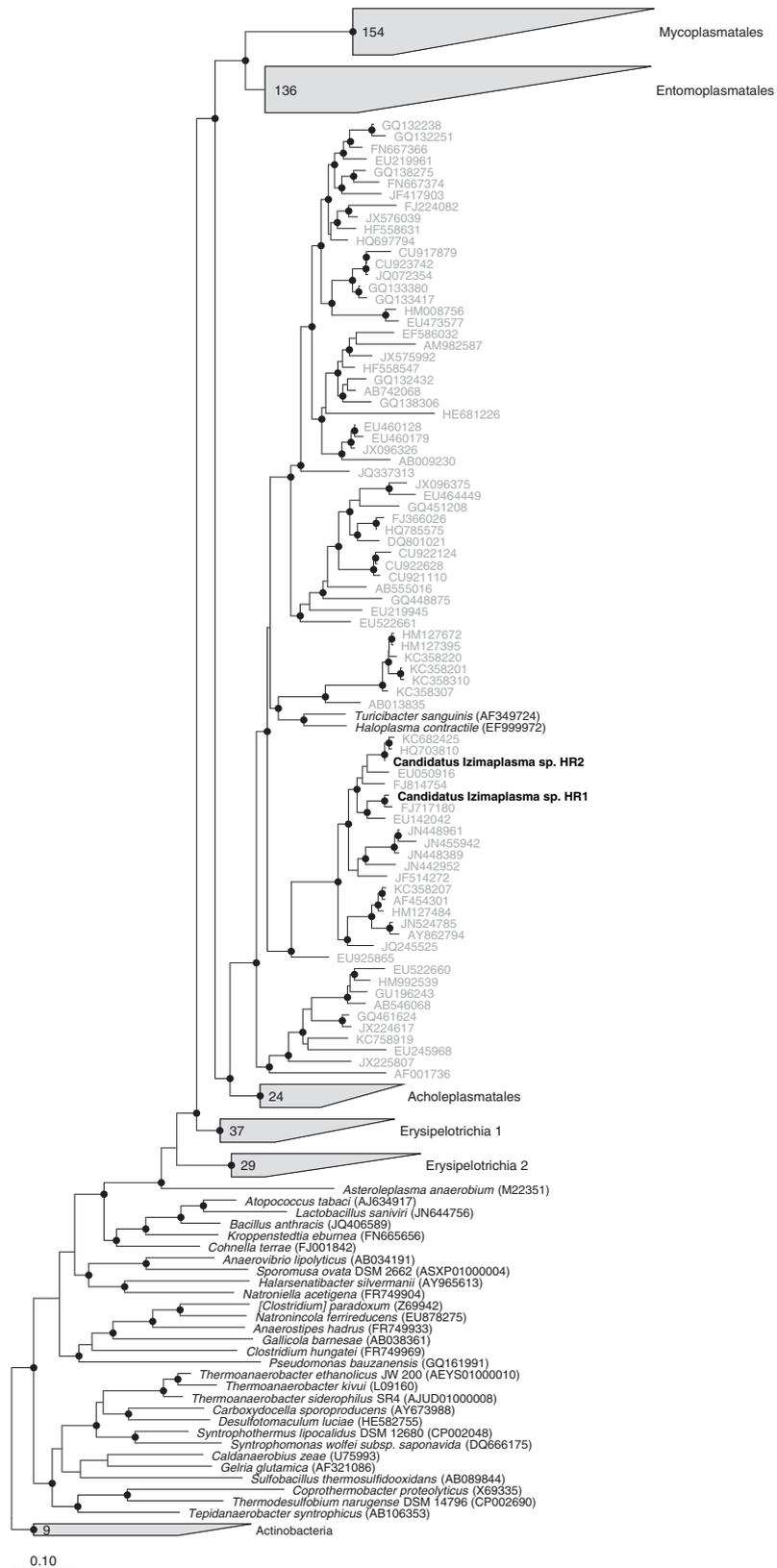


Figure 2 Maximum likelihood phylogenetic tree of 16S rRNA gene sequences from cultured *Tenericutes*, *Erysipelotrichia*, some other *Firmicutes* representatives and some *Actinobacteria* used as the outgroup. Gray wedges represent monophyletic groups of a particular taxonomy. In this tree the class *Erysipelotrichia* are split into two groups, labeled 1 and 2. Sequences belonging to clade 'NB1-n' are shown in gray. All sequences are labeled with their NCBI accession numbers. Nodes with greater than 70% bootstrap support are labeled with a black circle. Scale bar represents substitutions per site.

high protein coding percentage (Supplementary Figure S6) and a small number of repetitive elements, with *Izimaplasma* sp. HR1 containing a single multi-copy transposon, which facilitated the highly contiguous metagenomic assembly and genome finishing.

Reconstruction of metabolic pathways in the genomes resulted in a predictive model of the central carbon metabolism reliant on fermentation of simple sugars, similar to other *Tenericutes* (Figure 3). The two *Izimaplasma* genomes contained a complete set of genes to perform the Embden–Meyerhof–Parnas

(EMP) pathway for glycolysis and a complete pentose phosphate pathway. The tricarboxylic cycle was incomplete in the *Izimaplasma* genomes, with lactate being the probable end point of fermentation (Figure 3). Both *Izimaplasma* genomes predicted the ability to metabolize glucose, sucrose and maltose as carbon sources. In addition, *Izimaplasma* sp. HR1 has the genes to utilize fructose and trehalose, while *Izimaplasma* sp. HR2 contained the genes to metabolize galactose and cellulose (Figure 3). Both genomes also contained the arginine dihydrolase pathway that has been previously described in

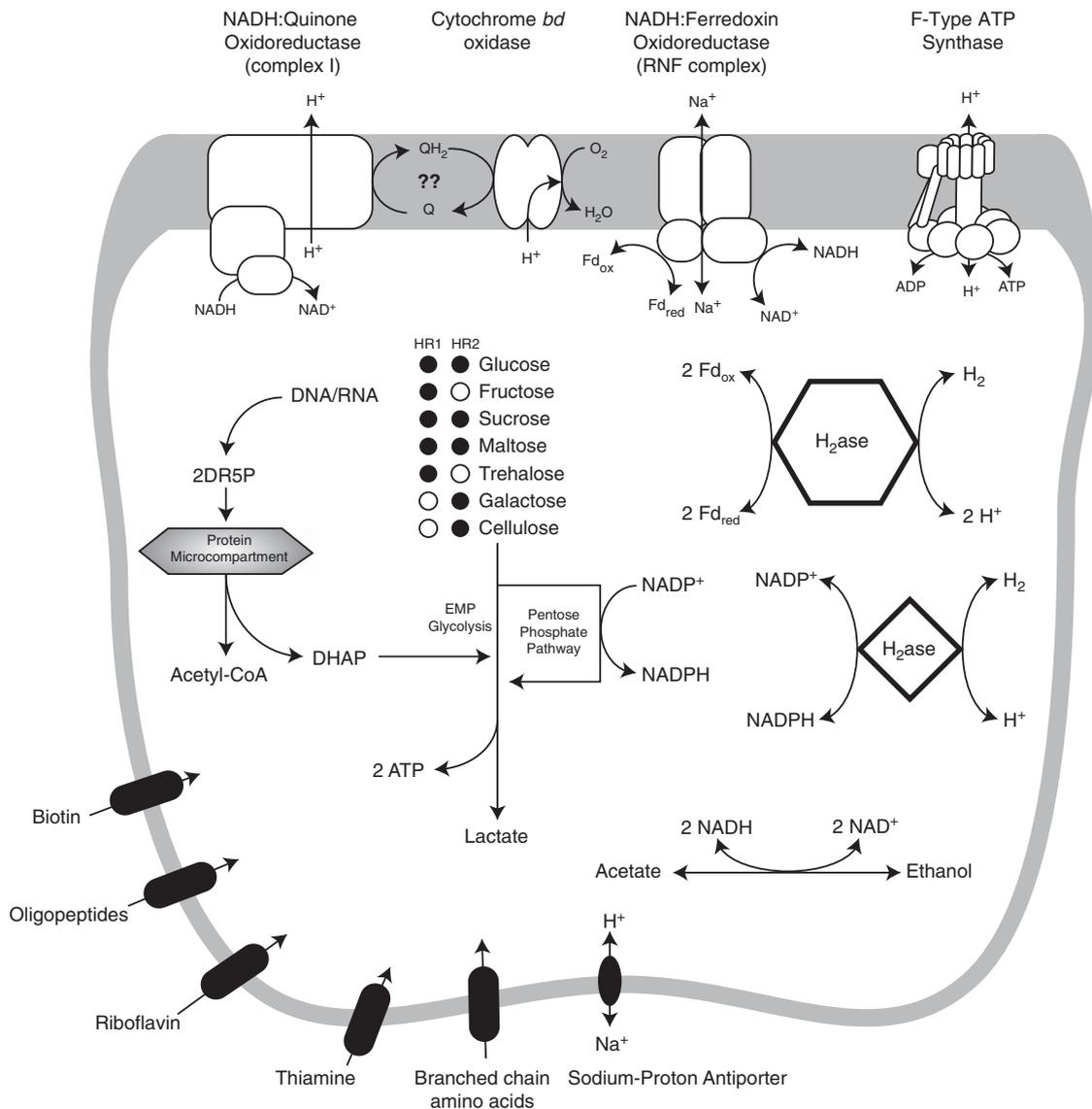


Figure 3 Model predictions of the central metabolism inferred from *Izimaplasma* genomes. Both genomes illustrate the potential for metabolizing a variety of simple sugar compounds that feed into the Embden–Meyerhof–Parnas glycolysis pathway (symbolized by black circles). Reducing equivalents (NADH) generated through glycolysis can be oxidized either by the conversion of acetate to ethanol, using a membrane-bound, sodium-pumping NADH:Ferredoxin oxidoreductase (RNF complex) or by a proton-pumping NADH:Quinone oxidoreductase (complex I). Both complex I and the cytochrome *bd* oxidase interact with the quinone pool, however no quinone synthesis genes were identified in the genome. Two types of hydrogenases act in catabolism to regenerate oxidized ferredoxin, or in anabolism to regenerate NADPH used during lipid and nucleotide synthesis. *Izimaplasma* sp. HR2 contains a protein microcompartment that links the degradation of DNA or RNA to the production of acetyl-CoA and intermediates of glycolysis. 2DR5P, 2-deoxy-D-ribose 5-phosphate; DHAP, dihydroxyacetone phosphate.

Mollicutes as a source of ATP through the oxidation of arginine to ammonia (Razin, 2006).

Reducing equivalents generated by anaerobic fermentation have the potential to be converted to hydrogen by *Izimaplasma*. Both of the *Izimaplasma* genomes contain two types of iron hydrogenases along with hydrogenase maturation genes (*hydEFG*). The first hydrogenase couples with ferredoxin and could be used to remove excess reducing equivalents during anaerobic fermentation (Figure 3). The second hydrogenase is linked with NADP and suggests that it may generate NADPH for the production of biomolecules (Figure 3; Malki *et al.*, 1995). Hydrogenases are not present in mollicute genomes but are found in *Erysipelotrichales*, *H. contractile* and *T. sanguinis* (Supplementary Figure S8).

Both *Izimaplasma* genomes contained homologs of the sodium translocating NADH:Ferredoxin oxidoreductase (RNF complex) and the proton translocating NADH:Quinone oxidoreductase (complex I). The RNF complex has been described from the genomes of some *Acholeplasmatales* and many of the *Erysipelotrichales*; however, complex I has not been found in any previously described mollicute (Supplementary Figure S8). The *Izimaplasma* genomes encoded three copies of the RNF complex, two of which contained the same operon structure (*rnfCDGEAB*) seen in clostridium species (Biegel *et al.*, 2011). The third copy of the *rnf* gene cluster lacked the *rnfC* subunit and instead contained homologs of cytochrome *bd* oxidase (*appCD*). RnfC is an iron–sulfur, flavin mononucleotide containing protein that is responsible for oxidizing NADH (Reyes-Prieto *et al.*, 2014), without this subunit this copy of the RNF complex may not be functional or may need to be compensated by the RnfC subunit found in the other two operons.

Phylogenetic reconstruction of the cytochrome *bd* oxidase protein AppC revealed that the closest relative of the *Izimaplasma* cytochrome *bd* oxidase was to *Methanosarcina acetivorans*, a member of the archaeal order *Methanosarcinales* (Supplementary Figure S9). However, the low (~30%) percentage identity and long branch length between the two proteins suggests that this was not due to a recent horizontal gene transfer between these two organisms. A search of other AppC in all of the metagenomic contigs failed to identify a more closely related AppC than the *M. acetivorans* protein, which suggests that *Izimaplasma* has not acquired these genes through a recent lateral gene transfer from other members of the sediment community.

The subunits for complex I were split into two separate operons, with the majority of the subunits (*nuoKNLHBCD*) comprising the P- and Q-modules found adjacent to a sodium-proton antiporter (*mnhEFGB*) and the subunits of the N-module (*nuoEFG*) found in a separate operon. Four subunits (*nuoAIJM*) from the P- and Q-modules could not be identified, however the adjacent sodium-proton

antiporter is the evolutionary progenitor of the P-module of complex I (Moparathi and Hägerhäll, 2011), it is possible that these genes function as part of complex I in *Izimaplasma*. Comparisons of the complex I subunits to previously sequenced genomes showed highest similarities to members of the *Thermotogae* and *Thermoanaerobacteriales*. The larger operon containing *nuoKNLHBCD* were most closely related to *Marinitoga piezophilla* (phylum *Thermotogae*; average 60% amino acid identity), which also contains the same gene order as the operon in *Izimaplasma*, whereas the operon containing the N-module (*nuoEFG*) were most closely related to *Mahella australiensis* (phylum *Firmicutes*, order *Thermoanaerobacteriales*).

In other organisms, complex I and cytochrome *bd* oxidase interact with the quinone pool, however *Izimaplasma* lacked a complete quinone synthesis pathway. Both *Izimaplasma* genomes contain the non-mevalonate (MEP/DOXP) pathway to generate the quinone precursors Isopentenyl-PP and Dimethylallyl-PP, but do not contain the rest of the pathway. Complex I and cytochrome *bd* oxidase may be non-functional without internal quinone production, an alternative electron carrier, or an external source of quinone.

Protein microcompartment for nucleotide metabolism

Both *Izimaplasma* genomes contained similar metabolic potential, however *Izimaplasma* sp. HR2 contained an additional large operon, comprised of 14 genes that appeared to encode a protein microcompartment. Bacterial microcompartment genes have been identified in many microorganisms and appear to be horizontally transferred (Kerfeld *et al.*, 2010). These protein-based compartments encapsulate enzymes and cofactors to promote specific metabolic pathways including carbon fixation, 1,2-propanediol, ethanol and ethanamine metabolism and are found in ~20% of sequenced genomes (Stojiljkovic *et al.*, 1995; Sedorf *et al.*, 2008; Kerfeld *et al.*, 2010; Erbilgin *et al.*, 2014). With the exception of carbon fixation, the other metabolic pathways encapsulated within microcompartments produce aldehyde intermediates that may be toxic to the cell. It has been proposed that encapsulating these reactions within a microcompartment protects the cell from these intermediates and allows for a more efficient metabolism of these compounds (Kerfeld *et al.*, 2010). *Izimaplasma* sp. HR2 contained a 10-kbp operon containing structural genes for generating a microcompartment plus a number of catalytic genes (Figure 4). An aldehyde dehydrogenase was found in this operon that suggests that *Izimaplasma* sp. HR2 also uses a microcompartment for detoxification purposes. However, examination of the other catalytic genes in the microcompartment-associated operon pointed to the possibility of an alternate, previously

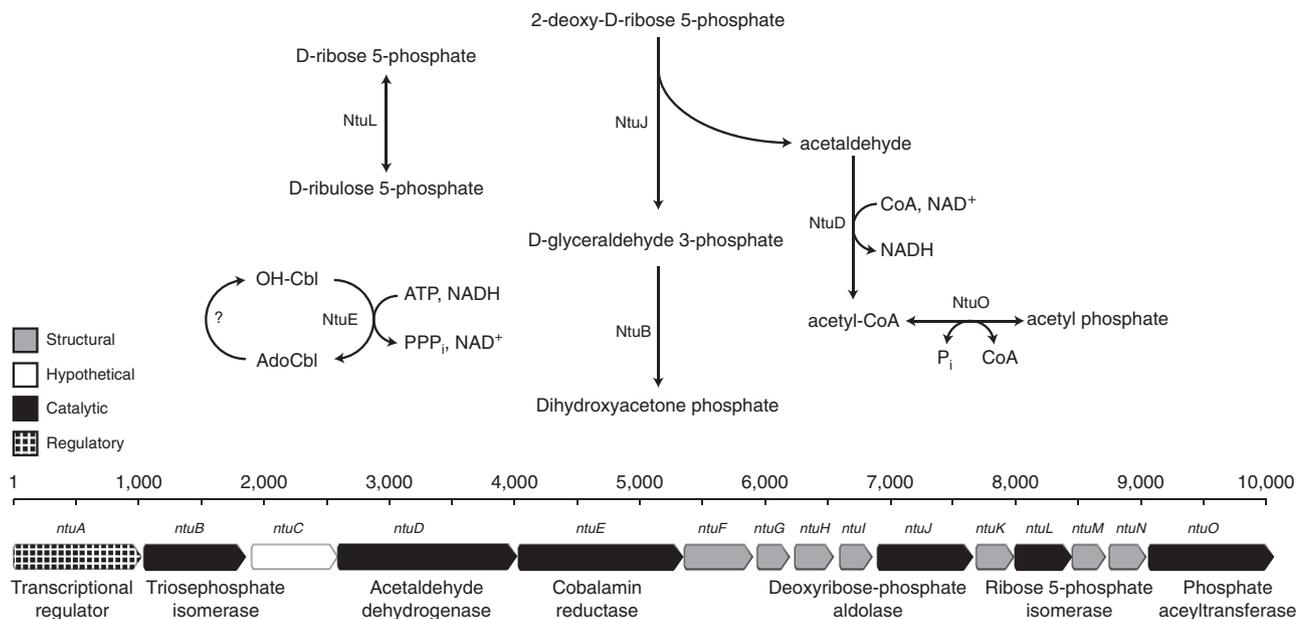


Figure 4 Proposed model of DNA/RNA degradation by the protein microcompartment found in *Izimaplasma* sp. HR2. The operon encoding the structural and catalytic genes of this microcompartment, which we propose to call the nucleotide utilizing operon (*ntu*), is shown below. The reactions catalyzed by this operon are shown above and labeled with the protein that catalyzes each reaction. Two proteins for cofactor recycling (*NtuE*, *NtuO*) are also found; however, the protein requiring a cobalamin cofactor is unknown (represented by ?). AdoCbl, adenosylcobalamin; OH-Cbl, hydroxocobalamin.

undescribed metabolism related to nucleic acid degradation. A complete pathway was predicted that would enable strain HR2 to convert 2-deoxy-D-ribose 5-phosphate, an intermediate in nucleotide metabolism to dihydroxyacetone phosphate, an intermediate in glycolysis. The first step in this reaction, the conversion of 2-deoxy-D-ribose 5-phosphate to D-glyceraldehyde 3-phosphate produces acetaldehyde as a byproduct (Figure 4). Acetaldehyde can then be converted to acetyl-CoA using the acetaldehyde dehydrogenase (*ntuD*) and utilized along with dihydroxyacetone phosphate as carbon and energy sources. Two genes in the *Izimaplasma* sp. HR2 microcompartment operon (*ntuEO*) could be utilized for the purposes of recycling cofactors vitamin B12 and Coenzyme A (CoA), which is consistent with previous reports that microcompartments contain their own pool of cofactors (Huseby and Roth, 2013). Despite only being present in *Izimaplasma* sp. HR2 genome, the protein microcompartment operon does not appear to add novel metabolic functionality as the ability to metabolize 2-deoxy-D-ribose 5-phosphate is also present in the *Izimaplasma* sp. HR1 genome. Furthermore, homologs of phosphate acetyltransferase (*ntuO*) and deoxyribose-phosphate aldolase (*ntuI*) genes are found in the majority of mollicute genomes (Supplementary Figure S8) but not associated with structural microcompartment genes. Although these genes appear consistent with the degradation of nucleotides within the microcompartment, cultivation of a representative of HR2 may be required to better

understand the function of this ultrastructural feature.

Visual identification and enrichment cultivation of *Izimaplasma*

The occurrence of members of clade ‘NB1-n’ in the #5579 incubation was assessed by FISH microscopy using a newly designed oligonucleotide probe ‘Izzy659’ targeting members of this clade (Supplementary Figure S2). A small number of coccoid-shaped cells (0.6–1 µm diameter) were positively hybridized in the seep sediment, observed either as single cells or doublets (Figure 5). To try and culture this organism, we used the predicted metabolism and phenotype from the HR1- and HR2-annotated genomes as a guide for designing enrichment medium, adding simple sugars (for example, glucose) as a carbon source and cell wall acting antibiotics to select for members of the *Tenericutes*. After a 6-week incubation at 10 °C, the community was dominated by two cell types: small coccoid cells (0.5–1.0 µm diameter) that positively hybridized to the ‘NB1-n’-specific probe, and larger coccoid cells that only hybridized to the general bacterial probe (Figure 5). This community composition remained stable for over five passages through new enrichment media and isolation attempts are still ongoing. Illumina 16S rRNA gene sequencing (iTAG) confirmed our FISH observations, with ~60% of the recovered reads in the incubations assigned to clade ‘NB1-n’ (Supplementary Figure S5). There were three ‘NB1-n’ operational taxonomic units

defined from the iTAG data, one of which was dominant, accounting for 90% of the 'NB1-n' sequences and was 98% identical to the region of the 16S rRNA gene recovered from the HR1 genome bin. Electron cryotomography of the enrichment culture confirmed the presence of coccoid cells lacking a cell wall, with a membrane architecture comparable to the Mollicute, *Mycoplasma pneumonia* (Figure 6). Representative cells of 'NB1-n' in the enrichment analyzed by electron cryotomography were 1 µm in diameter and devoid of any discernable intracellular structure.

Discussion

Members of the class *Mollicutes* are some of the smallest self-replicating organisms discovered in nature (Razin, 2006). Currently described isolates have been recovered from a variety of plant and animal hosts, and many members are important to veterinary and agricultural sciences as pathogens (Razin *et al.*, 1998; Razin, 2006). Environmental clades, identified solely through 16S rRNA gene diversity studies, have been affiliated with the class *Mollicutes*. Sequences from one of these clades,

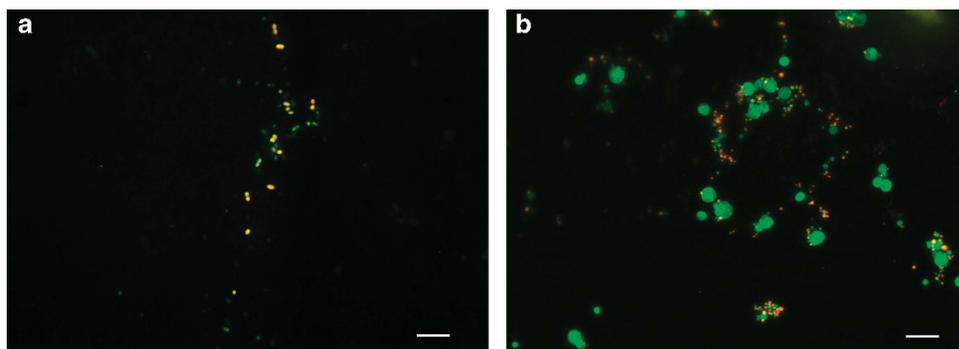


Figure 5 Fluorescence *in situ* hybridization (FISH) micrograph of microorganisms recovered from a methane seep sediment incubation. Cells targeted by the general bacterial probe, EUB338, are shown in green (FITC); cells targeted by the specific probe for *Izimaplasma* (Izzy659) are shown in yellow (combined FITC and Cy3 fluorescence). The scale bar in both images is 10 µm. (a) FISH image of microorganisms from the initial sediment sample used for the *Izimaplasma* enrichment. (b) Microbial diversity after 6 weeks of cultivation in glucose-amended enrichment media.

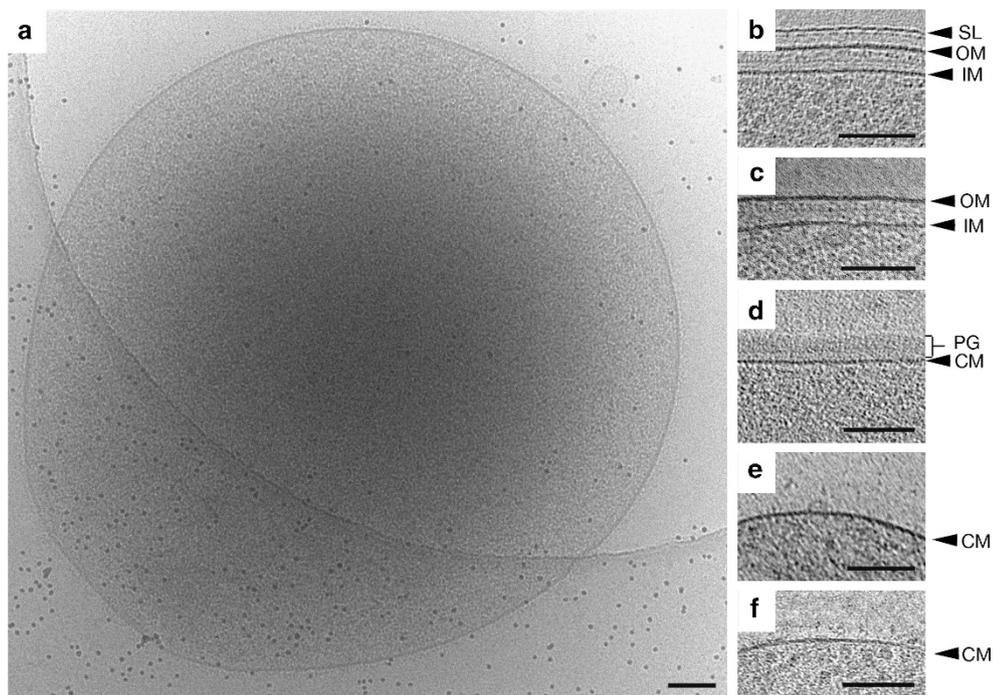


Figure 6 (a) Electron cryomicrograph of an *Izimaplasma* cell. (b–f) Bacterial cell envelope architectures from slices of cryotomograms. (b) Gram-negative cell wall with S-layer: *Caulobacter crescentus*; (c) Gram-negative cell wall without S-layer: *Vibrio cholera*; (d) Gram-positive cell wall: *Listeria monocytogenes*; (e) cell-wall less: *Mycoplasma pneumonia*; (f) cell-wall less: *Izimaplasma*. CM, cell membrane; IM, inner membrane; OM, outer membrane; PG, peptidoglycan; SL, S-layer. Scale bars, 100 nm.

'NB1-n', have been previously reported from anaerobic digesters, hypersaline microbial mats, marine sediments and rumen (Supplementary Figure S2). These environments, with the exception of rumen, are distinct from the plant- and animal-affiliated habitats that are well known environments for *Mollicutes* and raises the possibility that members of clade 'NB1-n' are free-living rather than host-associated organisms. Our work here using a combination of metagenomics and enrichment cultivation confirms the free-living nature of the *Tenericutes*-associated Izimaplasma, demonstrating that these microorganisms are capable of growth under cold, anaerobic conditions in the absence of a eukaryotic host.

The taxonomy of *Mollicutes* has been a source of debate. *Mollicutes* are currently classified as the sole class within the phylum *Tenericutes* based on their unique phenotypic properties and the phylogeny of non-ribosomal phylogenetic markers (Ludwig *et al.*, 2009). Conversely, phylogenetic analysis of the 16S rRNA gene, 23S rRNA gene and a concatenated alignment of ribosomal proteins place the *Mollicutes* within the phylum *Firmicutes*, closely associated with the firmicute class *Erysipelotrichia* (Davis *et al.*, 2013; Yutin and Galperin, 2013). The observation that *Mollicutes* do not share a common root with the *Firmicutes* for some non-ribosomal proteins (Ludwig and Schleifer, 2005) is not fully supported with current genomic data (Supplementary Figures S3 and S4). The RpoB phylogeny places *Mycoplasma* near the base of the bacterial tree, however the *Acholeplasma* branch with the *Firmicutes*. It has been proposed that the acholeplasmas were the first *Mollicutes* to diverge from an ancestral firmicute and that the *Mycoplasma* diverged through extensive genome reduction (Brown *et al.*, 2011). The positioning of the *Mycoplasma* near the base of the RpoB tree may be the result of accelerated evolution that can occur when bacteria become dependent on host organisms that obfuscates molecular phylogeny analysis (McCutcheon and Moran, 2012). Despite the variability of phylogenetic placement of the *Mollicutes*, the Izimaplasma genomes branched consistently with members of the *Acholeplasmatales*. The next most phylogenetically related organisms, *H. contractile* and *T. sanguinis*, consistently branched together when using ribosomal or non-ribosomal genes, which suggest that they may be more closely related than suggested by their current taxonomic classifications. Visual identification of Izimaplasma cells using FISH and electron microscopy showed that Izimaplasma were not associated with eukaryotes and carried the defining membrane architecture of all previously studied *Tenericutes*—distinct from the Gram-positive cell wall found within members of the *Firmicutes* (Figures 5 and 6).

The size and metabolic completeness of the Izimaplasma genomes is comparable to free-living streamlined marine bacteria like *Prochlorococcus* and *Pelagibacter* (Supplementary Figure S6). In

contrast, bacteria that are more reliant on host organisms typically undergo genome reduction through the removal of core metabolic properties, and also tend to accumulate pseudogenes and have a higher overall percentage of non-coding DNA (reviewed in Giovannoni *et al.*, 2014). There are a number of unique metabolic features that distinguish Izimaplasma from the currently described host-associated *Mollicutes* (Supplementary Figure S8). Homologs to the Izimaplasma hydrogenases and the RNF complex were only detected in related bacteria including *H. contractile* and *T. sanguinis* as well as some representatives from the *Erysipelotrichia* (Supplementary Figure S8). The presence of cytochrome *bd* oxidase and complex I in Izimaplasma is unique among all currently described *Mollicutes* (Supplementary Figure S8). The Izimaplasma genes and their closest homologs from archaea, *Thermotogae* and *Thermoanaerobacteriales* showed low-sequence identity suggesting that these genes have not been recently laterally transferred, despite Izimaplasma being recovered from seep sediments dominated by methanotrophic ANME archaea. One possible explanation for these observations is that the ancestor of current Izimaplasma accumulated membrane-bound electron complexes from multiple donor organisms. It is also possible that a single, as yet unsequenced microorganism, was the single donor to Izimaplasma.

The presence of hydrogenases and electron transport complexes may provide Izimaplasma with multiple means of cycling redox-active electron carriers like ferredoxin and NAD(P)H. The presence of cytochrome *bd* oxidase and complex I suggests a simplified electron transport chain in Izimaplasma, where electrons from NADH are transferred to the quinone pool and are subsequently utilized by cytochrome *bd* oxidase to reduce oxygen (Figure 3). This model has implications for the potential function of the F-Type ATPase in Izimaplasma, as in many *Mollicutes*, the F-type ATPase functions to maintain a proton gradient by hydrolyzing ATP, rather than through ATP generation (Pollack *et al.*, 1997). In Izimaplasma, the presence of multiple membrane-bound ion pumping complexes could drastically change how Izimaplasma generates ATP. The activity of these ion pumping complexes may allow for a sufficient proton gradient for the ATPase to generate ATP.

One inconsistency with this model is the absence of a complete quinone synthesis pathway in the Izimaplasma genomes. If indeed this predicted complex I and *bd* cytochrome oxidase are functional in Izimaplasma, a source of quinone must be acquired from an external source—perhaps passed through associations with other organisms in the methane seep microbial community. Membrane vesicles, which are produced by many Gram-negative bacteria (Beveridge, 1999) have been shown to fuse with mollicute membranes (Grant and McConnell, 1973) and could represent a mechanism to

deliver quinone to Izimaplasma (Bergsma *et al.*, 1982). Alternatively, Izimaplasma may contain an, as yet undescribed, pathway to synthesize quinone or potentially use an alternative membrane-soluble electron carrier. The expression and functionality of these enzymes can only be determined from further biochemical analysis; however, the genetic evidence does suggest that these are not pseudogenes. There are no internal stop codons or frameshifts in either operon and the Izimaplasma-associated cytochrome *bd* oxidase contains characteristic amino acids for quinone and heme binding (Zhang *et al.*, 2004; Mogi *et al.*, 2006; Yang *et al.*, 2007). The selective pressure for Izimaplasma to retain these electron transport complexes and their physiological role in these organisms can only be speculated. In model organisms, such as *Escherichia coli*, cytochrome *bd* oxidase functions under microaerophilic conditions to protect the anaerobic metabolism from oxygen poisoning (Borisov *et al.*, 2011). It may be that Izimaplasma are exposed to variable levels of oxygen *in situ* within the seep environment, where bioturbation and methane bubble ebullition can mix oxygenated seawater into the shallow seafloor sediments. Here the cytochrome *bd* complex may serve as an efficient way for Izimaplasma to protect itself from oxygen and generate proton motive force to generate extra ATP.

The most striking difference between the two genomes was the presence of a microcompartment operon in Izimaplasma sp. HR2. Analysis of the enzymatic genes in this operon suggested that the compartment is used for the efficient metabolism of purines and pyrimidines. This proposed metabolism is unique among currently known microcompartments but there are similar features as the reaction generates acetaldehyde, which has been proposed in other microcompartments to be a toxic intermediate that requires sequestering from the cytoplasm (Kerfeld *et al.*, 2010). The presence of the microcompartment suggests that Izimaplasma sp. HR2 may be more efficient at metabolizing nucleic acids. A comparison with the Izimaplasma sp. HR1 genome showed that the complement of purine and pyrimidine metabolic genes is identical between the two genomes, which suggests that Izimaplasma sp. HR1 is also capable of degrading nucleic acids but this may occur at a slower rate. Electron cryomicroscopy did not reveal any discernable microcompartments in cells from the enrichment culture, however the dominant operational taxonomic unit in this mixed culture was closely related to Izimaplasma sp. HR1, which lacks the microcompartment genes. Additional genome sequencing is needed to assess how widespread these microcompartment genes are within the Iziamplasma.

The deep phylogenetic placement of Izimaplasma, basal to the *Acholeplasmatales*, may warrant the creation of a novel class of free-living wall-less bacteria within the phylum *Tenericutes*. These organisms could represent an intermediate

in the reductive evolution from *Firmicutes* to *Mollicutes*. The lack of an obvious host organism suggests that cell-wall loss may have occurred before symbiotic relationships with Eukaryotes were formed. However, recent discovery of symbiotic *Mollicutes* of Fungi that have retained their cell wall suggests that cell-wall loss has occurred multiple times (Desirò *et al.*, 2014; Naito *et al.*, 2015; Torres-Cortés *et al.*, 2015). There is also the possibility that Izimaplasma have non-obligate symbiotic relationships with micro-eukaryotes that were undetected in the incubations or sediment samples analyzed. The successful enrichment and passage of Izimaplasma relatives from methane seep sediment incubations suggests that culturing efforts and sequencing of environmental DNA from marine sediments may result in pure culture isolates for further taxonomic classification and physiological experiments.

The metabolic reconstruction and enrichment cultivation results indicate that Izimaplasma do not directly take part in the methane and sulfur cycles that are dominant in methane seeps. Instead, members of the Izimaplasma, representing a rare member of the seep microbial community, appear to be heterotrophs engaging primarily in fermentative metabolism, consisting of simple sugars, small organic acids, arginine and possibly nucleotides generated during the degradation of organic matter. These genomic predictions enabled the design of anaerobic culture media to selectively enrich for Izimaplasma from methane seep sediments. Unlike previously described mollicute genomes, Izimaplasma contain homologs to cytochrome *bd* and to complex I that may form a simplified electron transport chain. Their predicted role in the community is likely in the fermentation of products from the degradation of organic matter to produce lactate and possibly other small molecules such as acetate or ethanol, which in turn may be utilized by other members of the methane seep microbial community.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work is funded by the Gordon and Betty Moore Foundation through Grant GBMF3780 and the US Department of Energy, Office of Science, Office of Biological Environmental Research under award numbers DE-SC0003940 and DE-SC0010574 (to VJO). Additional support was provided by Caltech's Center for Environmental Microbial Interactions and the Howard Hughes Medical Institute. We thank D Smith, S Scheller, S Zinder, J Hemp and D Newman for helpful discussions and M Imelfort for bioinformatics assistance. GWT was supported by an ARC Queen Elizabeth II fellowship (ARC-DP1093175).

References

- Antunes A, Alam I, El Dorry H, Siam R, Robertson A, Bajic VB *et al.* (2011). Genome sequence of *Haloplasma contractile*, an unusual contractile bacterium from a deep-sea anoxic brine lake. *J Bacteriol* **193**: 4551–4552.
- Antunes A, Rainey FA, Wanner G, Taborda M, Pätzold J, Nobre MF *et al.* (2008). A new lineage of halophilic, wall-less, contractile bacteria from a brine-filled deep of the Red Sea. *J Bacteriol* **190**: 3580–3587.
- Bergsma J, Meihuizen KE, Van Oeveren W, Konings WN (1982). Restoration of NADH oxidation with menaquinones and menaquinone analogues in membrane vesicles from the menaquinone-deficient *Bacillus subtilis* aroD. *Eur J Biochem* **125**: 651–657.
- Beveridge TJ. (1999). Structures of gram-negative cell walls and their derived membrane vesicles. *J Bacteriol* **181**: 4725–4733.
- Biegel E, Schmidt S, González JM, Müller V. (2011). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell Mol Life Sci* **68**: 613–634.
- Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. (2011). The cytochrome bd respiratory oxygen reductases. *Biochim Biophys Acta* **1807**: 1398–1413.
- Brown DR, May M, Bradbury JM, Johansson K-E. (2011). Class I. *Mollicutes*. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL *et al.* (eds). *Bergey's Manual of Systematic Bacteriology: The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*, vol. 4. Springer: New York, pp 568–573.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K *et al.* (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Cuív PÓ, Klaassens ES, Durkin AS, Harkins DM, Foster L, McCorrison J *et al.* (2011). Draft genome sequence of *Turicibacter sanguinis* PC909, isolated from human feces. *J Bacteriol* **193**: 1288–1289.
- Davis JJ, Xia F, Overbeek RA, Olsen GJ. (2013). Genomes of the class *Erysipelotrichia* clarify the firmicute origin of the class *Mollicutes*. *Int J Syst Evol Microbiol* **63**: 2727–2741.
- Desirò A, Salvioli A, Ngonkeu EL, Mondo SJ, Epis S, Faccio A *et al.* (2014). Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. *ISME J* **8**: 257–270.
- Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SAFT, Harhangi HR, Schmid M *et al.* (2011). FACIL: fast and accurate genetic code inference and logo. *Bioinformatics* **27**: 1929–1933.
- Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Erbilgin O, McDonald KL, Kerfeld CA. (2014). Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl Environ Microbiol* **80**: 2193–2205.
- Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA *et al.* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52–56.
- Giovannoni SJ, Cameron Thrash J, Temperton B. (2014). Implications of streamlining theory for microbial ecology. *ISME J* **8**: 1553–1565.
- Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M *et al.* (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103**: 425–430.
- Grant CW, McConnell HM. (1973). Fusion of phospholipid vesicles with viable *Acholeplasma laidlawii*. *Proc Natl Acad Sci USA* **70**: 1238–1240.
- Huseby DL, Roth JR. (2013). Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J Bacteriol* **195**: 2864–2879.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. (2014). GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**: e603.
- Kerfeld CA, Heinhorst S, Cannon GC. (2010). Bacterial microcompartments. *Annu Rev Microbiol* **64**: 391–408.
- Kremer JR, Mastrorarde DN, McIntosh JR. (1996). Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**: 71–76.
- Kube M, Siewert C, Migdoll AM, Duduk B, Holz S, Rabus R *et al.* (2013). Analysis of the complete genomes of *Acholeplasma brassicae*, *A. palmae* and *A. laidlawii* and their comparison to the obligate parasites from “Candidatus Phytoplasma”. *J Mol Microbiol Biotechnol* **24**: 19–36.
- Laslett D, Canback B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **32**: 11–16.
- Lazarev VN, Levitskii SA, Basovskii YI, Chukin MM, Akopian TA, Vereshchagin VV *et al.* (2011). Complete genome and proteome of *Acholeplasma laidlawii*. *J Bacteriol* **193**: 4943–4953.
- Lechner M, Findeiß S, Steiner L, Marz M, Stadler P, Prohaska S. (2011). Proteinortho: detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics* **12**: 124.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN]. Available at <http://arxiv.org/abs/1303.3997v2>.
- Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Ludwig W, Schleifer K-H. (2005). Molecular phylogeny of bacteria based on comparative sequence analysis of conserved genes. In: Sapp J (ed). *Microbial Phylogeny and Evolution: Concepts and Controversies*. Oxford University Press: New York, USA, pp 70–98.
- Ludwig W, Schleifer KH, Whitman WB. (2009). Revised road map to the phylum Firmicutes. In: De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA *et al.* (eds). *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 3. Springer: New York, USA, pp 1–13.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Malki S, Saimmaime I, De Luca G, Rousset M, Dermoun Z, Belaich JP. (1995). Characterization of an operon encoding an NADP-reducing hydrogenase in *Desulfovibrio fructosovorans*. *J Bacteriol* **177**: 2628–2636.

- Mason OU, Case DH, Naehr TH, Lee RW, Thomas RB, Bailey JV *et al.* (2015). Comparison of archaeal and bacterial diversity in methane seep carbonate nodule and host sediments, Eel River Basin and Hydrate Ridge, USA. *Microb Ecol* **70**: 766–784.
- McCutcheon JP, Moran NA. (2012). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13–26.
- Mogi T, Akimoto S, Endou S, Watanabe-Nakayama T, Mizuochi-Asai E, Miyoshi H. (2006). Probing the ubiquinol-binding site in cytochrome bd by site-directed mutagenesis. *Biochemistry* **45**: 7924–7930.
- Moparthi VK, Hägerhäll C. (2011). The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J Mol Evol* **72**: 484–497.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**: W182–W185.
- Naito M, Morton JB, Pawlowska TE. (2015). Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. *Proc Natl Acad Sci USA* **112**: 7791–7796.
- Nawrocki EP, Eddy SR. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.
- Ogawa Y, Ooka T, Shi F, Ogura Y, Nakayama K, Hayashi T *et al.* (2011). The genome of *Erysipelothrix rhusiopathiae*, the causative agent of swine erysipelas, reveals new insights into the evolution of firmicutes and the organism's intracellular adaptations. *J Bacteriol* **193**: 2959–2971.
- Orphan VJ, Hinrichs KU, Ussler W 3rd, Paull CK, Taylor LT, Sylva SP *et al.* (2001). Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl Environ Microbiol* **67**: 1922–1934.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.
- Pernthaler J, Glöckner F-O, Schönhuber W, Amann R. (2001). Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. In: Paul JH (ed). *Marine Microbiology*, vol. 30. Academic Press: Cambridge, MA, USA, pp 207–226.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786.
- Pollack JD, Williams MV, McElhaney RN. (1997). The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit Rev Microbiol* **23**: 269–354.
- Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
- Razin S, Yogeve D, Naot Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**: 1094–1156.
- Razin S. (2006). The genus *Mycoplasma* and related genera (Class Mollicutes). In: *The Prokaryotes*. Springer: USA, pp 836–904.
- Reyes-Prieto A, Barquera B, Juárez O. (2014). Origin and evolution of the sodium-pumping NADH: ubiquinone oxidoreductase. *PLoS One* **9**: e96696.
- Saunders SE, Burke JF. (1990). Rapid isolation of miniprep DNA for double strand sequencing. *Nucleic Acids Res* **18**: 4948.
- Seedorf H, Fricke WF, Veith B, Brüggemann H, Liesegang H, Strittmatter A *et al.* (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Proc Natl Acad Sci USA* **105**: 2128–2133.
- Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.
- Smith DG, Russell WC, Ingledew WJ, Thirkell D. (1993). Hydrolysis of urea by *Ureaplasma urealyticum* generates a transmembrane potential with resultant ATP synthesis. *J Bacteriol* **175**: 3253–3258.
- Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stojiljkovic I, Bäumlér AJ, Heffron F. (1995). Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. *J Bacteriol* **177**: 1357–1366.
- Tivol WF, Briegel A, Jensen GJ. (2008). An improved cryogen for plunge freezing. *Microsc Microanal* **14**: 375–379.
- Torres-Cortés G, Ghignone S, Bonfante P, Schüßler A. (2015). Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proc Natl Acad Sci USA* **112**: 7785–7790.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL *et al.* (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**: D581–D591.
- Wu D, Jospin G, Eisen JA. (2013). Systematic identification of gene families for use as “markers” for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS One* **8**: e77033.
- Yang K, Zhang J, Vakkasoglu AS, Hielscher R, Osborne JP, Hemp J *et al.* (2007). Glutamate 107 in subunit I of the cytochrome bd quinol oxidase from *Escherichia coli* is protonated and near the heme d/heme b595 binuclear center. *Biochemistry* **46**: 3270–3278.
- Yutin N, Galperin MY. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol* **15**: 2631–2641.
- Zhang J, Hellwig P, Osborne JP, Gennis RB. (2004). Arginine 391 in subunit I of the cytochrome bd quinol oxidase from *Escherichia coli* stabilizes the reduced form of the hemes and is essential for quinol oxidase activity. *J Biol Chem* **279**: 53980–53987.
- Zheng SQ, Keszthelyi B, Branlund E, Lyle JM, Braunfeld MB, Sedat JW *et al.* (2007). UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J Struct Biol* **157**: 138–147.

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