The ISME Journal (2012) 6, 430–439

www.nature.com/ismej

ORIGINAL ARTICLE

A 'universal' type II chaperonin PCR detection system for the investigation of Archaea in complex microbial communities

Bonnie Chaban and Janet E Hill
Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Bacteria and Archaea are evolutionarily and biochemically distinct domains found together in many environments. Robust 'universal' PCR primer sets targeting both the bacterial 16S rRNA gene and the type I chaperonin gene have been established. However, 'universal' PCR primers for Archaea are currently limited to the 16S rRNA gene. We investigated the type II chaperonin (known as the thermosome, TF55, CCT or TCP-1) as a potential universal target (UT) for Archaea. Reproducible amplification of thermosome gene sequences from all major phyla tested was achieved through the application of a mixture or 'cocktail' of two forward and two reverse primers. Phylogenies based on the ~750-bp thermosome UT were congruent with 16S rRNA gene phylogenies while exhibiting longer branch lengths, improving resolution of closely related taxa. 'Universal' thermosome primers were applied to profiling the archaeal community of dairy cow rumen and results compared with profiles based on the 16S rRNA gene and methyl co-enzyme M reductase (methanogen-specific) gene. Clone libraries generated from each target gene, as well as a pyrosequencing profile of one thermosome rumen library, revealed that all three targets consistently detected *Methanobrevibacter smithii, Methanobrevibacter ruminantium* and *Methanosphaera stadtmanae* as the dominant constituents; however, thermosome gene sequences were more diverse than either of the other targets providing a higher resolution description of the archaeal community. These findings demonstrate that a 'universal' thermosome PCR protocol is a powerful metagenomic tool for detecting and characterizing Archaea and archaeal communities.
The ISME Journal (2012) **6,** 430–439; doi:10.1038/ismej.2011.96; published online 21 July 2011
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: 16S rRNA gene; Archaea; chaperonin; rumen; thermosome; 'universal' PCR

Introduction

The prokaryotic world is divided into the domains Bacteria and Archaea. Although these two groups are evolutionarily and biochemically distinct, both are important ecological contributors and are often found together in the same environments (Chaban *et al.*, 2006; Martiny *et al.*, 2006). Thus, a thorough understanding of the microbial component of an environment requires consideration of constituents from both prokaryotic domains. Limitations in culturing efficiencies (Kaeberlein *et al.*, 2002) have lead microbiologists to rely on molecular techniques to investigate the composition of microbial communities. However, because of the genetic distinction between Bacteria and Archaea, molecular targeting of these microbial populations is necessarily domain specific.

Correspondence: JE Hill, Department of Veterinary Microbiology, University of Saskatchewan, 52 Campus Drive, Saskatoon, Saskatchewan, Canada S7N 5B4.
E-mail: janet.hill@usask.ca
Received 24 March 2011; revised 14 June 2011; accepted 17 June 2011; published online 21 July 2011

No PCR primers can be absolutely universal, but the term 'universal' is commonly used to describe PCR primers designed to amplify phylogenetically informative gene sequences from virtually all members of a domain. Among these genes, the 16S rRNA gene is the most commonly targeted (Fox *et al.*, 1980). The appeal of the 16S rRNA gene lies in its 'universal' nature, slow rate of evolution (allowing phylogenetic comparisons of distantly related organisms) and its sequence structure of alternating conserved and variable regions. These properties have facilitated the design of PCR primers that can amplify either the bacterial or archaeal 16S rRNA gene (Baker *et al.*, 2003). As such, the phylogeny of the 16S rRNA gene has become the basis of prokaryotic molecular taxonomy (Woese and Fox, 1977; Woese *et al.*, 1985). To complement the use of the 16S rRNA gene for phylogenetic and metagenomic studies, the Ribosomal Database Project was established (Cole *et al.*, 2009).

Despite the utility of the 16S rRNA gene, there are well-documented shortcomings in assessing microbial diversity and for phylogenetic analysis. While the slow rate of evolution allows for broad taxonomic comparison, this same characteristic results

in nearly identical sequences for closely related organisms, reducing the resolution of phylogenetic trees and making differentiation of these taxa difficult (Wang *et al.*, 2007; Schellenberg *et al.*, 2009; Weng *et al.*, 2009). As well, the regular occurrence of insertions and deletions (indels) in rRNA genes has resulted in challenges for multiple sequence alignments, the first step in phylogenetic analysis (DeSantis *et al.*, 2006). These challenges become exponentially more complex when large data sets, like those from microbial ecology studies, are considered. Finally, there will always be questions as to whether one gene can truly represent an organism's phylogeny, especially in light of horizontal gene transfer, from which the 16S rRNA gene is not exempt (Yap *et al.*, 1999).

An alternative to the 16S rRNA gene has been the utilization of conserved protein-coding genes. These have advantages over 16S rRNA genes in that they are usually present in single copies in prokaryotic genomes, are subject to low rates of indel events and accumulate silent mutations due to codon degeneracy, resulting in better species resolution (Santos and Ochman, 2004). Genes such as *recA* (Weng *et al.*, 2009), *rpoB* (Meintanis *et al.*, 2008; Glazunova *et al.*, 2009; Weng *et al.*, 2009), *recN* (Zeigler, 2005; Arahal *et al.*, 2008), *gyrB* (Wang *et al.*, 2007; Glazunova *et al.*, 2009) and *cpn*60 (Hill *et al.*, 2006a; Glazunova *et al.*, 2009) have been used when the 16S rRNA gene sequence could not provide species resolution. Of these, the type I chaperonin gene, *cpn*60 (also known as *hsp*60 or *groEL*) is the most developed alternative. It is the only target other than 16S rRNA gene that can be accessed with 'universal' PCR primers and a curated sequence database, cpnDB, is available (http://www.cpndb.ca; Hill *et al.*, 2004, 2006b; Schellenberg *et al.*, 2009). The *cpn*60 gene provides greater discriminating power than 16S rRNA gene for closely related taxa and the uniform size and sequence heterogeneity of the *cpn*60 'universal target' (UT) simplify sequence comparisons and other bioinformatics tasks.

Unfortunately, 'universal' PCR protocols for the Archaea have been limited to the archaeal 16S rRNA gene. The major reason for this has been a lack of sequence data from the archaeal domain. With the wealth of complete archaeal genomes now available, it is possible to investigate and evaluate protein-coding genes as potential archaeal UTs.

A promising candidate archaeal UT is the type II chaperonin. Type II chaperonins, also known as thermosomes, TF55, CCT or TCP-1, are found in Archaea and the eukaryotic cytosol (Trent *et al.*, 1991; Kubota *et al.*, 1995; Large and Lund, 2009; Large *et al.*, 2009). In Archaea, organisms possess one to three parologous thermosome genes, giving rise to α, β and γ subunits. The thermosome gene is an appealing target as its bacterial homologue, *cpn*60, has been demonstrated to be an excellent target for species detection, identification and quantification of individual bacterial species and strains as well as the metagenomic characterization of complex microbial communities (Hill *et al.*, 2005; Dumonceaux *et al.*, 2006b; Schellenberg *et al.*, 2009). In addition, the infrastructure needed to translate the thermosome sequence into a useful identification tool is already in place since thermosome sequences have been included in the cpnDB database since its creation (Hill *et al.*, 2004). In the current study, we set out to accomplish three goals: compare 16S rRNA gene-based and thermosome-based phylogenies of the Archaea; design a 'universal' PCR protocol to amplify the thermosome gene; and evaluate the thermosome 'universal' PCR protocol for its ability to detect and distinguish members of the archaeal community compared with established 16S rRNA gene and methanogen-specific (*mcrA* gene) PCR protocols. To accomplish the third objective, rumen samples from dairy cows fed two distinct diets were profiled using all three genetic targets by clone library analysis, and thermosome sequences were further analysed by pyrosequencing.

Materials and methods

Reference DNA sequences

Reference DNA sequences used in this study were taken from two sources: 16S rRNA genes and methyl co-enzyme M reductase α subunit (*mcrA*) genes were from NCBI Genomes (http://www.ncbi.nlm.nih.gov/genomes), while thermosome sequences were taken from cpnDB (http://www.cpndb.ca/; Supplementary Table 1). Eighty-four archaeal strains were identified for which both 16S rRNA gene and thermosome sequences were available. A single, representative 16S rRNA gene sequence was chosen from each strain as well as the thermosome genes from each subunit (from one to three parologues per genome, depending on the species). In addition, full-length *mcrA* genes from 25 methanogen strains (plus the homologous *mrtA* gene from *Methanosphaera stadtmanae* DSM 3091) were also included in the analysis.

Phylogenetic analysis

Multiple sequence alignments of 16S rRNA gene sequences were assembled at the Ribosomal Database Project website using the secondary structure alignment program Infernal (Nawrocki *et al.*, 2009), while thermosome and *mcrA* gene sequences alignments were constructed with ClustalW (Thompson *et al.*, 1997) with a gap opening penalty of 50 and gap extension penalty of 5. Alignments were manually inspected and minor edits were made when necessary with GeneDoc (Nicholas *et al.*, 1997) before trees were constructed using PHYLIP (Felsenstein, 1989). DNA distance matrices were calculated based on the F84 maximum likelihood option and neighbour-joined trees were assembled

as the consensus of 100 replicates. Final trees were visualized with TreeView (Page, 1996).

Archaeal genomic DNA

Archaeal genomic DNA from *Methanococcus voltae* PS, *Methanococcus vannielii* SB, *Methanococcus maripaludis* S2, *Methanotorris igneus* Kol 5, *Sulfolobus solfataricus*, *Sulfolobus* sp., *Thermoplasma acidophilum* and *Halobacterium salinarum* (formerly *Halobacterium halobium*) NRC 34008 was generously donated by Ken Jarrell, Queen's University at Kingston, Canada while *Haloferax volcanii* WR341 and *Hf. volcanii* WR536 were generously donated by Jerry Eichler, Ben-Gurion University of the Negev, Israel. Genomic DNA from *Hb. salinarum* (formerly *Halobacterium cutirubrum*) ATCC 33170, *Hb. salinarum* (formerly *Hb. salinarium*) ATCC 33171, *Thermococcus gorgonarius* ATCC 700654, *Thermococcus pacificus* ATCC 700653 and *Thermococcus zilligii* ATCC 700529 was acquired from the American Type Culture Collection (Manassas, VA, USA).

Rumen sample collection and processing

Rumen contents (both solid and liquor) were collected from fistulated dairy cows housed at the University of Saskatchewan dairy barn. Six cows on a 'regular' high milk production dairy cow diet and two cows on a modified 'dry' diet were sampled $\sim$2 h after feeding and DNA was extracted immediately. Samples were thoroughly mixed before a 300-μl volume of solid and liquor was removed for processing. Total DNA extraction from the samples was accomplished as previously described (Dumonceaux *et al.*, 2006a). Final DNA extracts were suspended in Tris-EDTA buffer, pH 8.0 and stored at −20 °C.

PCR primers and conditions

Thermosome-specific archaeal primers were designed from a multiple sequence alignment of 166 thermosome sequences (Table 1; Figure 1). The thermosome primer pair, JH0175 and JH0178, is theoretically appropriate for the entire archaeal domain and performed best on low to mid-range GC content organisms. An alternative version of these primers, JH0268 and JH0269, contains no degeneracies and was engineered to target high GC organisms.

Standard thermosome PCR consisted of 1 × PCR reaction buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 2.5 mM $MgCl_2$, 200 μM dNTP, 400 nM of each forward and reverse primer, 2.5 U Platinum Taq DNA Polymerase (Invitrogen, Burlington, Canada) and 1.0 ng template DNA, carried out in a final volume of 50 μl. A Mastercycler thermocycler (Eppendorf, Mississauga, Canada) or a MyiQ thermocycler (Bio-Rad, Mississauga, Canada) was used with initial denaturing at 98 °C for 3 min, followed by 40 cycles of 30 s at 98 °C, 1 min at 54 °C and 1 min at 72 °C, followed by a final extension at 72 °C for 10 min. Annealing temperature gradients (45–66 °C) were tested with both thermosome primer sets (individually and in combination) with several archaeal genomic DNAs (spanning the range of GC contents from 28% to 66%) and 54 °C was determined to be the optimal annealing temperature (data not shown). For the generation of JH0175/JH0178 PCR products from rumen samples, a pool of PCR products was generated from four PCRs with annealing temperatures of 56.7 °C, 53.6 °C, 49.2 °C and 46.3 °C, respectively. For the generation of thermosome PCR products from regular diet rumen samples using a 7:1 molar ratio cocktail of JH0175/JH0178:JH0268/JH0269, the standard thermosome PCR program (annealing temperature of 54 °C only) was used. The primer cocktail experiments conducted to determine that a 7:1 molar ratio mix of JH0175/JH0178:JH0268/JH0269 was the optimal ratio to allow for amplification of a broad spectrum of GC content sequences in a single PCR are detailed in Supplementary materials.

The 16S rRNA gene primer set and appropriate PCR protocol were taken from Baker *et al.* (2003), while the methyl co-enzyme M reductase α subunit (*mcrA*) gene-specific primer set and appropriate PCR protocol were taken from Mihajlovski *et al.* (2008) (Table 1). The Mihajlovski *et al. mcrA* gene primers are a slight modification of the original *mcrA* gene primer set proposed by Luton *et al.* (2002).

Table 1 PCR primers used in this study

Primer name	*Primer sequence (5′–3′)*	*Target*	*Source*
JH0175	GGI CCI MRR GGI ITI GAY AAR ATG	Universal target thermosome	This study
JH0178	GCI AII TCR TCI ATI CCY TTY TG		
JH0268	GGC CCG AAG GGC ATG GAC AAG ATG	High GC universal target thermosome	This study
JH0269	GGC ATG TCG TCG ATG CCC TTC TG		
A571F	GCY TAA AGS RIC CGT AGC	Archaeal 16S rRNA gene	Baker *et al.*, 2003
UA1240R	TTM GGG GCA TRC IKA CCT		
MM_01	TAY ATG TCI GGY GGT GTH GG	Methyl co-enzyme M reductase α subunit (*mcrA*) gene	Mihajlovski *et al.*, 2008
MM_02	ACR TTC ATI GCR TAG TTI GG		

I = inosine, M = A or C, R = A or G, Y = C or T, K = G or T and H = A or C or T.

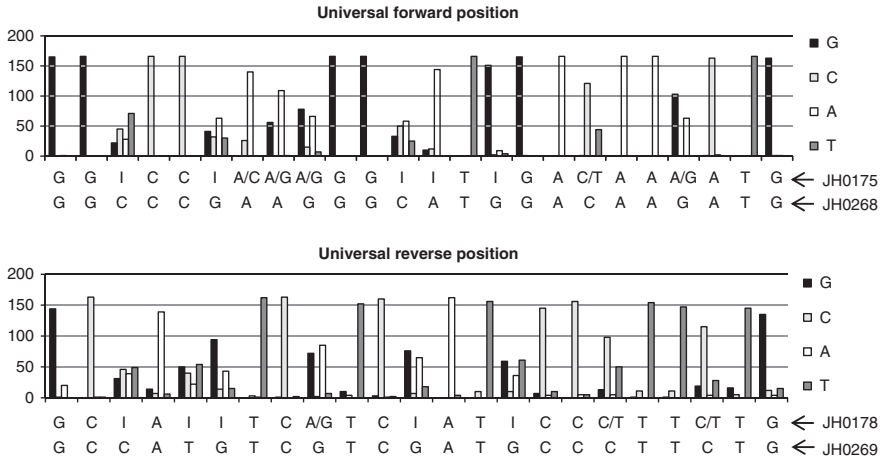

Figure 1 Nucleotide frequencies for the thermosome (group II chaperonin) primer annealing sites. The frequency of each nucleotide in each position in the 166 sequence alignment is indicated. Where more than two different nucleotides were common, inosine (I) was used in JH0175/JH0178 to reduce degeneracy. The graphs are shown such that the *x*-axis depicts the primer sequences in their 5′–3′ orientation. The 'high GC' JH0268/JH0269 primers are depicted directly below their counterparts to highlight how and where degeneracy was removed.

Clone library construction and sequencing

For comparison of the archaeal communities in dairy cow rumen, PCR products from each target gene (thermosome, 16S rRNA, *mcrA*) and each diet (regular and dry) were cloned into the vector pGEM-T Easy (Invitrogen) as previously described (Hill *et al.*, 2005). A total of 96 white colonies were picked randomly from each library (192 colonies for the 7:1 tcp regular library) and sequenced.

454 pyrosequencing

In addition to the clone library, the thermosome PCR product pool generated from the dry diet rumen samples (tcpdry) was sequenced on a 454 GS FLX Titanium instrument (454 Life Sciences, Branford, CT, USA). PCR products were sequenced directly as an untagged sample within a larger multiplexed run. Sequences were identified as originating from the tcpdry pool if 12–24 bp of either the forward or the reverse primer (from the 5′-end) was detected at the beginning of the sequence.

Clone library and pyrosequencing data processing

Raw sequence data were trimmed for quality and PCR primer/vector sequences and assembled into clusters of identical sequences using APED (http://sourceforge.net/projects/aped). A combination of BLAST and Smith–Waterman alignments (watered-BLAST; Schellenberg *et al.*, 2009) was used to identify the best match (nearest neighbour) for each individual sequence read from an appropriate reference sequence collection (Supplementary Table 1). Results were filtered to remove matches shorter than 100 bp to ensure reliable identifications. The distribution of nearest neighbour percent identities within the clone library data sets showed a clear distinction between target sequences (99–75% for all three targets) and non-target sequences (49% and lower; better non-target identities were obtained from the GenBank nr database). The thermosome pyrosequencing data set was filtered for nearest neighbour percent identities $>50\%$ (based on the thermosome clone library analysis), with sequences having between 50% and 75% identity to a thermosome nearest neighbour confirmed as thermosome by Blastx (Altschul *et al.*, 1990) search with the GenBank nr database. Sequences generated in this study have been deposited in GenBank under the accession numbers HQ268028–HQ268244 and JF717634–JF717664.

Statistical analysis

Statistical analyses were done using SPSS software (SPSS Inc., Chicago, IL, USA). Differences between dairy cow archaeal communities from different diets as determined by each target gene were analysed for significance using the Wilcoxon signed ranks test, while the differences between the archaeal communities determined from different target genes for the same diet were analysed for significance using the Friedman's test.

Results

Design of 'universal' PCR primers for the archaeal thermosome

Full-length thermosome sequences ($n=166$, including α, β and γ subunits) were compiled and aligned to identify a universally amplifiable region of the thermosome gene. Degenerate 'universal' PCR primers JH0175 and JH0178 (Figure 1) were designed corresponding to positions 145–168 and 895–917 of the *Nanoarchaeum equitans* Kin4-M thermosome gene sequence, respectively. Primers were predicted to amplify a UT product of 752–803 bp, depending on the template species.

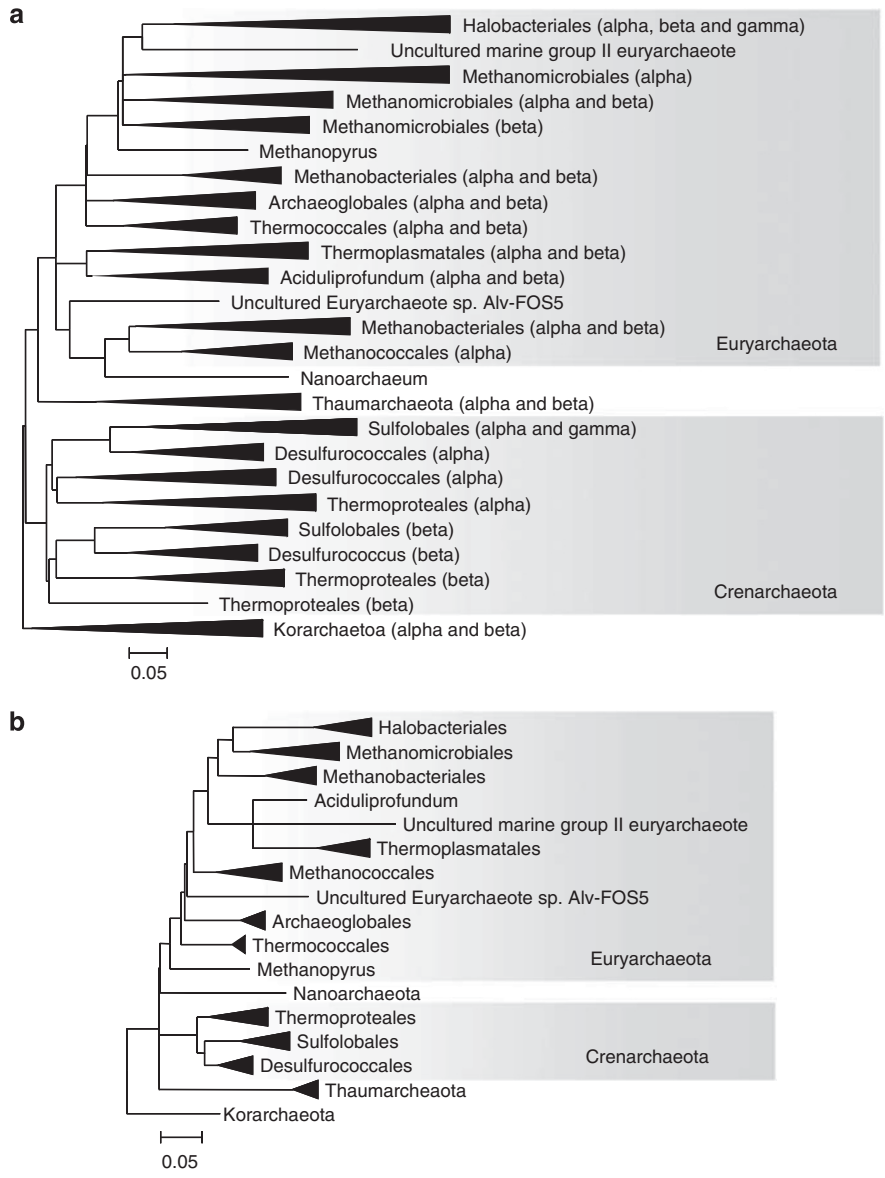

Figure 2 Phylogenetic trees of the archaeal domain based on the (**a**) UT region of thermosome gene or the (**b**) full-length 16S rRNA gene. Trees are neighbour-joined consensus trees based on 100 replicates, collapsed to the order level. Uncollapsed trees with bootstrap values are presented in Supplementary Figure 1.

Thermosome phylogeny

Thermosome sequences were trimmed to the UT region and used to construct a phylogeny (Figure 2a; Supplementary Figure 1B). The phylogeny of the thermosome UT was compared with the thermosome full-length gene phylogeny and the resulting trees were virtually identical (Supplementary Figure 1A and B). The phylogeny derived from the thermosome UT region shares the same overall tree topology when compared with the full-length 16S rRNA gene phylogeny at the order level (Figure 2b; Supplementary Figure 1C). Only the *Methanobacteriales* order separates into two distinct clusters by thermosome phylogeny, with the species *Methanothermobacter thermoautotrophicus* Delta H clustering separately from the rest of the order (visible as two Methanobacteriales (α and β) modes in Figure 2a and expanded in Supplementary Figure 1A and B). Pairwise identities for full-length 16S rRNA gene sequences (1316–1539 bp) ranged from 54% to 99% (median 73%; Supplementary Figure 2). Pairwise identities for thermosome UT sequences (705–756 bp; amplification primers were removed for analysis) ranged from 35% to 100% (median 57%; Supplementary Figure 2). The multiple subunit nature of the thermosome genes resulted in distinct clustering of α, β and γ subunits within the phylogenetic tree. This phenomenon, illustrating the recurrent paralogy of the thermosome gene, has been discussed by Archibald *et al.* (1999) and the clustering of subunits seen in this analysis is consistent with distributions found in the previous, smaller study.

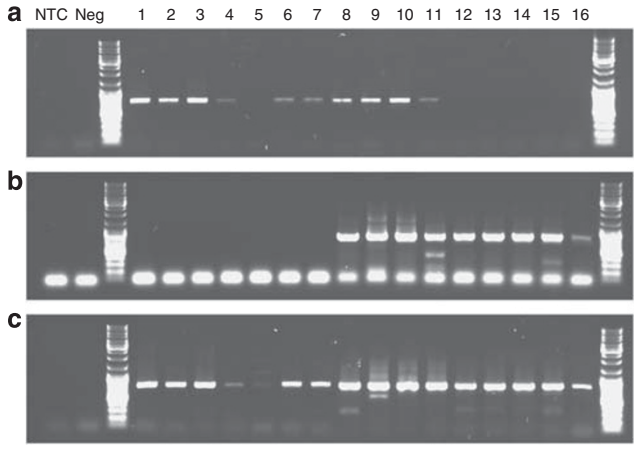

Figure 3 Thermosome PCR products from (**a**) the JH0175/JH0178 primer set, (**b**) the JH0268/JH0269 primer set and (**c**) the 7:1 cocktail of JH0175/JH0178:JH0268/JH0269 using 1.0 ng per reaction of genomic DNA from archaeal isolates as templates. The GC content of the species (or genus range, if no complete genome sequence is available for that species) is given in parentheses after the species name. Lanes are (NTC) PCR no template control; (Neg) *Escherichia coli* DH5α; (1) *Mc. voltae* (28% GC); (2) *Mc. vannielii* (31% GC); (3) *Mc. maripaludis* (33% GC); (4) *Mt. igneus* (38% GC); (5) *Ms. hungarei* (45% GC); (6) *S. solfataricus* (36% GC); (7) *Sulfolobus* sp. (33–36% GC); (8) *Tp. acidophilum* (46% GC); (9) *Tc. gorgonarius* (40–54% GC); (10) *Tc. pacificus* (40–54% GC); (11) *Tc. zilligii* (40–54% GC); (12) *Hb. salinarum* (formerly *Hb. halobium*) (66% GC); (13) *Hb. salinarum* (formerly *Hb. cutirubrum*) (66% GC); (14) *Hb. salinarum* (formerly *Hb. salinarium*) (66% GC); (15) *Hf. volcanii* WR341 (66% GC) and (16) *Hf. volcanii* WR536 (66% GC).

'Universal' PCR amplification of thermosome sequences from individual archaeal species

The JH0175/JH0178 primer set was tested on a range of purified archaeal genomic DNA extracts (Figure 3a). The predicted PCR product was amplified from most templates, with the notable exception of the halophiles (*Hb. salinarum* and *Hf. volcanii* strains), which are 66% GC (Figure 3a). To improve amplification of thermosome sequences from high GC organisms, the primers were modified to remove degenerate positions in favour of nucleotides most common in halophile sequences. When tested on the panel of genomic DNA, this high GC primer set, JH0268/JH0269, amplified thermosome sequences from the halophiles, and also members of the *Thermoplasma* and *Thermococcus* genera (Figure 3b). These results suggested that a mixture of 'universal' and 'high GC' thermosome primers would give optimal amplification across the domain. A range of primer cocktails were tested to determine what ratio of JH0175/JH0178 primers to JH0268/JH0269 primers for use in thermosome PCR most faithfully amplified a complex community (Supplementary material; Supplementary Figure 3). The 7:1 molar ratio of JH0175/JH0178:JH0268/JH0269 (350 nM per reaction of each JH0175 and JH0178 and 50 nM per reaction of each JH0268 and JH0269) gave the best overall representation (Figure 3c; Supplementary Figure 3).

Amplification of thermosome, 16S rRNA gene and mcrA *gene sequences from rumen contents*

To evaluate the performance of the thermosome PCR primers on complex samples, rumen contents from dairy cows on two different diets were obtained, total DNA was isolated and the archaeal communities present were determined using thermosome, 16S rRNA gene and *mcrA* 'universal' primers. Although only the thermosome and 16S rRNA primer sets are designed to be domain-wide, the archaeal community of the rumen is known to be methanogen dominated (Shin *et al.*, 2004; Wright *et al.*, 2007; Zhou *et al.*, 2009), making the methanogen-specific, protein-encoding *mcrA* gene a worthwhile target for comparison. In addition, the *mcrA* gene is a common target for archaeal rumen research (Tatsuoka *et al.*, 2004; Denman *et al.*, 2007). Clone libraries were made for each gene target from each of the diets, resulting in six libraries. A regular diet clone library was also generated using the 7:1 thermosome primer cocktail. Clones were selected randomly from each library and sequenced. Thermosome PCR amplicons (JH0175/JH0178 only) from the dry food diet were also subjected to pyrosequencing. Sequences obtained were processed for quality and the nearest neighbour was determined for each sequence.

Table 2 reports the compositions of the various libraries by nearest neighbour. Regardless of the target gene used, *Methanobrevibacter smithii*, *Methanobrevibacter ruminantium* and *Ms. stadtmanae* were the dominant species detected. Within the clone library data (based on the six libraries where the same primers were used for both diets), there were no statistically significant differences between the archaeal populations detected in the two diets examined, regardless of the gene targeted (thermosome, $P=0.593$; 16S rRNA, $P=0.715$; *mcrA*, $P=0.610$). As well, the target used had no significant effect on the archaeal population detected within a diet (regular diet, $P=0.513$; dry diet, $P=0.405$).

Phylogenetic trees of clone library sequences from JH0175/JH0178 only, thermosome 7:1 primer cocktail, 16S rRNA gene and *mcrA* are shown in Supplementary Figure 4A–D, respectively. Both α and β subunit sequences were detected and are clearly distinguishable for *Mb. smithii* and *Mb. ruminantium* in the thermosome clone libraries (Supplementary Figure 4A and B).

When the thermosome PCR products from the dry diet were sequenced to greater depth by 454 pyrosequencing, a comparable archaeal community profile was observed (Table 2). As expected, the increase in community sampling revealed taxa not seen in the clone libraries. Interestingly, among the additional sequences, three non-identical thermosome sequences were detected that appear to be non-methanogen in origin (closest neighbour sequence was Desulfurococcales species *Staphylothermus hellenicus*, 67.3%, 63.2% and 63.2% identical; Table 2).

Table 2 Species identified as a nearest neighbour for sequences detected in rumen samples by diet and target gene

Nearest neighbour	*Regular diet*				*Dry diet*				*Total*
	tcp 7:1[a] *lib*	*tcp lib*[a]	*16S lib*[b]	*mcrA lib*[c]	*tcp lib*	*16S lib*	*mcrA lib*	*tcp pyro*[d]	
Methanobrevibacter smithii	46[e]	24	37	21	39	52	26	276	521
Methanobrevibacter ruminantium	55	11	28	45	12	8	23	146	328
Methanosphaera stadtmanae	9	17	1	8	8	4	16	88	151
Methanosphaerula palustris	0	0	0	0	0	6	0	4	9
Methanocorpusculum labreanum	0	0	0	0	0	0	12	1	13
Methanothermobacter thermoautotrophicus	0	0	0	0	0	0	0	3	3
Methanococcus maripaludis	0	0	0	0	0	0	1	0	1
Methanoculleus marisnigri	0	0	0	1	0	0	0	0	1
Methanococcoides burtonii	0	0	0	1	0	0	0	0	1
Methanosarcina acetivorans	0	0	0	0	0	0	0	1	1
Staphylothermus hellenicus	0	0	0	0	0	0	0	3	3
Total sequences	110	52	66	76	59	69	78	522	1034

[a]Thermosome gene target generated from either a cocktail of 7:1 JH0175/JH0178:JH0268/JH0269 primers (tcp 7:1) or JH0175/JH0178 primers alone (tcp), clone library results.
[b]16S rRNA gene target, clone library results.
[c]Methyl co-enzyme M reductase gene target, clone library results.
[d]Thermosome gene target (JH0175/JH0178 primers alone), pyrosequencing results.
[e]Number of sequence reads.

Comparison of intraspecies resolution of the three gene targets

Phylogenetic trees of *Mb. smithii*-like sequences from all three gene targets were constructed to evaluate the intraspecies resolution of each target (Figure 4). The 16S rRNA gene sequence target (trimmed to 617–620 bp common to all reads) generated 16 unique sequences from 81 sequence reads (Figure 4a). Pairwise sequence identities ranged from 93% to 99% (median 98%; Supplementary Figure 5). Thirty-three unique *Mb. smithii mcrA* sequences (436–442 bp) were detected, representing 32 unique subunit 1 types (from 46 reads) and 1 subunit 2 type (from a single read) (Figure 4b). Pairwise identities within *mcrA* α subunits were 84–99% (median 95%), while the single β subunit was 78% identical to the *Mb. smithii mcrA* type 2 subunit. Between subunits, type 1 and type 2 shared an average of 67% identity, with a range of 65–70% identity (Supplementary Figure 5). The longest branch lengths were observed in the thermosome tree (based on alignment of 707–711 bp), with 23 unique α subunit sequences (from 43 reads) and 14 unique β subunit sequences (from 16 reads) (Figure 4c). Pairwise DNA sequence identities were 75–99% (median 82%) for α subunits, 82–99% (median 85%) for β subunits and 59–67% (median 62%) between α and β subunits (Supplementary Figure 5).

Discussion

The objective of this study was to develop a 'universal' archaeal PCR protocol, based on the conserved archaeal thermosome (type II chaperonin) gene, for application in archaeal species identification, metagenomic and phylogenetic studies. The strength of the thermosome gene as a target is that it combines universality (like the 16S rRNA gene) with the greater sequence diversity associated with a protein-encoding gene (like the *mcrA* gene). The benefits of targeting a protein-encoding gene have been seen in bacteria, where the *cpn*60 (type I chaperonin) UT has advantages over 16S rRNA gene in terms of discrimination of closely related taxa and bioinformatics (Schellenberg *et al.*, 2009).

PCR primers were designed with the intent to amplify at least one thermosome subunit sequence from each known archaeal species. However, given the conservation across α, β and γ subunits at the primer annealing sites, the thermosome primers presented here are expected to amplify most or all subunits from most of the species examined. Experimentation revealed that while the JH0175/JH0178 primer pair could amplify thermosome sequence from all Archaea tested (Supplementary Figure 3), amplification of high GC sequences was relatively inefficient (Figure 3a). Difficulty with PCR amplification of high GC content templates has been reported with both specific primers (Varadaraj and Skinner, 1994) and degenerate primers containing inosine (Hill *et al.*, 2006b). This problem was encountered in the *cpn*60 UT PCR protocol and was overcome by the application of a 'cocktail' of inosine containing primers with primers specifically designed to favour high GC sequences (Hill *et al.*, 2006b). We applied a similar approach and designed primers JH0268 and JH0269 to enhance amplification of high GC organisms (Figure 3b). A primer cocktail containing a 7:1 molar ratio of primers JH0175/JH0178 and JH0268/JH0269 resulted in balanced amplification of the complete spectrum of organisms tested (Figure 3c; Supplementary Figure 3).

The thermosome UT region proved to be a useful phylogenetic target. With approximately half the

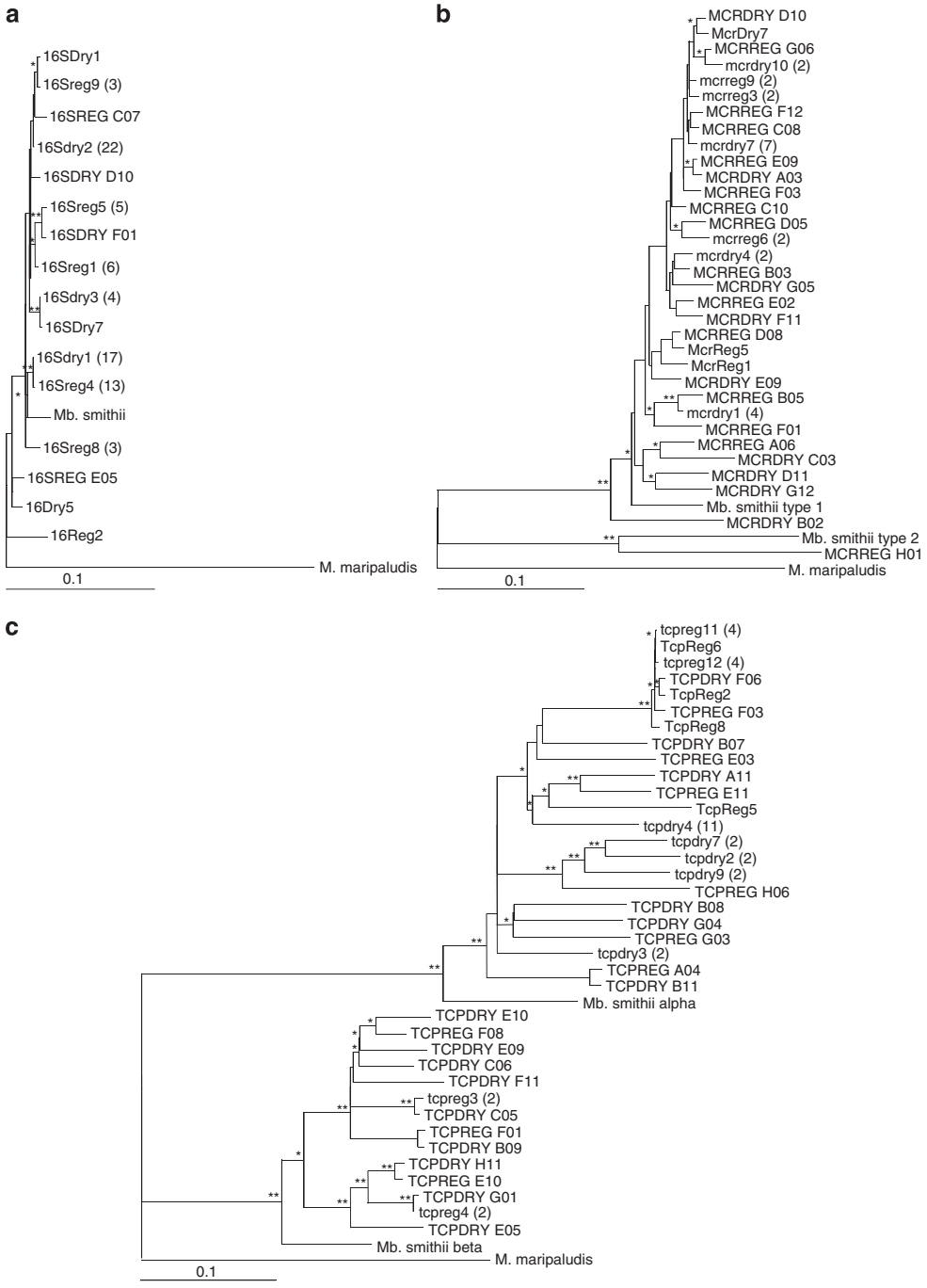

Figure 4 Phylogenetic trees of DNA sequences identified as *Mb. smithii* from rumen clone libraries based on the (**a**) 16S rRNA gene, (**b**) *mcrA* gene or (**c**) thermosome gene ('universal' JH0175/JH0178 amplification only). Trees are neighbour-joined consensus trees based on 100 replicates. Nodes with bootstrap values greater than 50 (*) or 90 (**) are indicated. *Mb. smithii* reference sequences were taken from strain DSM 2374 and *Mc. maripaludis* C5 sequences were used as outliers. Numbers in parentheses indicate how many times a sequence was detected in the library when present more than once.

sequence length of the full 16S rRNA gene, the thermosome-based phylogenetic tree featured longer branch lengths and clearer separation of most species compared with the 16S rRNA gene-based tree (Figure 2; Supplementary Figure 1). For example, the three C strains of *Mc. maripaludis* (C5, C6 and C7) that were virtually identical by 16S rRNA gene sequence (99%) are distinguishable by the shorter thermosome sequence (96–97%; Supplementary Figure 1). The only instance where there was a difference distinguishable by 16S rRNA gene and not thermosome UT was between the two strains of *S. solfataricus* (98/2 and P2), where they had 99% identity over the 1497-bp of 16S rRNA gene sequence and 100% identity within each class of thermosome subunit over 726 bp.

When applied to the complex microbial community of the rumen, the thermosome PCR detected the same major species as the 16S rRNA gene and *mcrA* PCR protocols: *Mb. smithii*, *Mb. ruminantium* and *Ms. stadtmanae* (Table 2). The species of methanogens detected in this study were expected, with many of them previously reported from ruminants (Tatsuoka *et al.*, 2004; Wright *et al.*, 2004, 2007; Denman *et al.*, 2007). Interestingly, the pyrosequencing library contained three apparently non-methanogen thermosome sequences (Table 2). These sequences were too short for inclusion in the phylogenetic analysis; however, their observation is consistent with reports that non-methanogenic Archaea exist as part of the rumen community (Shin *et al.*, 2004; Wright *et al.*, 2007). Examination of the sequences classified as *Mb. smithii* by each target gene illustrated differing degrees of strain variation within this species (Figure 4). There was an average of 98% pairwise identity between *Mb. smithii*-like 16S rRNA gene sequences. Conversely, thermosome α and β subunits shared only 82% and 85% pairwise identity within each of the α and β subunits, respectively, with further subgroupings within subunits apparent in the phylogenetic tree (Figure 4c). Whether these subgroups represent ecotypes, as has been shown for *cpn*60 ecotypes (Vermette *et al.*, 2010), is an area for future study.

Our results demonstrate that (1) the thermosome gene and portions thereof can be taxonomically informative and are more diverse than 16S rRNA gene sequences; (2) a 'universal' PCR protocol could be designed to target the thermosome gene; (3) thermosome 'universal' PCR performed comparably to established 16S rRNA gene and *mcrA* gene 'universal' PCR protocols for the detection of archaeal members within rumen communities, while yielding greater sequence diversity. The advantage of the thermosome system is in its combined broad domain specificity and its superior discriminating power. The relatively homogenous length of the thermosome UT simplified bioinformatic analyses and the availability of a curated collection of thermosome sequences (in cpnDB) facilitates the application of thermosome sequences to metagenomic and phylogenetic studies. The occurrence of multiple subunits results in more complex profiles from community samples than those obtained with 16S rRNA gene or *mcr*A, which is potentially an advantage since these more complex profiles may make for more robust diagnostics. Currently, the ability to identify thermosome paralogues in communities containing poorly characterized, uncultured Archaea is limited by the lack of complete genome sequences for reference. However, given the rapid pace of accumulation of genomic and metagenomic sequence data from environmental Archaea, this limitation will likely be short lived. Meanwhile, the method described here can be implemented for the identification and phylogenetic placement of archaeal isolates, the development of species-specific detection and quantification methods and the characterization of archaeal communities.

Acknowledgements

We thank Ken Jarrell (Queen's University, Canada) and Jerry Eichler (Ben-Gurion University of the Negev, Israel) for generously providing archaeal DNA and Steve Hendrick for collecting rumen samples. We are grateful to Champika Fernando for statistic assistance and Matthew Links for advice and support in pyrosequencing data processing. BC was supported by a Saskatchewan Health Research Foundation postdoctoral fellowship.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Arahal DR, Sanchez E, Macian MC, Garay E. (2008). Value of recN sequences for species identification and as a phylogenetic marker within the family 'Leuconostocaceae'. *Int Microbiol* **11**: 33–39.

Archibald JM, Logsdon JM, Doolittle WF. (1999). Recurrent paralogy in the evolution of archaeal chaperonins. *Curr Biol* **9**: 1053–1056.

Baker GC, Smith JJ, Cowan DA. (2003). Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.

Chaban B, Ng SY, Jarrell KF. (2006). Archaeal habitats—from the extreme to the ordinary. *Can J Microbiol* **52**: 73–116.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Denman SE, Tomkins NW, McSweeney CS. (2007). Quantitation and diversity analysis of ruminal methanogenic populations in response to the antimethanogenic compound bromochloromethane. *FEMS Microbiol Ecol* **62**: 313–322.

DeSantis Jr TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM *et al.* (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**: W394–W399.

Dumonceaux TJ, Hill JE, Briggs SA, Amoako KK, Hemmingsen SM, Van Kessel AG. (2006a). Enumeration of specific bacterial populations in complex intestinal communities using quantitative PCR based on the *chaperonin*-60 target. *J Microbiol Methods* **64**: 46–62.

Dumonceaux TJ, Hill JE, Hemmingsen SM, Van Kessel AG. (2006b). Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. *Appl Environ Microbiol* **72**: 2815–2823.

Felsenstein J. (1989). PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**: 164–166.

Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA *et al.* (1980). The phylogeny of prokaryotes. *Science* **209**: 457–463.

Glazunova OO, Raoult D, Roux V. (2009). Partial sequence comparison of the rpoB, sodA, groEL and gyrB genes

within the genus Streptococcus. *Int J Syst Evol Microbiol* **59**: 2317–2322.

Hill JE, Hemmingsen SM, Goldade BG, Dumonceaux TJ, Klassen J, Zijlstra RT *et al.* (2005). Comparison of ileum microflora of pigs fed corn-, wheat-, or barley-based diets by *chaperonin*-60 sequencing and quantitative PCR. *Appl Environ Microbiol* **71**: 867–875.

Hill JE, Paccagnella A, Law K, Melito PL, Woodward DL, Price L *et al.* (2006a). Identification of *Campylobacter* spp. and discrimination from *Helicobacter* and *Arcobacter* spp. by direct sequencing of PCR-amplified cpn60 sequences and comparison to cpnDB, a chaperonin reference sequence database. *J Med Microbiol* **55**: 393–399.

Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM. (2004). cpnDB: a chaperonin sequence database. *Genome Res* **14**: 1669–1675.

Hill JE, Town JR, Hemmingsen SM. (2006b). Improved template representation in cpn60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environ Microbiol* **8**: 741–746.

Kaeberlein T, Lewis K, Epstein SS. (2002). Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.

Kubota H, Hynes G, Willison K. (1995). The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur J Biochem* **230**: 3–16.

Large AT, Goldberg MD, Lund PA. (2009). Chaperones and protein folding in the archaea. *Biochem Soc Trans* **37**: 46–51.

Large AT, Lund PA. (2009). Archaeal chaperonins. *Front Biosci* **14**: 1304–1324.

Luton PE, Wayne JM, Sharp RJ, Riley PW. (2002). The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* **148**: 3521–3530.

Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

Meintanis C, Chalkou KI, Kormas KA, Lymperopoulou DS, Katsifas EA, Hatzinikolaou DG *et al.* (2008). Application of rpoB sequence similarity analysis, REP-PCR and BOX-PCR for the differentiation of species within the genus *Geobacillus*. *Lett Appl Microbiol* **46**: 395–401.

Mihajlovski A, Alric M, Brugere JF. (2008). A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the *mcrA* gene. *Res Microbiol* **159**: 516–521.

Nawrocki EP, Kolbe DL, Eddy SR. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**: 1335–1337.

Nicholas KB, Nicholas HBJ, Deerfield DWI. (1997). GeneDoc: analysis and visualization of genetic variation. *EMBNEW News* **4**: 14.

Page RD. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357–358.

Santos SR, Ochman H. (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* **6**: 754–759.

Schellenberg J, Links MG, Hill JE, Dumonceaux TJ, Peters GA, Tyler S *et al.* (2009). Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol* **75**: 2889–2898.

Shin EC, Choi BR, Lim WJ, Hong SY, An CL, Cho KM *et al.* (2004). Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16S rDNA sequence. *Anaerobe* **10**: 313–319.

Tatsuoka N, Mohammed N, Mitsumori M, Hara K, Kurihara M, Itabashi H. (2004). Phylogenetic analysis of methyl coenzyme-M reductase detected from the bovine rumen. *Lett Appl Microbiol* **39**: 257–260.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.

Trent JD, Nimmesgern E, Wall JS, Hartl FU, Horwich AL. (1991). A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**: 490–493.

Varadaraj K, Skinner DM. (1994). Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene* **140**: 1–5.

Vermette CJ, Russell AH, Desai AR, Hill JE. (2010). Resolution of phenotypically distinct strains of *Enterococcus* spp. in a complex microbial community using cpn60 universal target sequencing. *Microb Ecol* **59**: 14–24.

Wang LT, Lee FL, Tai CJ, Kasai H. (2007). Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the *Bacillus subtilis* group. *Int J Syst Evol Microbiol* **57**: 1846–1850.

Weng FY, Chiou CS, Lin PH, Yang SS. (2009). Application of recA and rpoB sequence analysis on phylogeny and molecular identification of *Geobacillus* species. *J Appl Microbiol* **107**: 452–464.

Woese CR, Fox GE. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**: 5088–5090.

Woese CR, Stackebrandt E, Macke TJ, Fox GE. (1985). A phylogenetic definition of the major eubacterial taxa. *Syst Appl Microbiol* **6**: 143–151.

Wright AD, Auckland CH, Lynn DH. (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Appl Environ Microbiol* **73**: 4206–4210.

Wright AD, Williams AJ, Winder B, Christophersen CT, Rodgers SL, Smith KD. (2004). Molecular diversity of rumen methanogens from sheep in Western Australia. *Appl Environ Microbiol* **70**: 1263–1270.

Yap WH, Zhang Z, Wang Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol* **181**: 5201–5209.

Zeigler DR. (2005). Application of a recN sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus*. *Int J Syst Evol Microbiol* **55**: 1171–1179.

Zhou M, Hernandez-Sanabria E, Guan LL. (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75**: 6524–6533.

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