

Molecular signatures for the phylum Synergistetes and some of its subclades

Vaibhav Bhandari · Radhey S. Gupta

Received: 4 April 2012 / Accepted: 30 May 2012 / Published online: 19 June 2012
© Springer Science+Business Media B.V. 2012

Abstract Species belonging to the phylum Synergistetes are poorly characterized. Though the known species display Gram-negative characteristics and the ability to ferment amino acids, no single characteristic is known which can define this group. For eight Synergistetes species, complete genome sequences or draft genomes have become available. We have used these genomes to construct detailed phylogenetic trees for the Synergistetes species and carried out comprehensive analysis to identify molecular markers consisting of conserved signature indels (CSIs) in protein sequences that are specific for either all Synergistetes or some of their sub-groups. We report here identification of 32 CSIs in widely distributed proteins such as RpoB, RpoC, UvrD, GyrA, PolA, PolC, MraW, NadD, PyrE, RpsA, RpsH, FtsA, RadA, etc., including a large >300 aa insert within the RpoC protein, that are present in various Synergistetes species, but except for isolated bacteria, these CSIs are not found in the protein homologues from any other organisms. These

CSIs provide novel molecular markers that distinguish the species of the phylum Synergistetes from all other bacteria. The large numbers of other CSIs discovered in this work provide valuable information that supports and consolidates evolutionary relationships amongst the sequenced Synergistetes species. Of these CSIs, seven are specifically present in *Jonquetella*, *Pyramidobacter* and *Dethiosulfovibrio* species indicating a cladal relationship among them, which is also strongly supported by phylogenetic trees. A further 15 CSIs that are only present in *Jonquetella* and *Pyramidobacter* indicate a close association between these two species. Additionally, a previously described phylogenetic relationship between the *Aminomonas* and *Thermanaerovibrio* species was also supported by 9 CSIs. The strong relationships indicated by the indel analysis provide incentives for the grouping of species from these clades into higher taxonomic groups such as families or orders. The identified molecular markers, due to their specificity for Synergistetes and presence in highly conserved regions of important proteins suggest novel targets for evolutionary, genetic and biochemical studies on these bacteria as well as for the identification of additional species belonging to this phylum in different environments.

Electronic supplementary material The online version of this article (doi:10.1007/s10482-012-9759-2) contains supplementary material, which is available to authorized users.

V. Bhandari · R. S. Gupta (✉)
Department of Biochemistry and Biomedical Sciences,
McMaster University, Hamilton, ON L8N 3Z5, Canada
e-mail: gupta@mcmaster.ca

Keywords Conserved indels · Molecular signatures · Synergistetes phylogeny and taxonomy · *Jonquetella*-*Pyramidobacter*-*Dethiosulfovibrio* clade · *Aminomonas*-*Thermanaerovibrio* clade · Lateral gene transfers

Introduction

The Synergistetes group of bacteria is a recently recognized phylum for which 40 organisms have been isolated and over three hundred 16S rRNA sequences are available (Hugenholtz et al. 2009; NCBI Taxonomy 2012). The phenotypic characteristics shared by the species from this phylum include their gram-negative cell wall, anaerobic existence, and rod/vibrioid cell shape (Jumas-Bilak et al. 2009). Although the presence of lipopolysaccharides, which is an important characteristic of diderm cell envelopes, in Synergistetes species has not yet been reported, they do contain genes for various proteins that are involved in lipopolysaccharide biosynthesis (Sutcliffe 2010). While a few species have been shown to be asaccharolytic, all Synergistetes have the ability to ferment amino acids (Magot et al. 1997; Baena et al. 1998, 1999a; Surkov et al. 2001; Hongoh et al. 2007; Downes et al. 2009; Jumas-Bilak et al. 2009). The Synergistetes inhabit primarily anaerobic environments including animal gastrointestinal tracts, soil, oil wells and wastewater treatment plants and they are also present in sites of human diseases such as cysts, abscesses and areas of periodontal disease (Godon et al. 2005; Kumar et al. 2005; de Lillo et al. 2006; Horz et al. 2006; Jumas-Bilak et al. 2007; Vartoukian et al. 2007; Zijngje et al. 2010). Due to their presence at illness related sites, the Synergistetes are suggested to be opportunistic pathogens but they can also be found in healthy individuals in the microbiome of the umbilicus and in normal vaginal flora (Vartoukian et al. 2007; Marchandin et al. 2010). Other species from this phylum have been identified as significant contributors in the degradation of sludge for production of biogas in anaerobic digesters and are potential candidates for use in renewable energy production through their production of hydrogen gas (McSweeney et al. 1993; Maune and Tanner 2012; Delbes et al. 2001; Riviere et al. 2009; Ziganshin et al. 2011).

Synergistetes species were first identified with the isolation of *Synergistes jonesii* from which the phylum name “Synergistetes” is derived. *S. jonesii* was isolated from the rumen of a goat in 1992 and described as a gram-negative staining, anaerobic, rod-shaped, commensal bacteria with the ability to degrade the toxic compound pyridinediol 3-hydroxy-4-1(*H*)-pyridone (Allison et al. 1992; McSweeney et al. 1993). *S. jonesii* 16S rRNA was not closely related to that of any bacteria

characterized at the time but the species was later misclassified as a member of the Deferrribacteres group (Garrity et al. 2004). Around the same time, the species *Thermoanaerovibrio acidaminovorans* was isolated from methanogenic sludge and indicated to be a member of the *Selenomonas* genus within the Firmicutes phylum (Guangsheng et al. 1992; Baena et al. 1999b). These misclassifications of Synergistetes continued, as species from the genera *Aminobacterium* and *Dethiosulfovibrio* were described as forming a deep-branching clade beside cluster V within Clostridia, consisting of a group composed of *Thermoanaerobacter* species (Magot et al. 1997; Baena et al. 1998). Several other organisms now considered Synergistetes were also placed among the Syntrophomonadaceae family of the Firmicutes (Garrity et al. 2004; Dahle and Birkeland 2006; Diaz et al. 2007). Eventually, efforts based on 16S rRNA sequences by Jumas-Bilak et al. (2009), identified the monophyletic nature of the Synergistetes within the bacterial domain and proposed that these “*Synergistia jonesii*-like” species form a distinct phylum, now named the Synergistetes (Jumas-Bilak et al. 2009). All characterized Synergistetes species are currently placed under the class Synergistia, the order Synergistiales and the family Synergistaceae (Jumas-Bilak et al. 2009). This family until recently was comprised of 11 genera, namely: *Aminiphilus*, *Aminobacterium*, *Aminomonas*, *Anaerobaculum*, *Cloacibacillus*, *Dethiosulfovibrio*, *Jonquetella*, *Pyramidobacter*, *Synergistes*, *Thermoanaerovibrio* and *Thermovirga* (Hugenholtz et al. 2009; Jumas-Bilak et al. 2009; NCBI Taxonomy 2012). Recently, a new genus *Fretibacterium* has also been described, which contains a single species *Fretibacterium fastidiosum* that was previously known as *Synergistes bacterium* SGP1. A candidate genus *Tammella*, composed of a group of related and uncultured species found within termite guts, has also been suggested to belong to the phylum Synergistetes (Hongoh et al. 2007; Hugenholtz et al. 2009).

While the Synergistetes are currently classified as belonging to a separate phylum based on their 16S rRNA sequences, no characteristic of these bacteria is known that can easily differentiate a Synergistetes species from other bacteria. Though all cultured Synergistetes can ferment amino acids, various species from other taxa also share this ability (Hou et al. 2004; Fonknechten et al. 2010). The availability of genome sequences has allowed for the employment of comparative genome approaches for the identification

of molecular markers that are specific for different bacterial groups at various taxonomic levels (Gupta 1998; Griffiths et al. 2005; Gupta and Bhandari 2011; Gupta and Shami 2011). Using genomic sequences, our lab has pioneered the discovery of conserved signature insertions/deletions (i.e. indels, CSIs) present in protein sequences that are specific for particular groups of organisms (Gupta 1998, 2009; Gao and Gupta 2005; Griffiths and Gupta 2006; Gupta and Bhandari 2011). The group specific presence of CSIs can be parsimoniously explained through rare genetic changes occurring in a common ancestor to the particular groups of species and then being passed down through vertical descent (Gupta 1998, 2000, 2009). Such CSIs, which are present in a related group of species and absent in other organisms, are useful as molecular markers for the identification of species belonging to a taxonomic group and the demarcation of the group's boundaries. Additionally, through comparison of sequences and based on the presence or absence of the indicated CSIs in outgroup species, a rooted phylogenetic relationship can be inferred among the species (Rivera and Lake 1992; Baldauf and Palmer 1993; Gupta 1998, 2001).

From the species identified as Synergistetes, complete or annotated draft genomes are now available for nine species (described below). In the present work, we have carried out detailed comparative analyses on protein sequences from these genomes to identify molecular markers (CSIs) that are specific for the phylum Synergistetes and some of its subgroups, as well as those that provide information regarding its relationship to other bacterial phyla. Our work has identified numerous CSIs that provide highly specific markers for all sequenced members of the Synergistetes phylum as well as a number of its sub-groups. Additionally, several CSIs that are commonly shared by Synergistetes and some species from other bacterial phyla suggest potential cases of lateral gene transfers. These CSIs provide novel and powerful means for the identification/circumscription of species from the phylum Synergistetes and for different types of studies on them.

Phylogenetic analysis of the genome sequenced Synergistetes

The complete genomes for *Aminobacterium (Amb.) colombiense* (Chertkov et al. 2010), *T. acidaminovorans*

(Chovatia et al. 2009) and *Thermovirga (Tv.) lienii* (Dahle and Birkeland 2006) have been published while annotated draft genomes were accessible for *Dethiosulfovibrio peptidovorans* (Labutti et al. 2010), *Aminomonas (Amm.) paucivorans* (Pitluck et al. 2010), *Anaerobaculum (An.) hydrogeniformans*, *Jonquetella anthropi* and *Pyramidobacter piscolens* (NCBI genomic database 2012). Limited sequence data for *F. fastidiosum*, which is currently referred to as *Synergistetes bacterium SGP1* in the NCBI database, was also available (Vartoukian et al. 2012). These species represent nine of twelve characterized genera from the phylum. Some characteristics of these organisms and their genomes are provided in Table 1.

The relationships of the species in the Synergistetes phylum have thus far been primarily analyzed through 16S rRNA sequence data. However, it is now recognized that trees based on a larger dataset of genes or proteins representing diverse functional categories are more reliable in resolving phylogenetic relationships than a single gene such as the 16S rRNA or a single protein (Rokas et al. 2003; Ciccarelli et al. 2006; Wu and Eisen 2008). Therefore, in order to visualize the relationship among the sequenced Synergistetes species, phylogenetic trees based upon concatenated sequences of ten housekeeping proteins were constructed. The 10 proteins that were used for phylogenetic analysis (viz. ArgRS, GyrB, Hsp70, ribosomal proteins L1 and L5, RpoB, RpoC, TrxB, UvrD and ValRS) are found in most bacteria and they have been extensively used for other phylogenetic studies (Bocchetta et al. 2000; Ciccarelli et al. 2006; Soria-Carrasco et al. 2007; Zhaxybayeva et al. 2009; Naushad and Gupta 2012). In addition to the Synergistetes species, the dataset that was employed for phylogenetic analyses also contained information for species from several other bacterial phyla including those in whose proximity the Synergistetes species were observed to branch in earlier studies (Guangsheng et al. 1992; Magot et al. 1997; Baena et al. 1998; Garrity et al. 2004; Diaz et al. 2007; Herlemann et al. 2009). The results for the multi-protein concatenated phylogenetic analysis are presented in Fig. 1a. In parallel, a 16S rRNA tree was also created to investigate the congruence with the protein tree (Fig. 1b).

In both the protein tree and the 16S rRNA tree, the Synergistetes species formed a monophyletic clade that was distinct from all other bacterial groups,

Table 1 Characteristics of the Synergistetes species with sequenced genomes

Synergistetes species	GC (%)	Isolation source	Asaccharolytic	Optimum temperature (°C)	Size (Mb)	Reference
<i>Aminobacterium colombiense</i>	45.3	Anaerobic lagoon of dairy wastewater treatment plant	Yes	37	1.98	DOE-JGI ^a
<i>Aminomonas paucivorans</i>	43	Anaerobic lagoon of dairy wastewater treatment plant	Yes	35	2.6	DOE-JGI ^a
<i>Anaerobaculum hydrogeniformans</i>	46.6	Oil-well production water	No	55	2.3	GSC-WashU ^b
<i>Dethiosulfovibrio peptidovorans</i>	54.4	Oil well	Yes	42	2.6	DOE-JGI ^a
<i>Jonquetella anthropi</i>	59.4	Peritoneal fluid, Breast and pelvic abscess, sebaceous cyst and wounds	Yes	37	1.7	DOE-JGI ^a
<i>Pyramidobacter piscocolens</i>	59	Human oral cavity	Yes	37	2.6	JCV ^c
<i>Thermanaerovibrio acidaminovorans</i>	63.8	Granular methanogenic sludge	No	55	1.85	DOE-JGI ^a
<i>Thermovirga lienii</i>	46.6	Oil-well production water	Yes	58	2.06	DOE-JGI ^a
<i>Fretibacterium fastidiosum</i>	63	Subgingival plaque	Yes	37	Unknown	Vartoukian et al. (2010)

^a DOE-JGI—these genomes have been sequenced by the United States Department of Energy Joint Genomic Institute

^b GSC-WashU—Genome Sequencing Center (GSC) at Washington University (WashU) School of Medicine

^c JCV—J. Craig Venter Institute

supporting their assignment into a separate phylum. The species such as *Syntrophomonas wolfei* or *Selemononas sputigena* in whose proximity some of the Synergistetes species were indicated to branch in earlier studies, branched distinctly from them. Although the relationships among other bacterial species/phyla differed within the two trees and they were mostly unresolved, within the Synergistetes clade both the concatenated protein tree and the rRNA tree displayed a similar branching order. In both trees, the Synergistetes species showed a split into two clades at the highest level. One clade is comprised of *Thermanaerovibrio* and *Aminomonas* sharing a distant relationship with the *Thermovirga* and *Anaerobaculum* species while the other clade is comprised of the five species from the genera *Aminobacterium*, *Jonquetella*, *Pyramidobacter*, *Dethiosulfovibrio* and *Fretibacterium*. The concatenated protein tree shows, with high statistical support, that the *J. anthropi* and *P. piscocolens* species branch together and that *D. peptidovorans* is the closest relative of these two species. The trees also show a well-supported, close relationship between *T. acidaminovorans* and *Amm. paucivorans*. The species *Amb. colombiense* and *F.*

fastidiosum are observed to robustly branch together, though their relationship to the *Dethiosulfovibrio*–*Pyramidobacter*–*Jonquetella* clade was strongly supported only by the NJ concatenated protein tree. The ML analysis and the rRNA tree weakly supported this relationship. The position of *An. hydrogeniformans* and *Tv. lienii* species within the phylum was poorly resolved in both the concatenated protein and the rRNA trees. In both the trees, short branches connect these species to the *Thermoanaerovibrio*–*Aminomonas* clade and their grouping in this clade was weakly supported by the ML tree and the rRNA tree.

Identification of CSIs that are specific for the Synergistetes species

For the identification of CSIs, BlastP searches against the non-redundant protein sequence (nr) database were carried out on all proteins from the genome of the species *Amb. colombiense* DSM 12261 and *T. acidaminovorans* DSM 6589 from the Synergistetes phylum (Altschul et al. 1997, 2005). Using the ClustalX program, multiple sequence alignments were

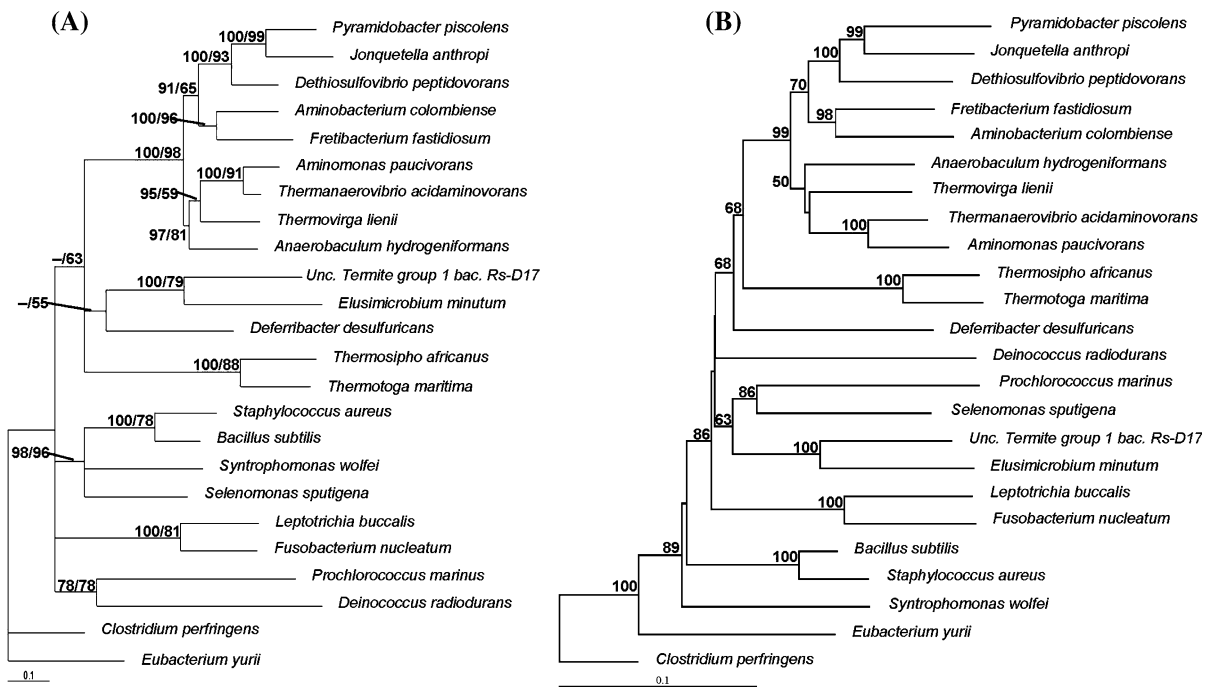


Fig. 1 Phylogenetic tree for sequenced Synergistetes species and representative species from some closely related bacterial phyla. **a** A neighbour joining (NJ) distance tree based upon concatenated sequences for 10 highly conserved and widely distributed proteins (ArgRS, GyrB, Hsp70, ribosomal proteins L1 and L5, RpoB, RpoC, TrxB, UvrD and ValRS). The numbers on the node indicate the % bootstrap score (or statistical support)

for each node in the NJ and maximum-likelihood analyses, respectively. The *dashes* (–) at nodes indicate that the statistical support for this particular branching relationship was <50 % in the NJ or ML analysis. **b** A NJ tree for the same species as shown in (a) based upon 16S rRNA sequences. The trees were constructed as described in our earlier work (Gupta and Bhandari 2011)

created for all proteins for which high scoring homologues were available from most Synergistetes species as well as several other groups of organisms. The aligned proteins were visually inspected to identify insertions or deletions that were flanked by conserved amino acids on both sides. Insertions and deletions that were not flanked by at least 4–5 conserved residues within the neighbouring 30–40 residues were not further considered as they do not provide useful molecular markers (Gupta 1998, 2001; Gupta and Bhandari 2011). More detailed BlastP searches (searching for 250 of the closest sequence matches) were then carried out on 50–80 aa long segments (longer in some cases) containing the indels and its flanking conserved regions to determine the species distribution for the identified indels. Indels predominantly found in Synergistetes species or those that were found in Synergistetes along with some other taxonomic group of organisms were retained and compiled into signature files. The signature files shown here contain sequence alignments of various

detected indels along with the flanking conserved regions for all Synergistetes and representative species from other taxonomic groups for which information was detected in the Blast searches. However, due to spatial considerations sequence information for only limited numbers of species from other groups are shown here. In a few cases, where more than one homologue of a protein was detected for the same species, sequence information for different homologues was included only if they showed differing characteristics (viz. one homologue contained the indel while the other(s) did not). All of the indels reported in this work are independent of each other and they are not part of indels for other larger clades.

CSIs that are specific for the Synergistetes phylum

CSIs in proteins brought about by rare genomic changes that are restricted to phylogenetically well-

defined groups are useful as molecular markers that provide means for evaluating evolutionary relationships (Gupta 1998; Rokas and Holland 2000). Our analyses of genome sequences from Synergistetes species have identified 32 CSIs that help define and demarcate the species of this phylum. Some characteristics for these Synergistetes specific CSIs are listed in Table 2 and two examples are provided in Fig. 2. Figure 2a depicts two inserts that are present in close proximity within the β' subunit of the RNA polymerase enzyme (RpoC), an essential enzyme responsible for transcription of genes in all organisms. The region of the protein shown is highly conserved among all organisms and it contains a 2 aa insert that is specifically present in all homologues of the RpoC enzyme from species of the phylum Synergistetes. Another example of a conserved indel that is specific for all Synergistetes species is shown in Fig. 2b. In this case, the α -subunit of DNA polymerase III contains a 1 aa insert that is specifically present in all Synergistetes species (Fig. 2b) but not in any other bacteria. The absence of amino acid residue both the CSIs shown in Fig. 2 in all other organisms except Synergistetes indicates that these CSIs constitute inserts rather than deletions. Within the RpoC protein, in the neighborhood of the conserved insert that is shown in Fig. 2a, another very large insert consisting of between 311 and 316 aa is also uniquely present in all sequenced Synergistetes species. The sequence region corresponding to this large insert is shown in Fig. 3. BlastP searches with this insert show no significant hits for any proteins from organisms outside of other Synergistetes, indicating that this insert is a distinctive characteristic of the species of this phylum. Because of its large size, this large insert likely forms a unique domain of the RpoC protein that is only found in the Synergistetes species.

Other indels present in all genome sequenced Synergistetes, and absent in species from other taxonomic groups, are depicted in Supplementary Figs. 1–12 and some characteristics of them are summarized in Table 2. These indels are present in proteins involved in important cellular processes such as DNA replication (e.g. DNA polymerase I), protein translation (30S ribosomal protein S1) and cell metabolism (2-oxoglutarate synthase). For some of these Synergistetes specific CSIs, protein homologues for one or more Synergistetes species were not detected (Supplementary Figs. 7–13). A 3 aa insert

in 2-oxoglutarate synthase (Supplementary Fig. 7) is an example of such an indel. The insert is present in all detected Synergistetes but in this case the homologue for *F. fastidiosum* was not found in BlastP searches. It is possible that the gene coding for this protein has been lost from this species due to genetic, environmental or physiological factors. However, as fully published genome sequences among the Synergistetes species are available for only *T. acidaminovorans*, *Tv. lienii* and *Amb. colombiense*, the lack of a protein homologue for some of these species could also be due to the fact that their entire genomes have not yet been sequenced and/or annotated. Nevertheless, since these CSIs are only found in the Synergistetes species and not in any other bacteria (0/250; top 250 blast hits), they also provide reliable molecular markers for this group.

The indels identified above are completely specific for the Synergistetes species. However, for a small number of other CSIs discovered in this work, along with their presence in all Synergistetes, these CSIs were also found in a small number (usually 1–2) of species belonging to other taxonomic groups. Two such examples are shown in Fig. 4. The first of these is another 2 aa long insert in the RpoC protein (Fig. 4a). This CSI is found in all Synergistetes in a highly conserved region of the protein, however it is also present in the species *Eubacterium yurii* from the Clostridia class of the phylum Firmicutes. The CSI is not present in any other organisms, including other Firmicutes species. Different possibilities exist for the presence of the CSI in a single species outside of the phylum. The shared presence of the CSI in *E. yurii*, a species not considered to be directly related to the Synergistetes (see Fig. 1), might be the result of a lateral gene transfer event wherein the Synergistetes gene containing the indel might have been introduced into *E. yurii*. Alternatively, it is possible that two separate genetic events led to the presence of similar CSIs in Synergistetes and *E. yurii*. A second example of such an indel is shown in Fig. 4b. Here a 2 aa insert is present within the 30S ribosomal protein S8 of Synergistetes species and an *uncultured Termite group 1 phylotype RS-D17* considered to belong to the phylum Elusimicrobia. As shown, *Elusimicrobium minutum* itself contains a 1 aa insert in a similar position in the protein. It is possible that the Elusimicrobia are a sister taxon of the Synergistetes and the indel has been passed on to both phyla through a

Table 2 Characteristics of the CSIs that are specific for the Synergistetes phylum

Protein name	Gene name	GenBank identifier	Fig. no.	Indel size	Indel position ^a	Species distribution of indel	
						Synergistetes ^b	Other organisms ^c
DNA-directed RNA polymerase, β' subunit	rpoC	294101621	Fig. 2a	2 aa ins	1197–1241	9/9	–
DNA polymerase III, α subunit	polC	288574785	Fig. 2b	1 aa ins	762–807	9/9	–
DNA-directed RNA polymerase, β' subunit	rpoC	294101621	Fig. 3	313 aa ins	1197–1601	9/9	–
30S ribosomal protein S1	rpsA	282856368	Suppl. Fig. 1	1 aa del	341–386	9/9	–
S-adenosylmethyltransferase MraW	mraW	294101810	Suppl. Fig. 2	1 aa ins	182–220	9/9	–
Ribose phosphate pyrophosphokinase	rpsA	294101745	Suppl. Fig. 3	5 aa ins	137–186	9/9	–
UvrD/REP helicase	uvrD	295112231	Suppl. Fig. 4	2–3 aa ins	61–125	9/9	–
GTP-binding protein TypA	typA	288574323	Suppl. Fig. 5	11–13 aa ins	219–287	9/9	–
2-Oxoglutarate synthase	korA	288574831	Suppl. Fig. 6	3 aa ins	126–167	8/8	–
MazG family protein	–	288574801	Suppl. Fig. 7	2 aa ins	154–183	8/8	–
Integrase family protein	–	294102004	Suppl. Fig. 8	1 aa del	264–294	8/8	–
FAD dependent oxidoreductase	–	288574196	Suppl. Fig. 9	2 aa ins	263–290	8/8	–
DNA gyrase, A subunit	gyrA	288573325	Suppl. Fig. 10	1 aa del	562–594	8/8	–
DEAD/DEAH box helicase domain protein	–	294101046	Suppl. Fig. 11	2–3 aa del	278–324	7/7	–
Serine O-acetyltransferase	cysE	294101111	Suppl. Fig. 12	4 aa ins	18–53	6/6	–
DNA directed RNA polymerase, β' subunit	rpoC	288574655	Fig. 4a	2 aa ins	44–93	9/9	<i>Eubacterium yurii</i>
Ribosomal protein S8	rpsH	294101642	Fig. 4b	2 aa ins	39–86	9/9	<i>Termite group 1 Rs-D17</i>
RecA protein	recA	294101403	Suppl. Fig. 13	2 aa ins	179–228	8/8	<i>Eubacterium yurii</i>
UvrD/REP helicase	uvrD	295112231	Suppl. Fig. 14	3–4 aa ins	88–122	9/9	<i>Kredonobacter racemifer</i>
DNA polymerase I	polA	260654849	Suppl. Fig. 15	3–4 aa ins	520–552	9/9	<i>Bacteriovorax marinus</i>
DNA-directed RNA polymerase, β subunit	rpoB	282857671	Suppl. Fig. 16	6–9 aa ins	977–1020	9/9	<i>Rubrobacter xylanophilus</i>
Polyribonucleotide nucleotidyltransferase	pnp	289523245	Suppl. Fig. 17	2 aa ins ^d	342–390	8/8	<i>Gemmatimonas aurantiaca</i>

Table 2 continued

Protein name	Gene name	GenBank identifier	Fig. no.	Indel size	Indel position ^a	Species distribution of indel	
						Synergistetes ^b	Other organisms ^c
Protein of unknown function DUF1385	–	288574776	Suppl. Fig. 18	2 aa del	249–289	7/7	<i>Elusimicrobium minutum</i> , <i>Thermotoga lettingae</i> , <i>Fervidobacterium nodosum</i>
Helicase domain protein	–	288574739	Suppl. Fig. 19	3 aa del	550–604	5/5	<i>Syntrophus aciditrophicus</i>
ABC transporter, periplasmic substrate binding protein	–	289523650	Suppl. Fig. 20	2 aa del	121–164	4/4	<i>Desulfobacterium autotrophicum</i>
3,4-Dihydroxy-2-butanone 4-phosphate synthase	ribB	294101989	Suppl. Fig. 21	3 aa ins	221–270	7/8	–
DEAD/DEAH box helicase domain protein	–	282856966	Suppl. Fig. 22	6–9 aa ins	28–74	7/8	–
DNA repair protein RadA	radA	289523258	Suppl. Fig. 23	2 aa del	129–184	7/8	–
Nicotinate nucleotide adenylyltransferase	nadD	294101840	Suppl. Fig. 24	2 aa ins	94–121	7/8	–
Orotate phosphoribosyltransferase	pyrE	269792697	Suppl. Fig. 25	1 aa del	82–128	8/9	–
Cell division protein FtsA	ftsA	260655410	Suppl. Fig. 26	1–2 aa ins	10–45	7/8	<i>Sutterella wadsworthensis</i>
Putative metal dependent phosphohydrolase	–	260654227	Suppl. Fig. 27	3–4 aa ins	208–250	8/9	<i>Syntrophothermus lipocalidus</i> , <i>Thermotogales bacterium mexGI</i>

^a The indel position provided indicates the region of the protein containing the CSI

^b Homologous sequences corresponding to the region containing the CSI's could not be detected for some of the eight Synergistetes species and this is indicated accordingly by the second of the two numbers

^c BLAST searches were carried out for the top 250 hits. The number of non-Synergistetes organisms, which were observed to contain the CSI, is indicated. Species containing a larger or a shorter CSI than indicated were not included in the total

^d The 2 aa insert indicated here consists of two separate 1 aa inserts in close proximity of each other

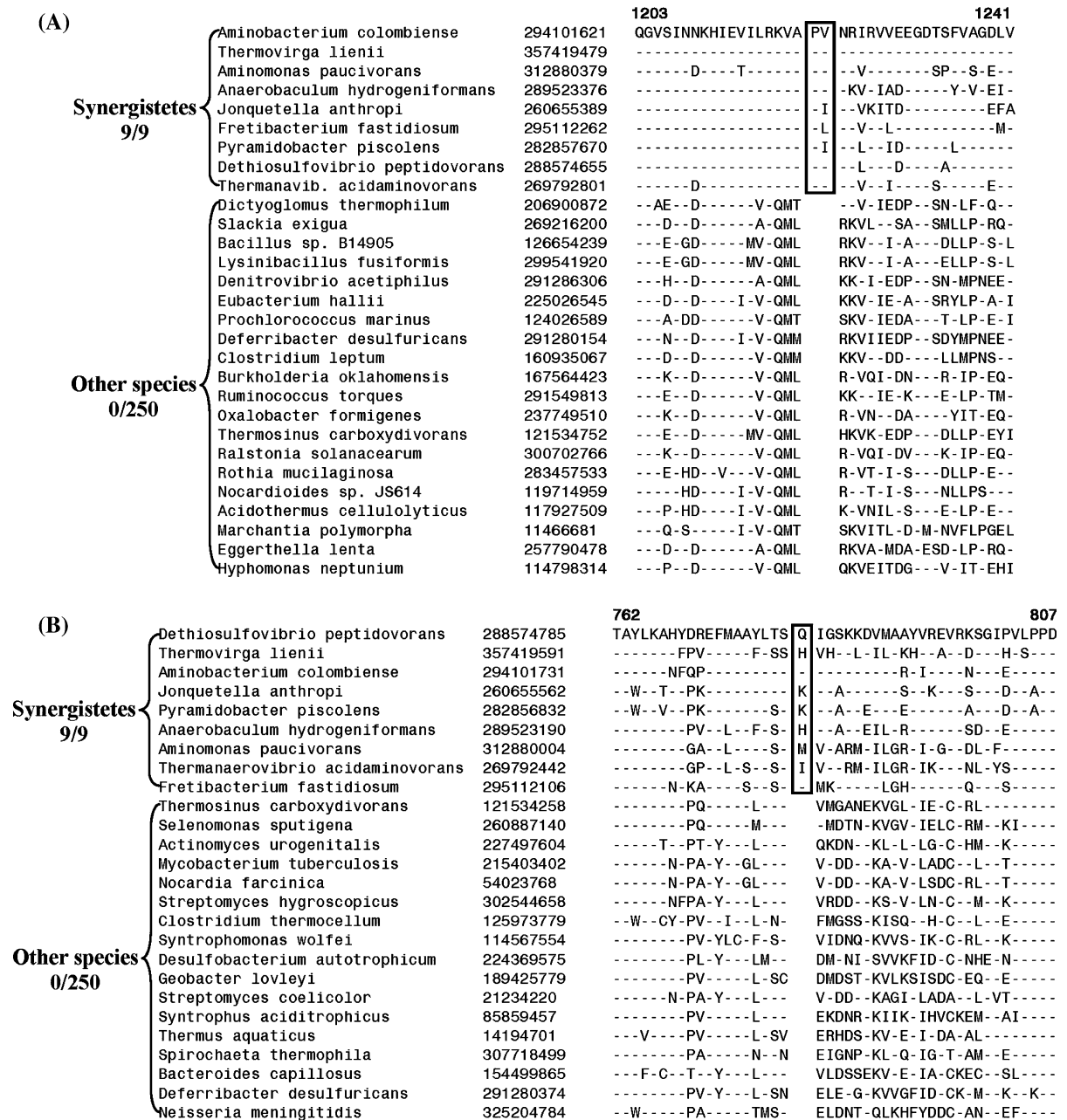


Fig. 2 Partial sequence alignments of conserved region within the **a** RNA polymerase β' subunit (RpoC) and **b** DNA polymerase III α subunit showing two CSIs (boxed) that are uniquely present in species from the Synergistetes phylum, but not in any other bacteria. The dashes (–) in this and all other alignments indicate identity with the corresponding amino acid on the top line. The numbers in the second column are the GenBank identifier numbers of the particular proteins. The numbers below the taxon identifiers indicate the number of

species detected with the indel and the total species of the respective taxon which were detected. Only representative species are shown in the alignments, however, no other species in the indicated number of blast hits contained the indel (0/250). Information for 12 other CSIs in widely distributed proteins that are specifically present in all sequenced species from the Synergistetes phylum is provided in Supplemental Figs. 1–11 and summarized in Table 2

1197	
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	QGVYRSQGSVINNKHIEVILRKVAPVNRIRVVEEGDTSFVAGDLVWTDIEEDENEAIRRENEKNIMEATRIFS QEVYESQGSVINNKHIEVILRKVAPINRVKITDEGDTSFVAGEFAWTSDEVEKIEEIKASNEKNL TEAVESLK QDVYRSQGSVINNDKHIEVILRKVAPVNRVRIIEGDSSFVAGELVWKEDLEAESRRRISEQN-ARFLEEASFLC QKVYRMQGVIEIGDKHIEVMVRQML--RKVRVIEAGDIDLPGSLI----- QNVYKSQGVAIIDKHIEVIVRQMT--SKVRIEDAGDTTFLPGELI----- QDVYRLQGVIKINDKHIEVILRQML--RKVEIAESGDSSFIKGDQM----- * * * * * :**** :*: : : : * : * : *
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	GRVVKSI AASRLTDTILQYQNMPLTEEAIRTLRPGYLISQMVLEGDEGSDLILVIGEA AFRKRMEGLELIEIT GKKLSDRGGALKGLLPDGA FDKPVTE DVLRKVLAPGGAVTELYFTDEEGP-LRVVVGEEA AFRKEMRGMELVEA GAVVKDAVG--SLDGTGITRDEELTMDKLRLLSPGVGASELICQDREG-LRVIVGEASFKRELEGLELLSD ----- -----
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	FKSEDGKE-IAAGTILTPGQLGIVTSGDPVSI CVRDHETIEKLV DSSYLAEIIVDNEVMAEKDHIFTQAMAA VTKGD-KVVPADPEP-LTVAQLSVITQGAPVPHMFRSVEKLRKQKQDEGYVAEAVIAESGLAKADQLLTEDLIE LPVADGCVISAGSR-LTASDASRIVAMGPQLLVKDLKAMEDMVGVEYLAEDVTVDGRVLSLKDRLFDLEVFQ ----- -----
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	ICFEHNIQAVKIWHSVERISVLDALQERLMNNIWRHRLTQAVDSEGNGLTDV SQMVDARIIRGLV DDEISAVD QIRRSSAPSVKIWKVTDITISVRDLQERLINRIWGRQLKLAIGADGQALSGSVHMVDGSSVVKGIVEGQIQGLV ELRSLPVESVRIWRNPERLDLCKDVYEYLIGNYLSQRVLRVITRDGAVTEPLDNRISMEIAEGIRSGEVEAIE ----- -----
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	LEG-EILTREKALIELNLTLYGKVLLEPVDVEKQILVDSGQEIINRAMIDL L VRSQAGEFVVRPLSARHDEK FADDSNTSRETELTEALSSVSGKVLLEDVKNSDGTVALQAGQIEGKKQLAKIVAADPTILTVRPVLDQTE TV LDGNNVSRERVLKALLTEKVYGVKVLLEPVRVDVGNVVPVSGREVSHQVMDQIVAACPGEMVVRPILAQGEHR -----DIHQ -----ELRQ -----ELTH
1617	
Aminobac. colombiense Jonquetella anthropi Therm. acidaminovorans Bacillus sp. B14905 Prochlorococcus marinus Pseudomonas fluorescens	TLIWDVTFVRKLRREGPKCRPFVHGITKAALATESFLSAASFQQTAAQLAGAAVKGEMD QLISRVSFVRRRLR L GPQWRPF IHGVTKAALATDSFLSAASFQQTAAQLAGAAVRNQVD RLIQRISFVRRRLRELPTWKPV LHGITKAALATDSFLSAASFQQTAAQLASAAVRGEVD FAEANADAVMNGKNPATCRPVLGITKASLETESFLSAASFQETTRVLTDAAIKGRD VEDTNQAISITGGAPSEFTPVLLGITKASLNTDSFISAASFQETTRVLTAAIEGKSD VLVENERLSTEDKFAKFTRVLLGITKASLSTESFISAASFQETTRVLTAAAVTGKRD : * : * * * : * * : * * * * * * : * : * : * : *

Fig. 3 Partial amino acid sequence alignment of the RpoC protein showing a large insert that is specifically present in all sequenced Synergistetes species. Partial sequence for the neighbouring regions is also shown in the alignment. The *dashes* in this particular alignment represent sequence gaps. The identical and conserved residues in this alignment are indicated by * and

semicolons (:), respectively. Blastp searches with the insert sequence (without the flanking region) show no significant hit for any protein except for the RpoC homologs from the Synergistetes species. Sequence information is shown for only a few Synergistetes, but this insert is present in all sequenced species

common ancestor. However, this postulation is not supported by the phylogenetic trees (Fig. 1) and it is possible that the CSI in these two taxa occurred independently or by means of LGT. The information for other CSIs where indels found in Synergistetes are also present in one or two species from other taxa is summarized in Table 2 and the sequence

alignments for these are presented in Supplementary Figs. 13–20.

A further seven CSIs, specific for species of the Synergistetes phylum, were discovered where one species from the phylum was detected to lack the indel. A 3 aa insert in the 3-4-dihydroxy-2-butanone 4-phosphate synthase (Supplementary Fig. 21) is an

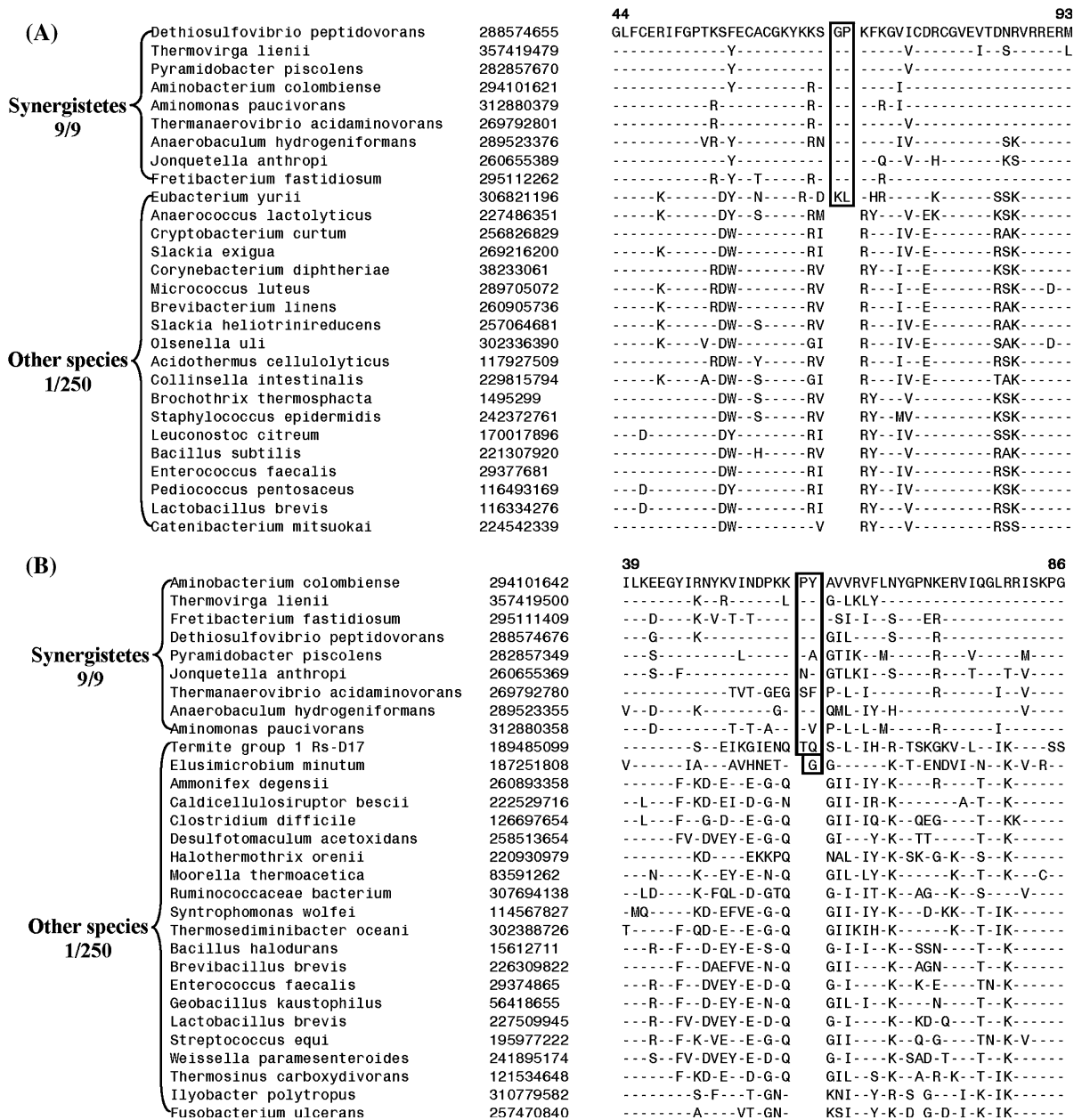


Fig. 4 Partial sequence alignments of RpoC and RpsH proteins showing two CSIs, which in addition to the Synergistetes species are also present in isolated other bacterial species. **a** Excerpt from RpoC sequence alignment depicting a 2 aa conserved insert which in addition to the Synergistetes is also present in *Eubacterium yurii*. **b** Sequence alignment of the ribosomal protein S8 (RpsH) showing a 2 aa insert which in

addition to the Synergistetes is also found in an *Uncultured Termite group 1 bacterium phylotype Rs-D17*. A 1 aa insert in this position is also present in *Elusimicrobia minutum*. Sequence information for 8 other CSIs in different protein containing an isolated exception is provided in Supplementary Figs. 13–20 and Table 2

example of such a CSI. The insert is present in all detected species of the Synergistetes except for *P. piscolens*. No species outside of the phylum contain the insert. The information for other such CSIs is

summarized in Table 2 and sequence alignments for them are presented in Supplementary Figs. 21–27. It is possible that these CSIs were also originally introduced in a common ancestor of the Synergistetes

phylum but they were lost in some species over time due to ecological/physiological pressures or by mechanisms such as LGT followed by gene loss. In some of the CSIs described above, in addition to the CSIs that were specific for the Synergistetes, indels of different lengths were also present in species from other taxonomic groups. Due to their different lengths, these CSIs have likely originated from independent genetic events.

CSIs that are specific for subgroups of the Synergistetes phylum

All Synergistetes species are currently classified as part of a single class (Synergistia), order (Synergistales) and family (Synergistaceae) (Jumas-Bilak et al. 2009; www.bacterio.cict.fr). The relationships among the species/genera of this phylum are not well understood. In the phylogenetic trees based upon concatenated protein sequences and the 16S rRNA a number of strongly supported relationships among the species within this phylum are observed (Fig. 1). Importantly, in the present work, our analyses of protein sequences from Synergistetes have led to discovery of several CSIs that are commonly shared only by species from this phylum and that are absent in all others. These CSIs independently support a specific evolutionary relationship among these species and they, in

conjunction with the results from phylogenetic analyses, can be used for determination of the relationships among the members of the phylum Synergistetes.

In the phylogenetic trees shown in Fig. 1, a clade consisting of *D. peptidovorans*, *J. anthropi* and *P. piscolens* is supported with high statistical support in both the concatenated protein tree and the rRNA tree. In our analysis, we have identified seven indels (Table 3) that are uniquely present in these three species supporting independently that these three species are closely related and form a distinct clade within the Synergistetes phylum. The first of these is a 4 aa deletion in the penicillin-binding protein 1A family protein which is involved in cell wall construction (Fig. 5). This deletion is found only in homologues of the protein from *D. peptidovorans*, *P. piscolens* and *J. anthropi* and all other Synergistetes, as well as non-Synergistetes species, lack this deletion. An additional 6 CSIs specific to these three organisms were discovered in the proteins tRNA modification enzyme TrmE, ribonucleoside diphosphate reductase, putative DEAD/DEAH box helicase, RpoB and the PlsC proteins. Information for these CSIs is summarized in Table 3 and their sequence alignments are presented in Supplementary Figs. 28–33. Among the three organisms which are part of this clade, *J. anthropi* and *P. piscolens* were observed as being more closely related to each other than either is to *D. peptidovorans*. This close association is underscored by a total of 15 CSIs, including an example that is shown

Table 3 Characteristics of the CSIs that are Specific for a Clade Consisting of *J. anthropi*, *P. piscolens* and *D. peptidovorans*

Protein name	Gene name	GenBank identifier	Figure no.	Indel size	Indel position ^a	Other species containing indel ^b
Penicillin binding protein, 1A family	–	288574813	Fig. 5	4 aa del	102–140	–
Ribonucleoside diphosphate reductase	nrdA	260654687	Suppl. Fig. 28	1 aa ins	193–225	–
Putative DEAD/DEAH box helicase	–	260655128	Suppl. Fig. 29	1 aa del	398–457	–
Putative DEAD/DEAH box helicase	–	260655128	Suppl. Fig. 30	6–8 aa ins	437–496	–
DNA directed RNA polymerase, β subunit	rpoB	282857671	Suppl. Fig. 31	13 aa ins	358–407	–
1-Acyl-sn-glycerol-3-phosphate acyltransferase	plsC	282855432	Suppl. Fig. 32	1 aa ins	57–84	<i>P. staleyi</i> , <i>T. mathranii</i>
tRNA modification GTPase TrmE	trmE	260655716	Suppl. Fig. 33	1 aa ins	263–299	<i>E. minutum</i> , <i>Termite group 1 Rs-D17</i>

^a The indel position provided indicates the region of the protein containing the CSI

^b BLAST searches were carried out for the top 250 hits. Organisms, other than *J. anthropi*, *P. piscolens* and *D. peptidovorans*, which were observed to contain the CSI are indicated. Species containing a larger or a shorter CSI than indicated were not included in the total

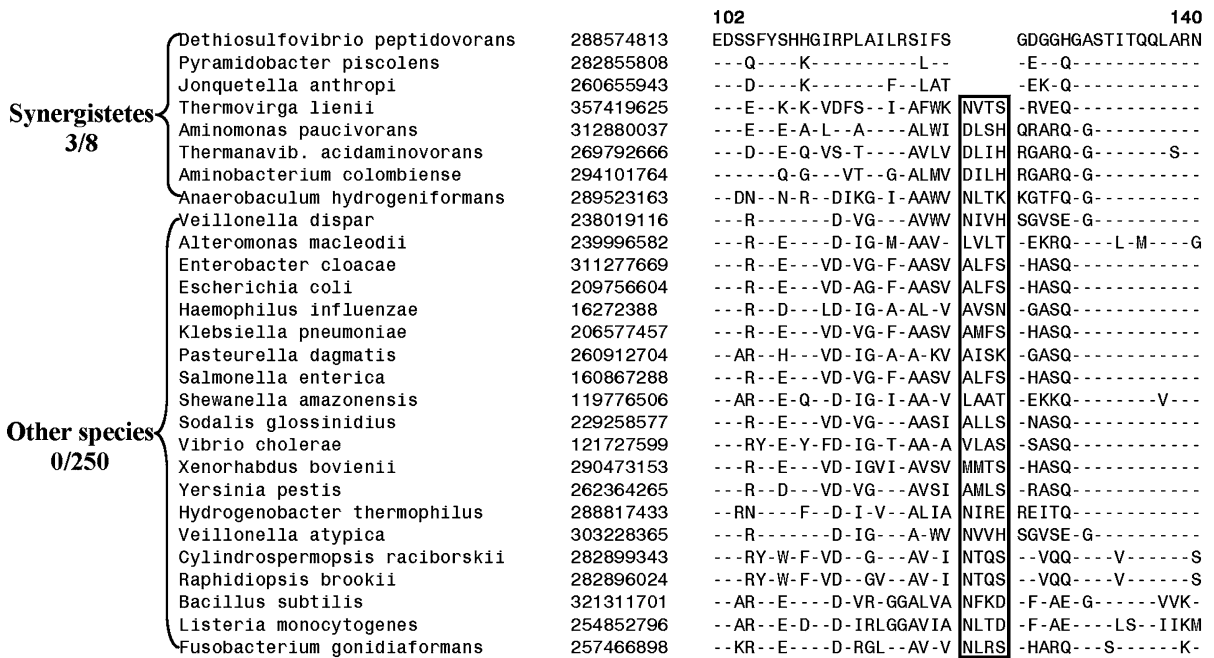


Fig. 5 Partial sequence alignment of a family 1A penicillin-binding protein containing a 4 aa deletion that is specific for *D. peptidovorans*, *P. piscolens* and *J. anthropi*. Sequence

information for five other CSIs that are specific for this clade of species is presented in Table 3 and Supplementary Figs. 28–33

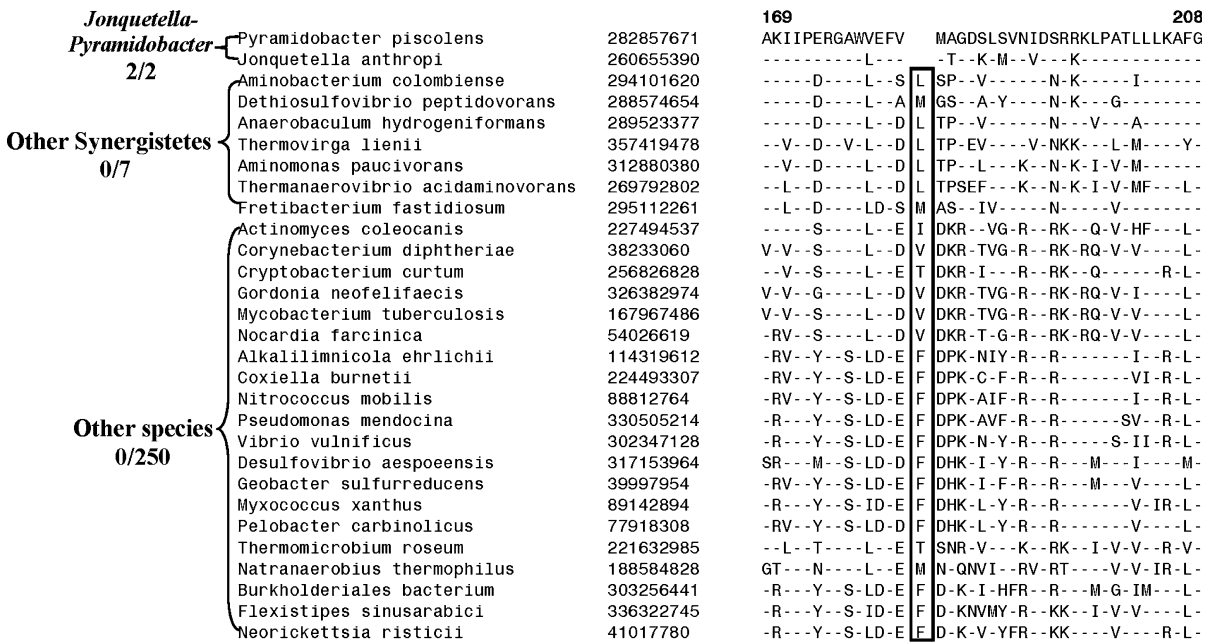


Fig. 6 Excerpts from sequence alignment for RNA polymerase β subunit (RpoB) showing a 1 aa deletion that is specifically present in *J. anthropi* and *P. piscolens*. The region contains a 1

aa deletion specific for the three species. Sequence information for 14 other CSIs that are specific for these two species is presented in Table 4 and Supplementary Figs. 34–47

Table 4 Characteristics of CSIs that are specific for *J. anthropi* and *P. piscicolens*

Protein name	Gene name	GenBank identifier	Figure no.	Indel size	Indel position ^a	Other species containing indel ^b
DNA-directed RNA polymerase, β subunit	rpoB	282857671	Fig. 6	1 aa del	169–190	–
Chlorohydrolase family protein	–	260654152	Suppl. Fig. 34	1 aa ins	241–272	–
Phospho- <i>N</i> -acetylmuramoyl-pentapeptide-transferase	mraY	260655416	Suppl. Fig. 35	1 aa ins	218–266	–
Recombination protein RecR	recR	260654559	Suppl. Fig. 36	1 aa ins	63–122	–
Transcriptional regulator NrdR	nrdR	260655521	Suppl. Fig. 37	1 aa del	84–133	–
Lipid A biosynthesis acyltransferase	–	282855413	Suppl. Fig. 38	1 aa ins	262–284	–
Glutamate synthase	gltB	282858093	Suppl. Fig. 39	6 aa ins	334–383	–
Acetate kinase	ackA	282856654	Suppl. Fig. 40	2 aa del	238–279	–
AcrB/D/F family transporter	–	260654486	Suppl. Fig. 41	1 aa del	111–154	–
Transcriptional regulator IclR	iclR	282857025	Suppl. Fig. 42	10–13 aa ins	139–202	–
Ribose phosphate pyrophosphokinase	kprS	282858157	Suppl. Fig. 43	1 aa ins	204–246	<i>Dichelobacter nodosus</i>
Phosphoribosyl-formylglycinamide synthase II	purL	260655584	Suppl. Fig. 44	4 aa del	102–150	<i>Hydrogenobaculum</i> sp. Y04AAS1
MazG family protein	–	282858161	Suppl. Fig. 45	3–4 aa ins	33–75	<i>Clostridium phytofermentans</i>
Glutamate synthase	gltB	282858093	Suppl. Fig. 46	8 aa ins	84–130	<i>Desulfovibrio africanus</i>
<i>S</i> -adenosylmethytransferase MraW ^c	mraW	289522914	Suppl. Fig. 47	1 aa ins	153–196	–

^a The indel position provided indicates the region of the protein containing the CSI

^b BLAST searches were carried out for the top 250 hits. Organisms, other than *J. anthropi* or *P. piscicolens*, which were observed to contain the CSI are indicated. Species containing a larger or a shorter CSI than indicated were not included in the total

^c All Synergistetes species were observed to contain the indel except *J. anthropi* and *P. piscicolens*, thus, differentiating these two species from the rest of the phylum

in Fig. 6, a 1 aa deletion in a conserved region of the enzyme RNA polymerase β subunit. Other indels that provide similar molecular evidence for the observed close relationships between these two genera are presented in Supplementary Figs. 34–47 and information for them is summarized in Table 4. The fidelity of these molecular markers can be tested on cultured but unsequenced members of the phylum Synergistetes and as more species belonging to these genera are sequenced, the identified CSIs should provide molecular markers for their induction into the clade formed by this sub-group of the phylum.

The phylogenetic trees also support a cladal relationship among two other species, *Amm. paucivorans* and *T. acidaminovorans*, which branch as sister organisms with high statistical support (Fig. 1). The clade harbouring these genera has been proposed to form a higher-level taxon within the phylum (Jumas-Bilak et al. 2009). In the present work, we have identified 7 CSIs that differentiate the species representing these two

genera from all other species and support a specific grouping of the genera *Thermanaerovibrio* and *Aminomonas* (Table 5). Among these CSIs is a 2 aa insert in enzyme *S*-adenosyl-methionine isomerase (Fig. 7). The information for 6 other CSIs supporting a specific relationship among these two species is provided in Table 5 and their sequence alignments are depicted in Supplementary Figs. 48–53. Two other CSIs identified in the present work, which include a 1–2 aa deletion in the ribosomal protein L13 (Supplementary Fig. 54) and 2 aa deletion in DNA gyrase B (Supplementary Fig. 55), are present in all detected Synergistetes species except *Thermanaerovibrio* and *Aminomonas*. The absence of these CSIs in the two species suggests that this clade may have diverged from the common Synergistetes ancestor before the other species of the phylum and the two indels may have been introduced after the divergence of this clade from the common Synergistetes ancestor. A loss of this signature from this clade after its divergence from other Synergistetes can also explain the

Table 5 Characteristics of the CSIs that are specific for a clade consisting of *T. acidaminovorans* and *Amm. paucivorans*

Protein name	Gene name	GenBank identifier	Figure no.	Indel size	Indel position ^a	Other species containing the indel ^b
S-adenosylmethionine/tRNA-ribosyltransferase-isomerase	queA	269792529	Fig. 7	2 aa ins	156–194	–
RecA protein	recA	269793250	Suppl. Fig. 48	1 aa ins	143–184	–
Glu/Leu/Phe/Val dehydrogenase	–	269791934	Suppl. Fig. 49	1 aa del	318–361	–
Uracil phosphoribosyltransferase ^c	upp	312880140	Suppl. Fig. 50	4–5 aa ins	133–184	–
Methyltransferase GidB	gidB	269791772	Suppl. Fig. 51	2 aa ins	75–123	<i>Sorghum bicolor</i>
Xanthine/uracil/vitamin C permease	–	269792033	Suppl. Fig. 52	5 aa ins	390–443	<i>Mesembryanthemum crystallinum</i>
Electron transport complex, RnfABCDGE type, C subunit	–	312880739	Suppl. Fig. 53	1 aa ins	166–213	<i>Saccharophagus degradans</i> , <i>marine gamma proteobacterium</i> , <i>Eubacterium cellulosolvens</i>
Ribosomal protein L13 ^d	rplM	294101309	Suppl. Fig. 54	1–2 aa del	108–138	–
DNA gyrase subunit B ^d	gyrB	294102629	Suppl. Fig. 55	2 aa del	191–234	<i>Acidaminococcus fermentans</i> , <i>Acetonea longum</i> , <i>Seinonella peptonophila</i>
Hypothetical protein Tac1_0455 ^e	–	269792069	Suppl. Fig. 56	2 aa ins	205–256	–

^a The indel position provided indicates the region of the protein containing the CSI

^b BLAST searches were carried out for the top 250 hits. The number of non-Synergistetes organisms, which were observed to contain the CSI, is indicated. Species containing a larger or a shorter CSI than indicated were not included in the total

^c BLAST searches were carried out for the top 250 hits. However, in the indicated case, no species outside of the Synergistetes phylum contained the protein homolog or the conserved region corresponding to the sequences flanking the indel

^d All Synergistetes species were observed to contain the indel except *Amm. paucivorans* and *T. acidaminovorans*, thus, differentiating these two species from the rest of the phylum

^e The CSI is also present in *Thermovirga lienii* species from the Synergistetes phylum

observation. These indels also support a close relationship among the genera *Thermanaerovibrio*–*Aminomonas* and information for them is also summarized in Table 5. These two species were observed to branch in a weakly supported clade with the *Tv. lienii* and *An. hydrogeniformans* (Fig. 1). However, only 1 CSI supporting the three-species-clade with *Tv. lienii* was identified in a protein of unknown function (Supplementary Fig. 56) and no CSI specific for all four organisms was discovered.

In the phylogenetic trees (Fig. 1), the species *Amb. colombiense* and *F. fastidiosum* were observed to branch with *J. anthropi*, *P. piscolens* and *D.*

peptidovorans. A specific relationship among these species is also supported by two of the identified CSIs. The first of these CSIs consists of a 1 aa del in GyrB that is uniquely present in all five of these species (Fig. 8a). Another CSI in orotidine 5'-phosphate decarboxylase, also consisting of a 1 aa deletion, is commonly shared by *Amb. Colombiense*, *J. anthropi*, *P. piscolens* and *D. peptidovorans* (Fig. 8b). A homologue for this protein was not detected for *F. fastidiosum*, whose genome has not been fully sequenced. Thus, it is likely that this CSI will also be present in this species and could provide an additional molecular marker for this clade.

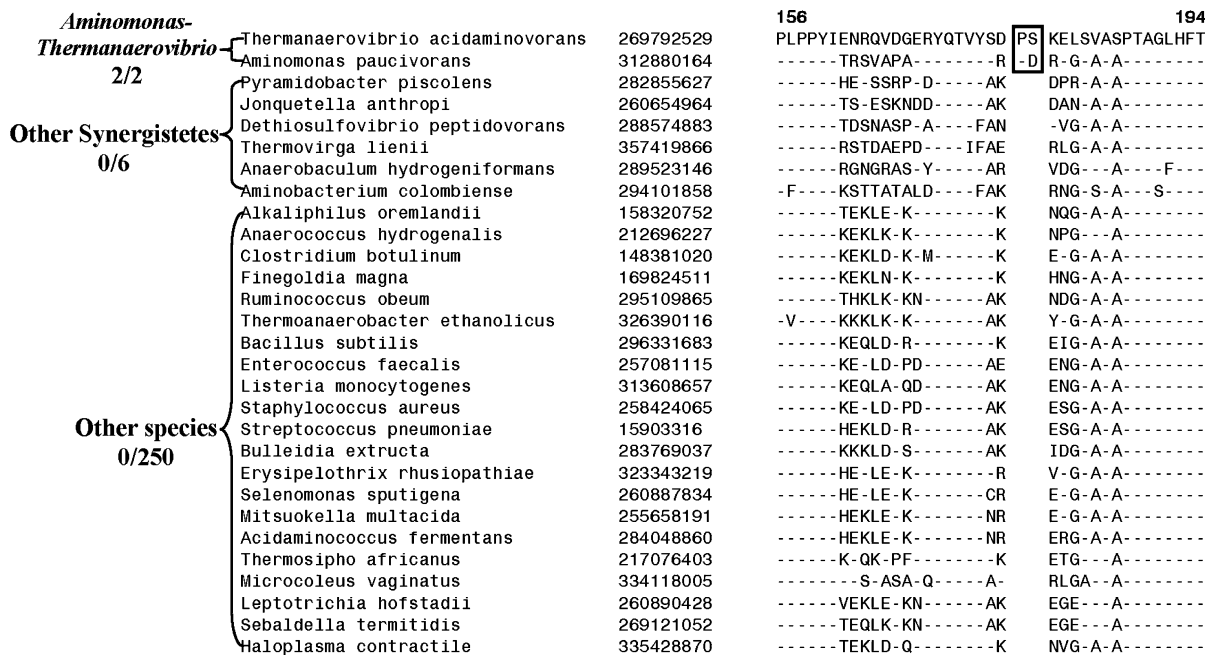


Fig. 7 Partial sequence alignment of *S*-adenosylmethionine/*t*RNA-ribosyltransferase-isomerase protein showing a 2 aa insert in a conserved region that is specific for *T.*

acidaminovorans and *Amm. paucivorans* species. Nine other CSIs that are specific for these two species have also been identified (Table 4 and Supplementary Figs. 34–47)

CSIs that are commonly shared by species of the Synergistetes phylum with other taxonomic groups

The Synergistetes is a taxonomic group that has only recently been identified as a separate phylum within the bacterial domain. Though it branches distinctly in the 16S rRNA trees with long branches separating it from other bacterial groups (Fig. 1; Jumas-Bilak et al. 2009), species from the phylum had previously been classified as part of *Syntrophomonadaceae* family in the Firmicutes (Baena et al. 1998, 1999a; Diaz et al. 2007), grouped with *Deferribacteres* (Garrity et al. 2004) and misclassified as *Selenomonas* (Guangsheng et al. 1992). The presence or absence of CSIs that associate these groups with the Synergistetes should prove helpful in determining whether any link exists between the Synergistetes and these other groups of bacteria.

In our analysis we have identified some CSIs that, along with being present in some or all the Synergistetes species, were present in other groups of organisms. Two examples of such indels are presented in Fig. 9. Figure 9a shows a 1 aa insert in the MiaB-family RNA modification enzyme that is uniquely

present in all detected Synergistetes as well as various species from the phylum *Chloroflexi*. All other bacteria lack this insert. Similarly, in the DNA polymerase III α -subunit, a 1 aa insert is present in all detected Synergistetes and also in various *Fusobacteria*, an *Opitutaceae* species as well as in *Thermomicrobium* (Fig. 9b). In phylogenetic trees constructed from these protein sequences, the Synergistetes species do not branch with species from these taxa (unpublished results) indicating that the shared presence of these CSIs is not due to their being sister taxa of Synergistetes or LGTs. The CSIs in these groups have thus likely originated independently. Other CSIs that the Synergistetes share with species from other taxonomic groups are listed in Table 6 and sequence information for them is provided in Supplementary Figs. 57–74. These other taxa include the *Fusobacteria* (Supplementary Figs. 57–61), *Elusimicrobia* (Supplementary Figs. 61, 62), class *Negativicutes* (Supplementary Figs. 63–66), *Acidobacteria* (Supplementary Fig. 67), *Proteobacteria* (Supplementary Figs. 68–70), *Aquificae* (Supplementary Fig. 71), *Erysipelotrichi* (Supplementary Fig. 72), *Actinobacteria* (Supplementary Fig. 73) and order *Lactobacillales* (Supplementary Fig. 74). The Synergistetes share the greatest number

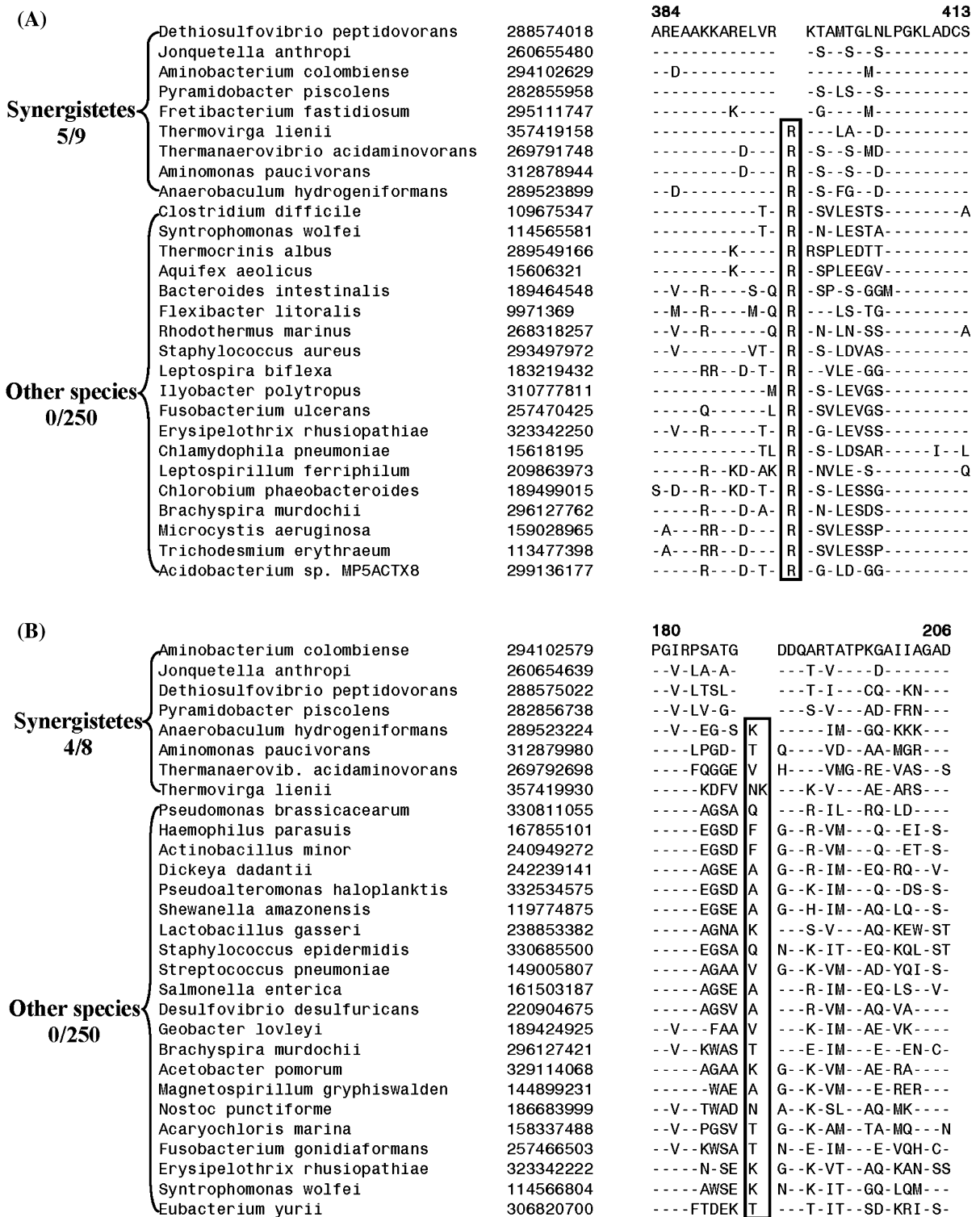


Fig. 8 Partial sequence alignments of a DNA Gyrase subunit B and b orotidine 5'-phosphate decarboxylase showing two CSIs in conserved regions that are specific for the species

D. peptidovorans, *J. anthropi*, *Amb. colombiense*, *P. piscolens*, *F. fastidiosum*, which form or define a higher clade within the Synergistetes group of species

Table 6 Conserved indels common to Synergistetes and shared with other groups

Protein	Beta-ketoacyl-acyl-carrier protein synthase II	Xaa-Pro dipeptidase	ATP-dependent protease La	Phosphoribosyl-formylglycin-amidime cycloligase	Translation elongation factor IA	GTP-binding protein Era
GenBank identifier	260654633	260654796	260654937	294102156	312880374	294101756
Accession no.	ZP_05860123	ZP_05860284	ZP_05860425	YP_003554014	ZP_07740174	YP_003553614
Indel/size	2 aa del	1 aa ins	1 aa ins	1 aa del	1 aa del	1 aa ins
Indel position ^a	273–312	220–260	625–672	253–299	231–269	75–121
Figure no.	Suppl. Fig. 57	Suppl. Fig. 58	Suppl. Fig. 59	Suppl. Fig. 60	Suppl. Fig. 61	Suppl. Fig. 62
Synergistetes species containing indel	<i>J. anthropi</i> and <i>P. piscotens</i>	<i>D. peptidovorans</i> , <i>P. piscotens</i> and <i>J. anthropi</i>	<i>J. anthropi</i> and <i>P. piscotens</i>	All detected Synergistetes	<i>Amm. paucivorans</i> , <i>T. acidaminovorans</i> , <i>D. peptidovorans</i>	All detected Synergistetes
Other species with indel	Epsilon- proteobacteria	Some Fusobacteria	Fusobacteria and <i>Treponema pallidum</i>	Two δ -proteobac. and Fusobacteria	Proteobacteria, two Elusimicrobia	Some Clostridia species, <i>Termite group I Rs-D17</i>
Protein	NADH dehydrogenase	Translation elongation factor Tu	Penicillin-binding protein 2	Ribonuclease, Rne/Rng family	DNA-directed RNA polymerase, subunit B	Ribosomal RNA large subunit methyltransferase J
GenBank identifier	312880263	294101321	288574639	269792810	288574654	289522985
Accession no.	ZP_07740063	YP_003553179	ZP_06392996	YP_003317714	ZP_06393011	ZP_06439839
Indel/size	1 aa del	1 aa ins	3–4 aa ins	1 aa ins	1 aa ins	1 aa ins
Indel position ^a	31–86	228–270	171–205	14–47	412–457	174–233
Figure no.	Suppl. Fig. 63	Suppl. Fig. 64	Suppl. Fig. 65	Suppl. Fig. 66	Suppl. Fig. 67	Suppl. Fig. 68
Synergistetes species containing indel	All detected Synergistetes	<i>Amb. colombiense</i> , <i>Tv. itenii</i> and <i>An. hydrogeniformans</i>	<i>P. piscotens</i> , <i>D. peptidovorans</i>	<i>A. paucivorans</i> , <i>T. acidamino- vorans</i>	All detected Synergistetes	All detected Synergistetes
Other species with indel	Negativicutes	Negativicutes	<i>Veillonella</i> species	<i>Veillonella</i> species	Acidobacteria and a few clostridia	Many alpha proteobacteria
Protein	GTP-binding protein Era	Orotidine 5'-phosphate decarboxylase	Ribosomal protein S5	Homoserine kinase	1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	3-Oxoacid CoA-transferase, subunit B
GenBank identifier	294101756	289523224	260655366	294101575	294101868	288573369
Accession no.	YP_003553614	ZP_06440078	ZP_05860854	YP_003553433	YP_003553726	ZP_06391726
Indel/size	4 aa ins	1 aa del	1 aa ins	1 aa ins	1 aa ins	2 aa del
Indel position ^a	213–259	149–177	105–144	222–258	171–218	43–83
Figure no.	Suppl. Fig. 69	Suppl. Fig. 70	Suppl. Fig. 71	Suppl. Fig. 72	Suppl. Fig. 73	Suppl. Fig. 74
Synergistetes species containing indel	All detected Synergistetes	Some Synergistetes	All detected Synergistetes	All detected Synergistetes	Most Synergistetes	All detected Synergistetes
Other species with indel	<i>Anaeromyxobacter</i> , <i>Prochlorococcus marinus</i>	<i>Pseudomonas</i> species and <i>A. caldus</i>	<i>Hydrogenothermaceae</i> species	<i>Erysipelotrichi</i> and <i>A. viridans</i>	Most detected Actinobacteria	Few <i>Streptococci</i>

^a The indel position provided indicates the region of the protein containing the CSI

(five) of these CSIs with the Fusobacteria and they share only 1–2 indels with most other taxonomic groups. In many cases where the Synergistetes share CSIs with other taxa, only some species from the Synergistetes or the other taxa contain the indel. The CSIs in these other groups may have arisen independently through separate genetic events or it is also plausible that their shared presence in some of these cases is due to LGTs.

Discussion and concluding remarks

The Synergistetes are a relatively unknown group of species living ubiquitously in anaerobic environments. Though characteristics for the isolated Synergistetes are known, such as their gram-negative morphology and their ability to ferment amino acids, no single molecular, morphological or physiological characteristic is known that distinguishes them as a group from other bacterial organisms. Utilizing the available genomic data for this group of organisms, we report here identification of over 60 novel CSIs specific for the species of the Synergistetes phylum. Of the various discovered CSIs, 32 were identified to be specific for all or most Synergistetes species (maximum of three exceptions unrelated to each other). These CSIs are present in widely distributed proteins with important cellular functions and they are rarely present in protein homologues of species outside of the phylum. As they are present in most or all Synergistetes and absent in bacteria from all other taxonomic groups, they provide strong evidence that species of the Synergistetes phylum constitute a monophyletic group that is distinct from all other prokaryotic taxa. These CSIs also provide novel molecular means for identification and circumscription of species from this phylum.

The bacteria belonging to the Synergistetes have been classified into 12 different genera (and a candidate genus) within the phylum. Despite the recognition of numerous species and genera, due to the lack of reliable biological characteristics that can identify the interrelationships among these bacteria, all genera are presently grouped into a single class, order and family. Numerous CSIs were discovered during the course of the study that were present in only certain clades of species within the Synergistetes phylum and absent from others. The group specificities of these CSIs are summarized in Fig. 10. Explicitly, 7 CSI were

detected to be specifically found in only the *J. anthropi*, *P. piscolens* and *D. peptidovorans* species; 15 CSIs were identified that are specific for the *J. anthropi* and *P. piscolens* species (or differentiate them from other Synergistetes) and 9 other CSIs differentiated the *T. acidaminovorans* and *Amm. paucivorans* from other members of the phyla. In addition, two of the discovered CSIs also supported a grouping together of the *J. anthropi*, *P. piscolens*, *D. peptidovorans*, *Amb. colombiense* and *F. fastidiosum* species. These relationships are also consistently observed in phylogenetic trees created for the Synergistetes group and the identified CSIs provide valuable markers that consolidate these relationships. Furthermore, it should be noted that in contrast to the CSIs supporting these relationships, very few, if any, CSIs that supported alternative relationships among these species were detected. Thus, the identified CSIs provide independent evidence for the existence of these clades and provide molecular means to demarcate and circumscribe these clades. The evidence based upon identified CSIs supports the division of the phylum Synergistetes into a number of distinct families (or other higher taxonomic groupings) and a formal proposal in this regard will be made in future work. Though the branching and interrelationships of several species within the phylum is well supported by multiple CSIs, the relationships of *Tv. lienii* and *An. hydrogeniformans*, and also to some extent *Amb. colombiense* and *F. fastidiosum* to other Synergistetes species were not resolved by the identified CSIs. This problem may be addressed as genome sequences for additional Synergistetes species become available.

As previously mentioned, the Synergistetes have often been misclassified as a lower ranked taxonomic group with bacteria belonging to other phylogenetic divisions. In our analysis, some CSIs were also discovered that were shared by Synergistetes species along with species from other taxonomic groups. Some of the organisms sharing such indels included species from the Fusobacteria, Chloroflexi, Proteobacteria, Acidobacteria, Aquificae and Firmicutes phyla. Most of these groups shared no more than 1–3 CSIs and in many cases only a few species within the groups contained the indels. Geissinger et al. (2009) presented a study suggesting a shared common ancestor for *Elusimicrobium* and Synergistetes and a recent study by Gupta (2011) also suggested that the *Negativicutes*, *Fusobacteria*, *Elusimicrobia* and

(A)		189	224
Synergistetes 7/7	Jonquetella anthropi	260654222	YGADLYKKRSLPKLLTELEK T LPQSVLRLFLYHPS
	Pyramidobacter piscolens	282855982	--S--FGRP---R--D---- E --RG-----
	Dethiosulfovibrio peptidovorans	288573284	--S--S-RG--SG--DAM-A E --EDI-----
	Aminobacterium colombiense	294101399	--R-V-G-P--IA--DS-TA S ---E-----L-----A
	Anaerobaculum hydrogeniformans	289524260	--M-WDGSSH-VE--DL-- H V-DGM-I-PL----
	Aminomonas paucivorans	312879348	--L-RGGRE--TD--DA--P S --GD-F---L----
	Thermanavib. acidaminovorans	269793254	--E-LGTD--ME--DQM-A- VRGHGV---L----T
Chloroflexi 6/6	Sphaerobacter thermophilus	269837471	----GI-NG--G--RMIAE E V-DLP---ML-IY--
	Roseiflexus castenholzii	156741270	--R--GL-DG-AI--D-ICA V --EN--V--M-AY-
	Roseiflexus sp. RS-1	148654700	--R--GL-DG-AL--D--CA V --KDR-V--M-AY-
	Chloroflexus aggregans	219848510	--R--GLRDG-AI--D--CQ V T-SDI-I--M-AY-
	Chloroflexus aurantiacus	163847415	--R--GL-DG-AT--A--CQ V T-PET-I--M-AY-
	Oscillochloris trichoides	309792044	--R--GLQDG-AT--E-MCQ I V-H-G-I--M-AY-
Other species 0/250	Thermotoga maritima	15644605	--I--R-QA--D--RR-NS -NGEF-I-VM----
	Clostridium botulinum	187933390	--S-I-G-KN-HV--K--S- IEGIK-I-VL-CY-
	Eubacterium yurii	306819660	--L-I-GEKK--N--R--S- IEGIR-I-FL-TY-
	Roseburia intestinalis	240144123	--V--GEK--HR--D--N- IKDLF-I-IM-CY-
	Syntrophomonas wolfei	114566788	--H-ISPQSA--T--R--S- -DGL-E-I--M----
	Thermoanaerobacter italicus	289578370	--I-I--FM--Q--K--SL I-NLK-I--L-AY-
	Fusobacterium varium	253583638	--I--EK--AR-MK--V- IDGLK--TY-MF-
	Ilyobacter polytropus	310779080	--I----KA--DVMKA-S- VEGIE-I-TY-MF-N
	Bacteroides finegoldii	255693887	--V----QM--E-IERISD I-GVE-I--H-AY-A
	Psychroflexus torquis	91215250	--L-----N-AE--R--V- VEGIE-I--H-AF-T
	Bacillus tusciae	295696216	--L--GR-R--D--KA-ND VDEL-R-I--H-AY-
	Brachyspira murdochii	296125911	--H-I--RLA--D--K--S- IEGIE-I-VL-QN-
	Aquifex aeolicus	15606200	--K--EYK-VE--EG-- -VEGIK-I--L--Y-T
	Thermocrinis albus	289547853	--K--HRKA-VQ--KK--E -EGIE-I--L--Y-T
	Prochlorococcus marinus	123967655	--Q-I-G-P--A--N--S- VVIP-I-IH-AY-T
	Leptospirillum rubarum	124514462	--S--EDGEG--R--E-IDR IGRIP-V--L-AY-T
	Dialister microaerophilus	313891928	--Q--RDGT--IL--KQ-V- I-EVK-I-----Y-T
(B)		244	274
Synergistetes 9/9	Dethiosulfovibrio peptidovorans	288574785	FYLRSAAEMWQIFG D DVPEALENTLKAERC
	Thermovirga lienii	357419591	----P----A---- K EL---T--VN----
	Jonquetella anthropi	260655562	----P----NF-- A E--D--T--VE----
	Aminobacterium colombiense	294101731	----P----DSL-- A EL-D--D--A-----
	Pyramidobacter piscolens	282856832	--F--Q--GY-- S EA-DS----R----
	Thermanaerovibrio acidaminovorans	269792442	----TP----SLL- G E-----R--LV----
	Aminomonas paucivorans	312880004	--F--P----SL-- A EL-D--R--QE--D--
	Anaerobaculum hydrogeniformans	289523190	----PV--K-- - ELSD--H--VD-S----
	Fretibacterium fastidiosum	295112106	----P--DA-- A E-----D--VA----
Fusobacteria 4/4	Fusobacterium sp. 3_1_5R	257452286	L--K-Y--YAVL- E QYQ--Q-SVE--K--
	Fusobacterium periodonticum	262066954	L--K-K--QRSLD E KFKH-I--NY--SL-
	Fusobacterium gonidiaformans	257465914	L--K-Y--YAVL- E QYQ--Q-SVE--K--
	Fusobacterium nucleatum	34763539	L--K-KD--KRFL- E KFEK-I--ANH--DL-
Other species 2/250	Opitutaceae bacterium	225163584	--K-RD--FK-- R EL--S-T--AV--M-
	Thermomicrobium roseum	221632747	L-FK-----RL-- A E-----L--VR-----
	Sphaerobacter thermophilus	269837330	L-FK-P----RL-- E-----IR-----
	Bifidobacterium bifidum	224282597	Y- IK----REL-K -H-D-CD--E-----
	Elusimicrobium minutum	187250629	L-FK--Q--AEL-S YI--VS--EV--K-
	Kribbella flavida	284030835	--VKT-A--REVWK -L--CD--L--Q-
	Eubacterium limosum	310828648	--K-PD--SRL-I -I-----N-----
	Clostridium perfringens	110803674	--K-P--DM-D YI-----E-
	Peptostreptococcus stomatis	307243299	--IK-R--EDL-A FS--V-----
	Desulfotomaculum acetoxidans	258514118	--K-P--AEL-P EL-Q-Q--VS--DK-
	Symbiobacterium thermophilum	51891797	L-VK----YWR-- H--G----S-A----
	Bacteroides capillosus	154499865	--IK-E--A-L-P -W--I--VVRV--K-
	Geobacter sulfurreducens	39996502	--VKTP--AAA-H YA--IS--V-----
	Desulfovibrio desulfuricans	220904502	L-YK-I--EKS-- H-----A--AR--A-
	Spirochaeta thermophila	307718499	--K-P--ARL-E -I-----VR-----
	Borrelia burgdorferi	15594924	--IK-Q--CEL-N -L-----VR--K-
	Dictyoglomus turgidum	217967773	--FK-PD--INL-K -I-----F-----
Selenomonas noxia	292670795	Y--K-EQ--AEL-A -Y-G-----V-----	
Leptospirillum rubarum	124514840	--KTPR--VRS-E EL--I--R--D-I	

◀ **Fig. 9** Examples of CSIs that are commonly shared by Synergistetes species and other groups of bacteria. **a** A CSI consisting of 1 aa insert in the MiaB-family of RNA modification enzyme that is commonly shared by different Synergistetes and Chloroflexi. **b** A 1 aa insert in a conserved region of DNA polymerase III, α subunit shared by all Synergistetes and Fusobacteria as well as two other bacteria belonging to the Chloroflexi and Verrucomicrobia phyla

Synergistetes phyla might be closely related to each other based on their cell membrane structure and shared indels in their DnaK and GroEL proteins (Geissinger et al. 2009; Gupta 2011). Though the *Elusimicrobia* and *Fusobacteria* share some CSIs with the Synergistetes species, no CSI was found that was

specifically shared by all detected Synergistetes and species from these taxa. Furthermore, the branching of these phyla in the protein trees (Fig. 1) does not support their close relationship with the Synergistetes. Hence, based upon these results, at present no clear relationship of the Synergistetes species to other bacteria phyla can be inferred. These results provide further evidence supporting the placement of Synergistetes species into a distinct phylum.

Due to their specificity, Synergistetes-specific CSIs provide interesting prospects for future research. Since these CSIs are present in conserved regions of various proteins, degenerate primers utilizing the conserved regions can be designed for use as a means for

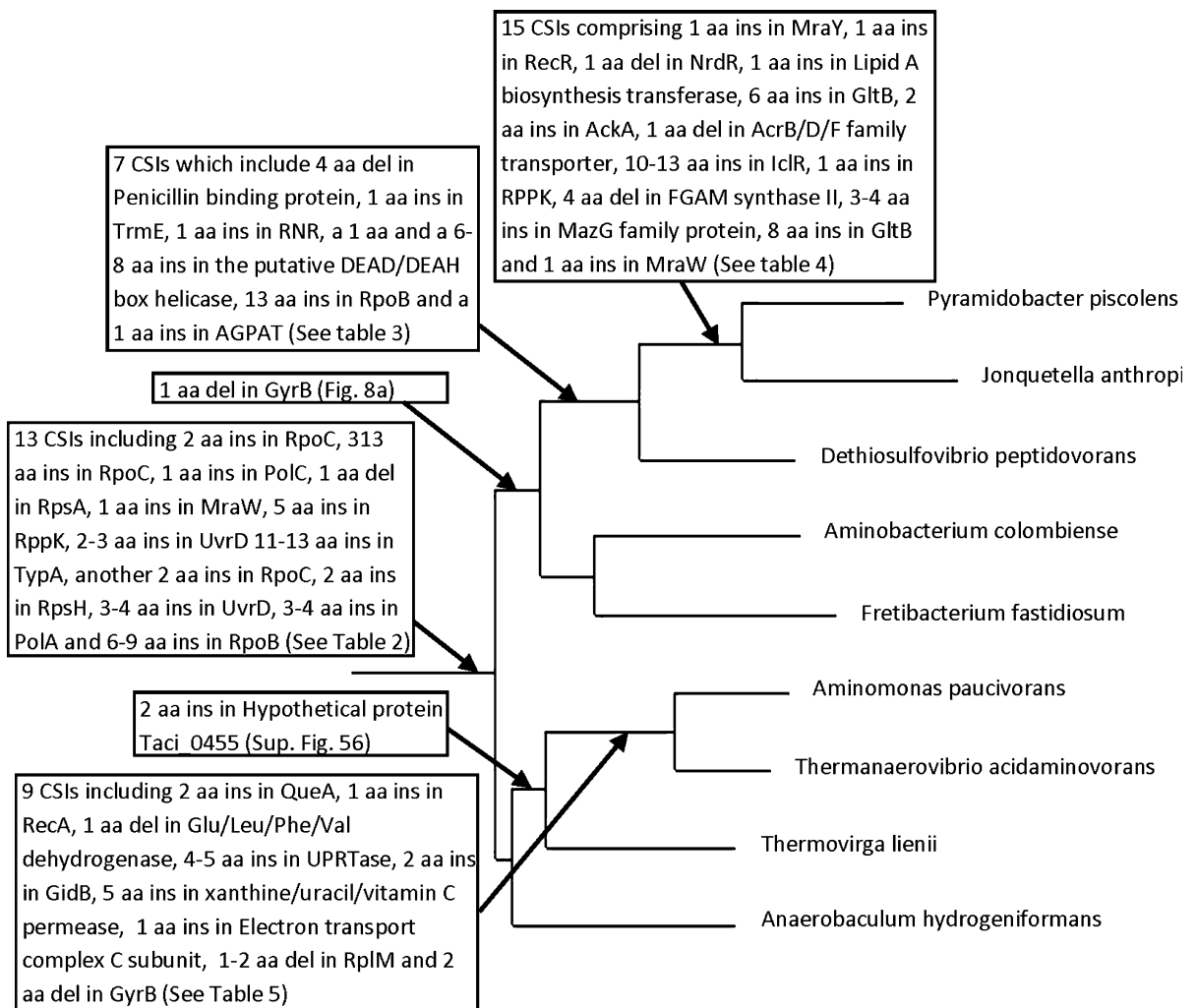


Fig. 10 A summary diagram portraying the species distribution of various identified Synergistetes-specific CSIs and the evolutionary stages where the genetic changes responsible for them have occurred

identification of various species of the phylum in different environments (Gao and Gupta 2005). This might prove to be especially useful, as it is surmised that universal primers utilized in detection of organisms in metagenomic studies may not efficiently identify some Synergistetes species (Hamady and Knight 2009). As molecular markers, the phylum-specific CSIs can be useful as identification tool for detection of known and unknown species in metagenomics experiments. These CSI can also assist in the classification of newly discovered bacteria into the phylum Synergistetes and its sub-groups.

Finally, some species of Synergistetes have also been notoriously difficult to culture/isolate (Vartoukian et al. 2010) and, for others, their biological nuances have just begun to be understood. It has been suggested that Synergistetes act in concert with other oral bacteria to degrade proteinaceous compounds in periodontitis lesions (Homer and Beighton 1992; Wei et al. 1999; Vartoukian et al. 2007). Prior functional studies on taxa-specific CSIs have shown that such indels are usually present in peripheral regions of proteins and they tend to be essential for the function of the proteins in the organisms where the CSIs occur (Itzhaki et al. 2006; Akiva et al. 2008; Hormozdiari et al. 2009; Singh and Gupta 2009). Hence, agents that bind to these CSIs and inhibit their cellular functions could provide novel therapeutics, which are specifically directed against this group of bacteria. Lastly, the molecular markers discovered in this study, due to their specificity for Synergistetes species provide novel and valuable means for understanding the contribution of this group of bacteria to the environment and to the microbial communities that they inhabit. Thus, analyses devoted to the understanding of the function of these CSIs should provide important insights into the biochemical and physiological properties that define the Synergistetes and their roles in different environments.

Acknowledgments This work was supported by a research grant from the Natural Science and Engineering Research Council of Canada.

References

- Akiva E, Itzhaki Z, Margalit H (2008) Built-in loops allow versatility in domain–domain interactions: lessons from self-interacting domains. *Proc Natl Acad Sci USA* 105:13292–13297
- Allison MJ, Maynerry WR, McSweeney CS, Stahl DA (1992) *Synergistes jonesii*, gen.nov., sp.nov.: a rumen bacterium that degrades toxic pyridinediols. *Syst Appl Microbiol* 15:522–529
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, Yu YK (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272:5101–5109
- Baena S, Fardeau ML, Labat M, Ollivier B, Thomas P, Garcia JL, Patel BK (1998) *Aminobacterium colombiense* gen. nov. sp. nov., an amino acid-degrading anaerobe isolated from anaerobic sludge. *Anaerobe* 4:241–250
- Baena S, Fardeau ML, Ollivier B, Labat M, Thomas P, Garcia JL, Patel BK (1999a) *Aminomonas paucivorans* gen. nov., sp. nov., a mesophilic, anaerobic, amino-acid-utilizing bacterium. *Int J Syst Bacteriol* 49(Pt 3):975–982
- Baena S, Fardeau ML, Woo TH, Ollivier B, Labat M, Patel BK (1999b) Phylogenetic relationships of three amino-acid-utilizing anaerobes, *Selenomonas acidaminovorans*, ‘*Selenomonas acidaminophila*’ and *Eubacterium acidaminophilum*, as inferred from partial 16S rDNA nucleotide sequences and proposal of *Thermanaerovibrio acidaminovorans* gen. nov., comb. nov. and *Anaeromusa acidaminophila* gen. nov., comb. nov. *Int J Syst Bacteriol* 49(Pt 3):969–974
- Baldauf SL, Palmer JD (1993) Animals and fungi are each other’s closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci USA* 90:11558–11562
- Bocchetta M, Gribaldo S, Sanangelantoni A, Cammarano P (2000) Phylogenetic depth of the bacterial genera Aquifex and Thermotoga inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J Mol Evol* 50:366–380
- Chertkov O, Sikorski J, Brambilla E, Lapidus A, Copeland A, Glavina DR, Nolan M, Lucas S, Tice H, Cheng JF, Han C, Detter JC, Bruce D, Tapia R, Goodwin L, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Spring S, Rohde M, Goker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP (2010) Complete genome sequence of *Aminobacterium colombiense* type strain (ALA-1). *Stand Genomic Sci* 2:280–289
- Chovatia M, Sikorski J, Schroder M, Lapidus A, Nolan M, Tice H, Glavina DR, Copeland A, Cheng JF, Lucas S, Chen F, Bruce D, Goodwin L, Pitluck S, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Chain P, Saunders E, Detter JC, Brettin T, Rohde M, Goker M, Spring S, Bristow J, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Eisen JA (2009) Complete genome sequence of *Thermanaerovibrio acidaminovorans* type strain (Su883). *Stand Genomic Sci* 1:254–261
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287

- Dahle H, Birkeland NK (2006) *Thermovirga lienii* gen. nov., sp. nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int J Syst Evol Microbiol* 56:1539–1545
- de Lillo A, Ashley FP, Palmer RM, Munson MA, Kyriacou L, Weightman AJ, Wade WG (2006) Novel subgingival bacterial phylotypes detected using multiple universal polymerase chain reaction primer sets. *Oral Microbiol Immunol* 21:61–68
- Delbes C, Moletta R, Godon J (2001) Bacterial and archaeal 16S rDNA and 16S rRNA dynamics during an acetate crisis in an anaerobic digester ecosystem. *FEMS Microbiol Ecol* 35:19–26
- Diaz C, Baena S, Fardeau ML, Patel BK (2007) *Aminiphilus circumscriptus* gen. nov., sp. nov., an anaerobic amino-acid-degrading bacterium from an upflow anaerobic sludge reactor. *Int J Syst Evol Microbiol* 57:1914–1918
- Downes J, Vartoukian SR, Dewhirst FE, Izard J, Chen T, Yu WH, Sutcliffe IC, Wade WG (2009) *Pyramidobacter piscolens* gen. nov., sp. nov., a member of the phylum ‘Synergistetes’ isolated from the human oral cavity. *Int J Syst Evol Microbiol* 59:972–980
- Fonknechten N, Chaussonnerie S, Tricot S, Lajus A, Andreesen JR, Perchat N, Pelletier E, Gouyvenoux M, Barbe V, Salanoubat M, Le Paslier D, Weissenbach J, Cohen GN, Kreimeyer A (2010) *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomics* 11:555
- Gao B, Gupta RS (2005) Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int J Syst Evol Microbiol* 55:2401–2412
- Garrity GM, Bell JA, Lilburn TG (2004) Taxonomic outline of the Prokaryotes. *Bergey’s manual of systematic bacteriology*. Release 5.0, 2nd edn. Springer, New York
- Geissinger O, Herlemann DP, Morschel E, Maier UG, Brune A (2009) The ultramicrobacterium “*Elusimicrobium minutum*” gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum. *Appl Environ Microbiol* 75:2831–2840
- Godon JJ, Moriniere J, Moletta M, Gaillac M, Bru V, Delgenes JP (2005) Rarity associated with specific ecological niches in the bacterial world: the ‘Synergistes’ example. *Environ Microbiol* 7:213–224
- Griffiths E, Gupta RS (2006) Molecular signatures in protein sequences that are characteristics of the phylum Aquificae. *Int J Syst Evol Microbiol* 56:99–107
- Griffiths E, Petrich AK, Gupta RS (2005) Conserved indels in essential proteins that are distinctive characteristics of Chlamydiales and provide novel means for their identification. *Microbiology* 151:2647–2657
- Guangsheng C, Plugge CM, Roelofsen W, Houwen FP, Stams AJM (1992) *Selenomonas acidaminovorans* sp. nov., a versatile thermophilic proton-reducing anaerobe able to grow by decarboxylation of succinate to propionate. *Arch Microbiol* 157:169–175
- Gupta RS (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 62:1435–1491
- Gupta RS (2000) The natural evolutionary relationships among prokaryotes. *Crit Rev Microbiol* 26:111–131
- Gupta RS (2001) The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *Int Microbiol* 4:187–202
- Gupta RS (2009) Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. *Int J Syst Evol Microbiol* 59:2510–2526
- Gupta RS (2011) Origin of diderm (gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek* 100:171–182
- Gupta RS, Bhandari V (2011) Phylogeny and molecular signatures for the phylum Thermotogae and its subgroups. *Antonie Van Leeuwenhoek* 100:1–34
- Gupta RS, Shami A (2011) Molecular signatures for the Crenarchaeota and the Thaumarchaeota. *Antonie Van Leeuwenhoek* 99:133–157
- Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* 19:1141–1152
- Herlemann DP, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A (2009) Genomic analysis of “*Elusimicrobium minutum*,” the first cultivated representative of the phylum “Elusimicrobia” (formerly termite group 1). *Appl Environ Microbiol* 75:2841–2849
- Homer KA, Beighton D (1992) Synergistic degradation of bovine serum albumin by mutans streptococci and other dental plaque bacteria. *FEMS Microbiol Lett* 69:259–262
- Hongoh Y, Sato T, Dolan MF, Noda S, Ui S, Kudo T, Ohkuma M (2007) The motility symbiont of the termite gut flagellate *Caduceia versatilis* is a member of the “Synergistes” group. *Appl Environ Microbiol* 73:6270–6276
- Hormozdiari F, Salari R, Hsing M, Schonhuth A, Chan SK, Sahinalp SC, Cherkasov A (2009) The effect of insertions and deletions on wirings in protein–protein interaction networks: a large-scale study. *J Comput Biol* 16:159–167
- Horz HP, Citron DM, Warren YA, Goldstein EJ, Conrads G (2006) Synergistes group organisms of human origin. *J Clin Microbiol* 44:2914–2920
- Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, Kawarabayasi Y, Donachie SP, Pikina A, Galperin MY, Koonin EV, Makarova KS, Omelchenko MV, Sorokin A, Wolf YI, Li QX, Keum YS, Campbell S, Denery J, Aizawa S, Shibata S, Malahoff A, Alam M (2004) Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* 101:18036–18041
- Hugenholtz P, Hooper SD, Kyrpides NC (2009) Focus: Synergistetes. *Environ Microbiol* 11:1327–1329
- Itzhaki Z, Akiva E, Altuvia Y, Margalit H (2006) Evolutionary conservation of domain–domain interactions. *Genome Biol* 7:R125
- Jumas-Bilak E, Carlier JP, Jean-Pierre H, Citron D, Bernard K, Damay A, Gay B, Teyssier C, Campos J, Marchandin H (2007) *Jonquetella anthropi* gen. nov., sp. nov., the first member of the candidate phylum ‘Synergistetes’ isolated from man. *Int J Syst Evol Microbiol* 57:2743–2748
- Jumas-Bilak E, Roudiere L, Marchandin H (2009) Description of ‘Synergistetes’ phyl. nov. and emended description of the phylum ‘Deferribacteres’ and of the family

- Syntrophomonadaceae, phylum ‘Firmicutes’. *Int J Syst Evol Microbiol* 59:1028–1035
- Kumar PS, Griffen AL, Moeschberger ML, Leys EJ (2005) Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol* 43:3944–3955
- Labutti K, Mayilraj S, Clum A, Lucas S, Glavina DR, Nolan M, Tice H, Cheng JF, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Mikhailova N, Pati A, Goodwin L, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Rohde M, Spring S, Goker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Lapidus A (2010) Permanent draft genome sequence of *Dethiosulfovibrio peptidovorans* type strain (SEBR 4207). *Stand Genomic Sci* 3:85–92
- Magot M, Ravot G, Campagnolle X, Ollivier B, Patel BK, Fardeau ML, Thomas P, Crolet JL, Garcia JL (1997) *Dethiosulfovibrio peptidovorans* gen. nov., sp. nov., a new anaerobic, slightly halophilic, thiosulfate-reducing bacterium from corroding offshore oil wells. *Int J Syst Bacteriol* 47:818–824
- Marchandin H, Damay A, Roudiere L, Teyssier C, Zorogniotti I, Dechaud H, Jean-Pierre H, Jumas-Bilak E (2010) Phylogeny, diversity and host specialization in the phylum Synergistetes with emphasis on strains and clones of human origin. *Res Microbiol* 161:91–100
- Maune MW, Tanner RS (2012) Description of *Anaerobaculum hydrogeniformans* sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus *Anaerobaculum*. *Int J Syst Evol Microbiol* 62: 832–838
- McSweeney CS, Mackie RI, Odenyo AA, Stahl DA (1993) Development of an oligonucleotide probe targeting 16S rRNA and its application for detection and quantitation of the ruminal bacterium *Synergistes jonesii* in a mixed-population chemostat. *Appl Environ Microbiol* 59:1607–1612
- Naushad HS, Gupta RS (2012) Molecular signatures (conserved indels) in protein sequences that are specific for the order Pasteurellales and distinguish two of its main clades. *Antonie Van Leeuwenhoek* 101:105–124
- NCBI genomic database (2012) <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>. Accessed 5 May 2012
- NCBI Taxonomy (2012) <http://www.ncbi.nlm.nih.gov/taxonomy>. Accessed 5 May 2012
- Pitluck S, Yasawong M, Held B, Lapidus A, Nolan M, Copeland A, Lucas S, Del Rio TG, Tice H, Cheng JF, Chertkov O, Goodwin L, Tapia R, Han C, Liolios K, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Pukall R, Spring S, Rohde M, Sikorski J, Goker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP (2010) Non-contiguous finished genome sequence of *Aminomonas paucivorans* type strain (GLU-3). *Stand Genomic Sci* 3:285–293
- Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257:74–76
- Riviere D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach J, Li T, Camacho P, Sghir A (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J* 3:700–714
- Rokas A, Holland PW (2000) Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* 15:454–459
- Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804
- Singh B, Gupta RS (2009) Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol Genet Genomics* 281:361–373
- Soria-Carrasco V, Valens-Vadell M, Pena A, Anton J, Amann R, Castresana J, Rossello-Mora R (2007) Phylogenetic position of *Salinibacter ruber* based on concatenated protein alignments. *Syst Appl Microbiol* 30:171–179
- Surkov AV, Dubinina GA, Lysenko AM, Glockner FO, Kuever J (2001) *Dethiosulfovibrio russensis* sp. nov., *Dethiosulfovibrio marinus* sp. nov. and *Dethiosulfovibrio acidaminovorans* sp. nov., novel anaerobic, thiosulfate- and sulfur-reducing bacteria isolated from ‘Thiodendron’ sulfur mats in different saline environments. *Int J Syst Evol Microbiol* 51:327–337
- Sutcliffe IC (2010) A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol* 18:464–470
- Vartoukian SR, Palmer RM, Wade WG (2007) The division ‘Synergistes’. *Anaerobe* 13:99–106
- Vartoukian SR, Palmer RM, Wade WG (2010) Cultivation of a Synergistetes strain representing a previously uncultivated lineage. *Environ Microbiol* 12:916–928
- Vartoukian S, Downes J, Palmer RM, Wade WG (2012) *Fretibacterium fastidiosum* gen. nov., sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol*. doi:10.1099/ijs.0.041038-0
- Wei GX, van der Hoeven JS, Smalley JW, Mikx FH, Fan MW (1999) Proteolysis and utilization of albumin by enrichment cultures of subgingival microbiota. *Oral Microbiol Immunol* 14:348–351
- Wu M, Eisen JA (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 9:R151
- Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM, DeBoy RT, Nelson KE, Nesbo CL, Doolittle WF, Gogarten JP, Noll KM (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc Natl Acad Sci USA* 106:5865–5870
- Ziganshin AM, Schmidt T, Scholwin F, Il’inskaya ON, Harms H, Kleinstaub S (2011) Bacteria and archaea involved in anaerobic digestion of distillers grains with solubles. *Appl Microbiol Biotechnol* 89:2039–2052
- Zijngje V, van Leeuwen MB, Degener JE, Abbas F, Thurnheer T, Gmur R, Harmsen HJ (2010) Oral biofilm architecture on natural teeth. *PLoS ONE* 5:e9321