Supplemental File 1A

Proteins used for Construction of Phylogenetic Tree

Protein	Length (aa)[#]	Protein	Length (aa)[#]
GTP-binding protein	362	30S ribosomal protein S5	201
Dimethyladenosine transferase (KsgA)	286	Ribosomal protein S13	126
30S ribosomal protein S7	156	50S ribosomal protein L13P	147
Elongation factor Tu	397	30S ribosomal protein S9	170
50S ribosomal protein L11	144	30S ribosomal protein S15	95
50S ribosomal protein L1	241	Ribosomal protein S19	93
RNA polymerase ß subunit (RpoB)	1161	Gyrase B	686
RNA polymerase ß' subunit (RpoC)	1299	Gyrase A	857
50S ribosomal protein L3	214	50S ribosomal protein L10/L16	176
Ribosomal protein L4	219	Preprotein translocase subunit SecY	437
Ribosomal protein L2	278	RNA polymerase subunit alpha	340
50S ribosomal protein L22	125	Threonyl-tRNA-synthetase	658
30S ribosomal protein S3	277	Molecular chaperone DnaK (Hsp70)	618
50S ribosomal protein L14	122	Chaperonin GroEL (Hsp60)	541
Ribosomal protein L5	185	O-sialoglycoprotein endopeptidase	374
30S ribosomal protein S8	132	Phosphatidate cytidylyltransferase	391
50S ribosomal protein L6	179	CDP-diglyceride synthase (CdsA)	391
		DNA polymerase III subunit delta (holB)	401

The phylogenetic trees for the actinobacterial species (Fig. 2) based upon concatenated sequence alignment of the above proteins were constructed as described in our earlier work (2, 6).

[#] Protein length is from *S. coelicolor* A3(2).

Supplemental File 1B

Identification of CSIs and CSPs that are Specific for *Actinobacteria* or its various Clades

To identify CSIs that are specific for *Actinobacteria*, Blastp searches were carried out on various proteins (or open reading frames) from the genome of *Corynebacterium glutamicum* ATCC 13032 to retrieve high scoring homologs for 10-15 divergent actinobacterial species as well as 5-6 sequences from other phyla of *Bacteria*. Multiple sequence alignments of these proteins were created using the ClustalX 1.83 program (8) and these alignments were visually inspected to identify indels in conserved regions (i.e. flanking 30-40 aa on either side contain minimally 4-5 identical residues) that were restricted to either some or all of the actinobacterial species in the sequence alignments (1). The indels that were not flanked by conserved regions were not further considered, as they do not provide reliable molecular markers (4-7). The species distribution of all potentially useful indels was further evaluated by detailed Blastp searches (minimally 250 top hits) on short sequence segments containing the indels and their flanking conserved regions (between 60-150 aa depending upon the lengths of the indels) to determine the species-specificity of the indels. In cases where the indels were large, two separate Blastp searches were performed, one with a sequence containing the indel and the other with a sequence lacking the indel, to determine the presence or absence of the indel in different organisms. The sequence information for various CSIs from representative *Actinobacteria*-species that contain the CSIs and some outgroup species were compiled and formatted into signature files using certain in house programs.

Although sequence information for only a limited number of species are shown in different alignments, unless otherwise indicated these CSIs are specific for the indicated groups/clades. The inference whether a given CSI is an insert or a deletion is based upon the presence or absence of the CSIs in other groups or phyla of *Bacteria* (*Actinobacteria*). If the CSI in question is lacking in the outgroup species, then the absence of indel is the ancestral character state of the protein and the indel in this case is an insert in the indicated group of bacteria. On the other hand, if the outgroup species contained the sequence region (CSI) that is missing in a given clade of *Actinobacteria*, then the indel in question is interpreted as a deletion.

To identify CSPs those are specific for either all *Actinobacteria* or several of its subclades, Blastp searches were carried out on each ORF from the genomes of *Bifidobacterium longum* NCC2705, B. dentium ATCC 27678, *Corynebacterium glutamicum* ATCC 13032, *Frankia sp. CcI3*, *Mycobacterium leprae* TN, *Mycobacterium avium* 104, *Rhodococcus jostii* RHA1 and *Streptomyces coelicolor* A3(2). These searches were performed against all organisms (i.e. non-redundant (nr) database) using the default parameters, without the low complexity filter (3). The proteins that were of interest were those for whom either all observed blast hits were from the indicated groups of *Actinobacteria*, or those where a large increase in E values is observed from the last hit belonging to *Actinobacteria* (or its specific clade) and the first hit from any other bacteria/actinobacteria and the E values for the latter hits are $>1e^{-03}$, indicating weak similarity that could occur by chance. However, higher E values could be significant for smaller proteins as the magnitude of the E value depends upon the length of the query sequence (3). Hence, the lengths of the query proteins and those of various hits were also taken into consideration when analyzing the results of these studies. In most cases, the lengths of various significant hits were very similar to those of the query proteins.

Reference List

1. **Gao, B. and R. S. Gupta**. 2005. Conserved indels in protein sequences that are characteristic of the phylum *Actinobacteria*. Int.J.Syst.Evol.Microbiol. **55**:2401-2412.

2. **Gao, B., R. Mohan, and R. S. Gupta**. 2009. Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria*. Int.J.Syst.Evol.Microbiol. **59**:234-247.

3. **Gao, B., R. Parmanathan, and R. S. Gupta**. 2006. Signature proteins that are distinctive characteristics of *Actinobacteria* and their subgroups. Antonie van Leeuwenhoek **90**:69-91.

4. **Gupta, R. S.** 1997. Protein phylogenies and signature sequences: evolutionary relationships within prokaryotes and between prokaryotes and eukaryotes. Antonie van Leeuwenhoek **72**:49-61.

5. **Gupta, R. S.** 2000. The phylogeny of *Proteobacteria*: relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol.Rev. **24**:367-402.

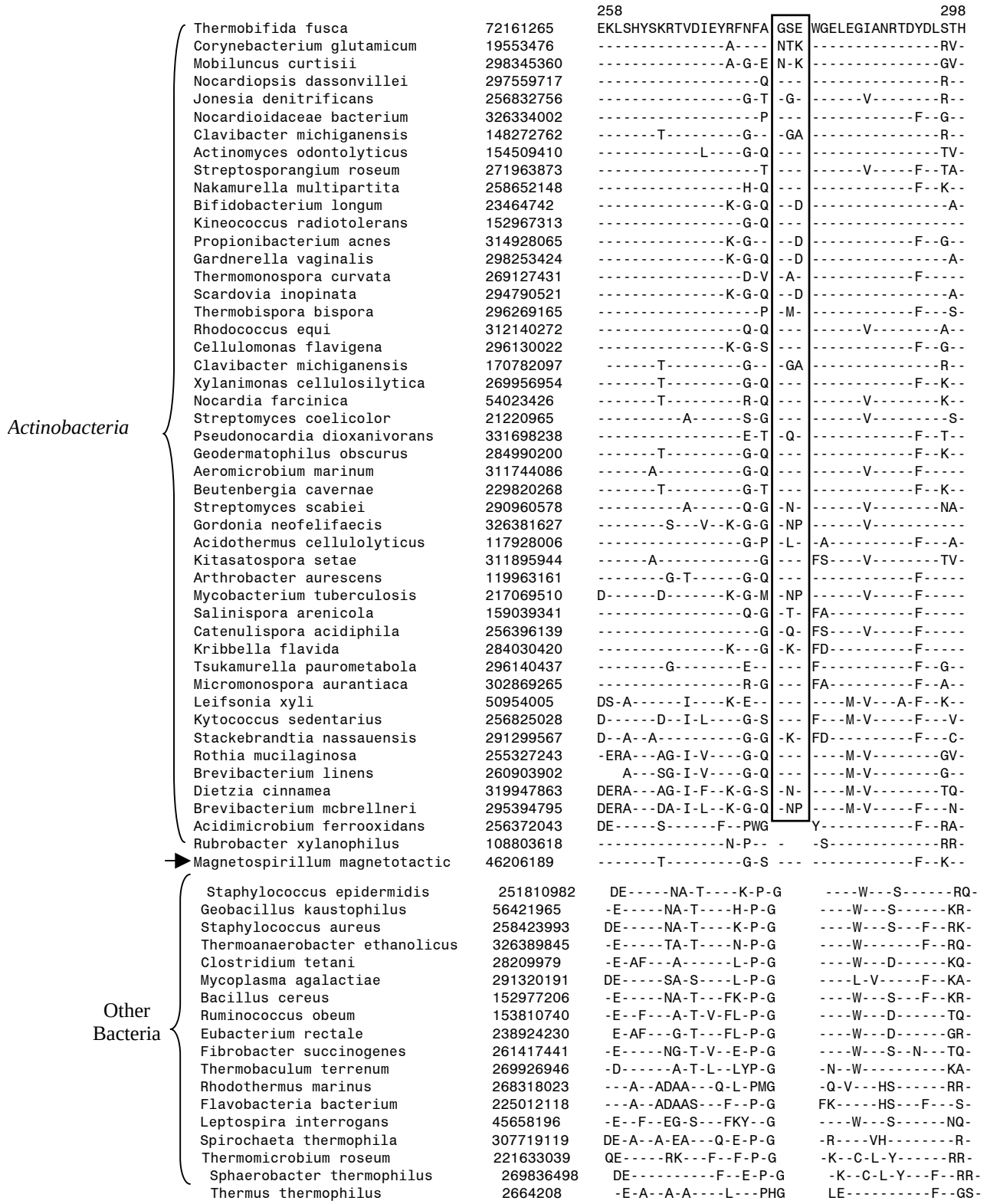
6. **Gupta, R. S.** 2009. Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. Int.J.Syst.Evol.Microbiol. **59**:2510-2526.

7. **Gupta, R. S. and A. Shami**. 2011. Molecular signatures for the *Crenarchaeota* and the *Thaumarchaeota*. Antonie van Leeuwenhoek **99**:133-157.

8. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson**. 1998. Multiple sequence alignment with Clustal x. Trends Biochem.Sci. **23**:403-405.

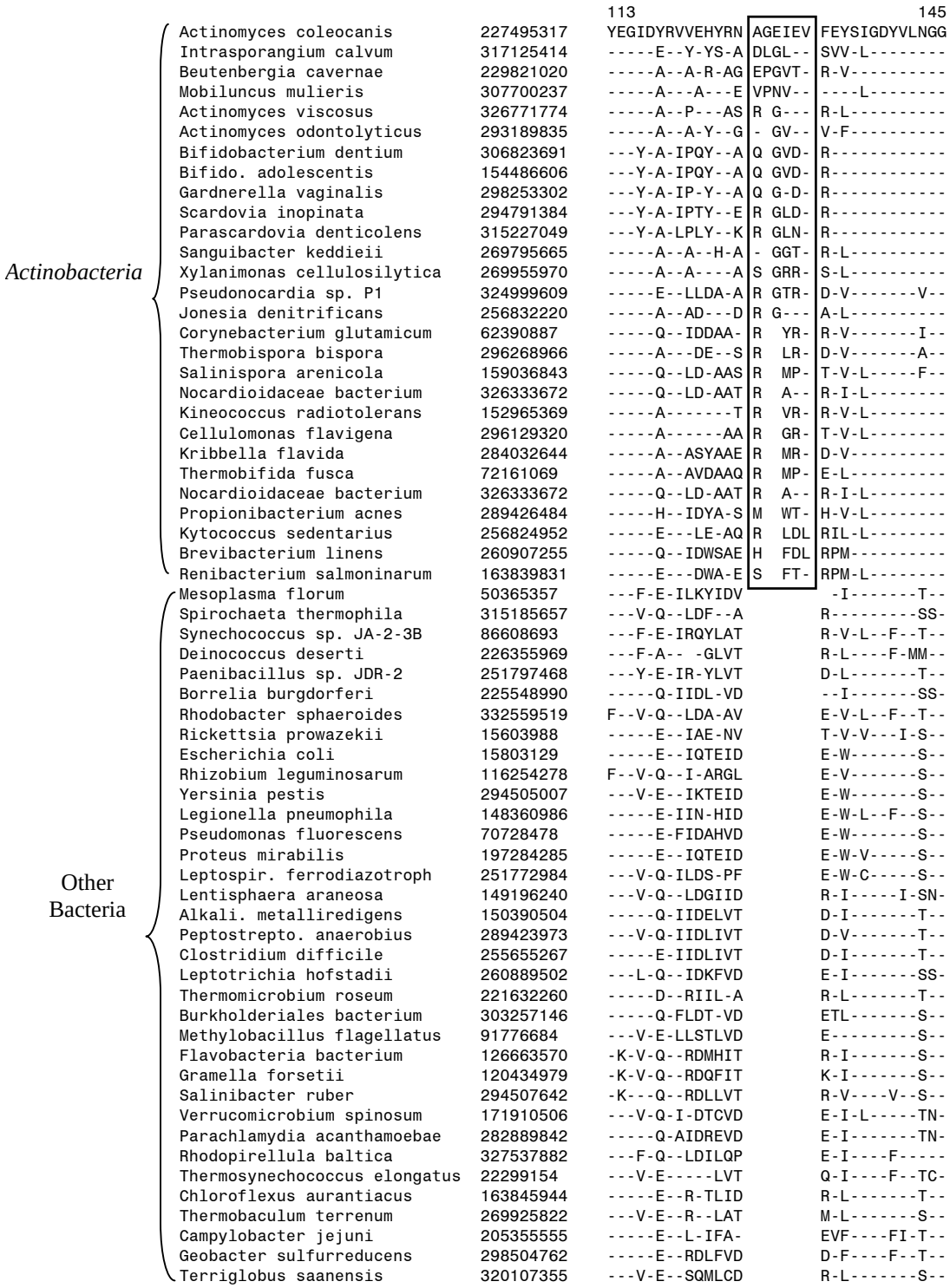
Supplemental File 2

A conserved insert in the Glycyl-tRNA synthetase that is largely specific for *Actinobacteria*. Sequence information for only limited numbers of actinobacteria is shown. The information regarding the distribution of this CSI in different genera of *Actinobacteria* is provided in the Supplemental File 7. Besides actinobacteria, this insert is also present in *M. magentosprillium.*



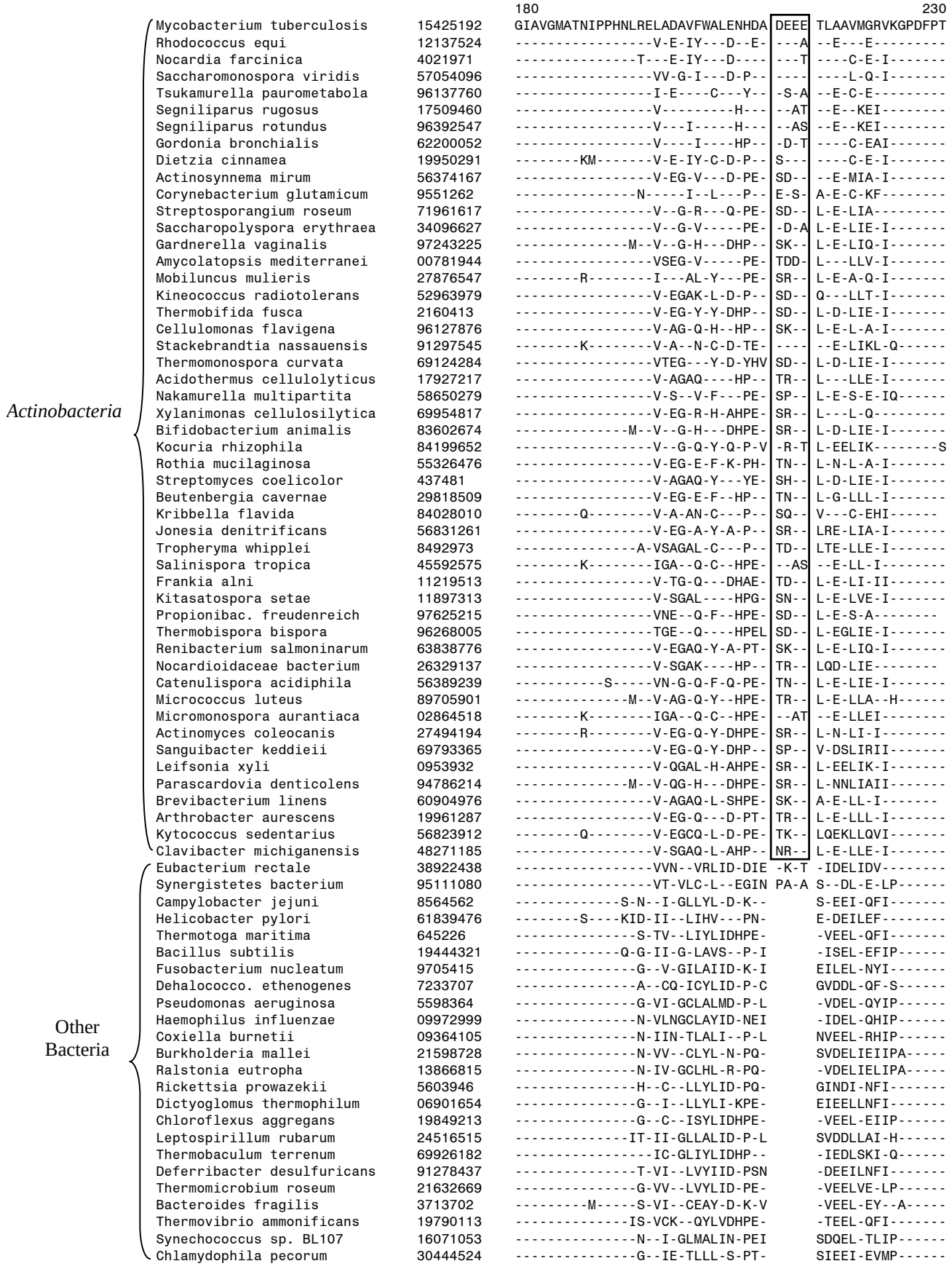
Supplemental File 3

A conserved insert in the tRNA (Guanine-1)-methyltransferase (TrmD) protein that is largely specific for *Actinobacteria*. Sequence information for only limited numbers of *Actinobacteria* is shown. The information regarding distribution of this CSI in different *Actinobacteria genera* is provided in the Supplemental File 7.



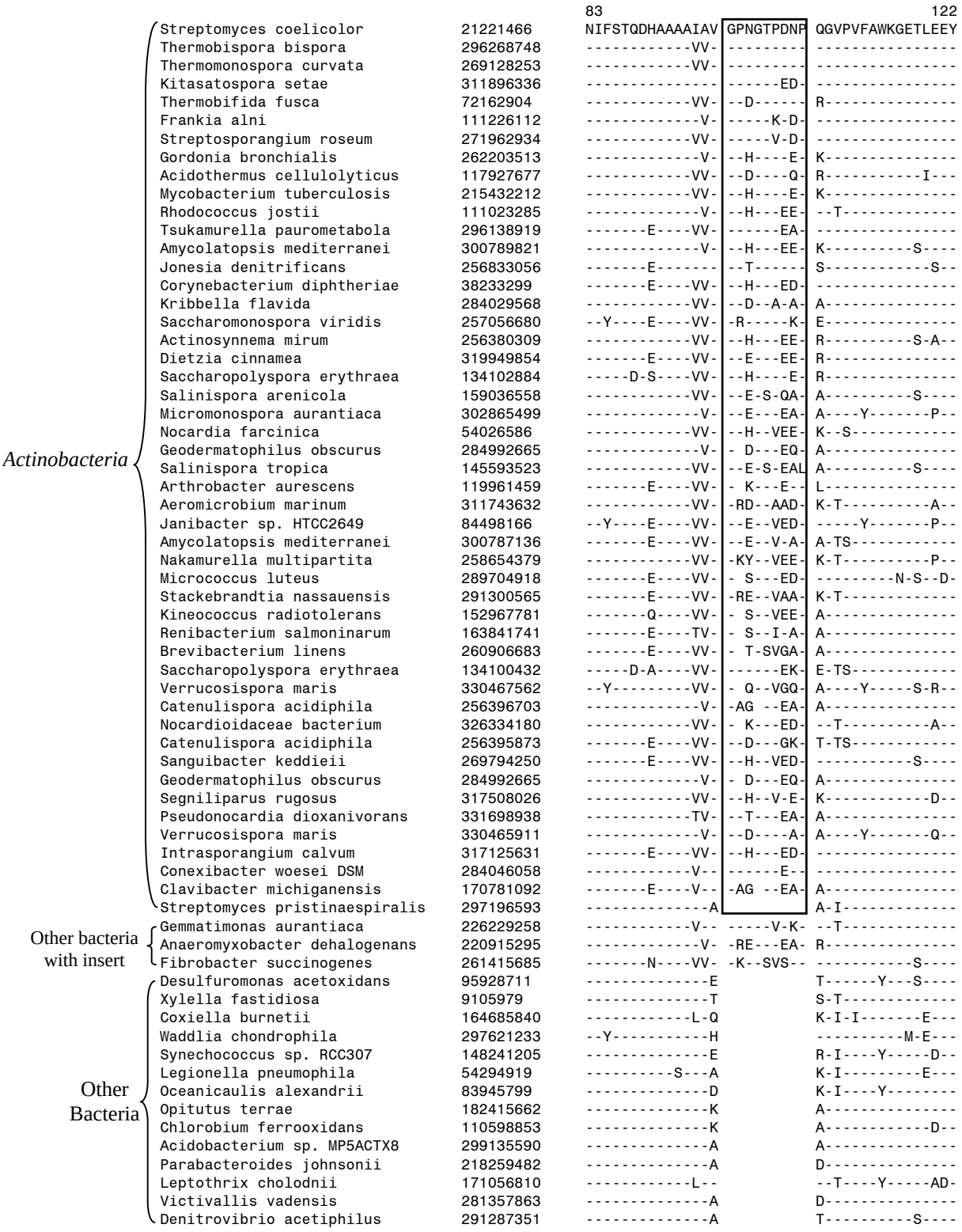
Supplemental File 4

A conserved insert in the DNA Gyrase A protein that is largely specific for **Actinobacteria**. Sequence information for only limited numbers of actinobacteria is shown. The information regarding the distribution of this CSI in other *Actinobacteria* genera is provided in the Supplemental File 7.



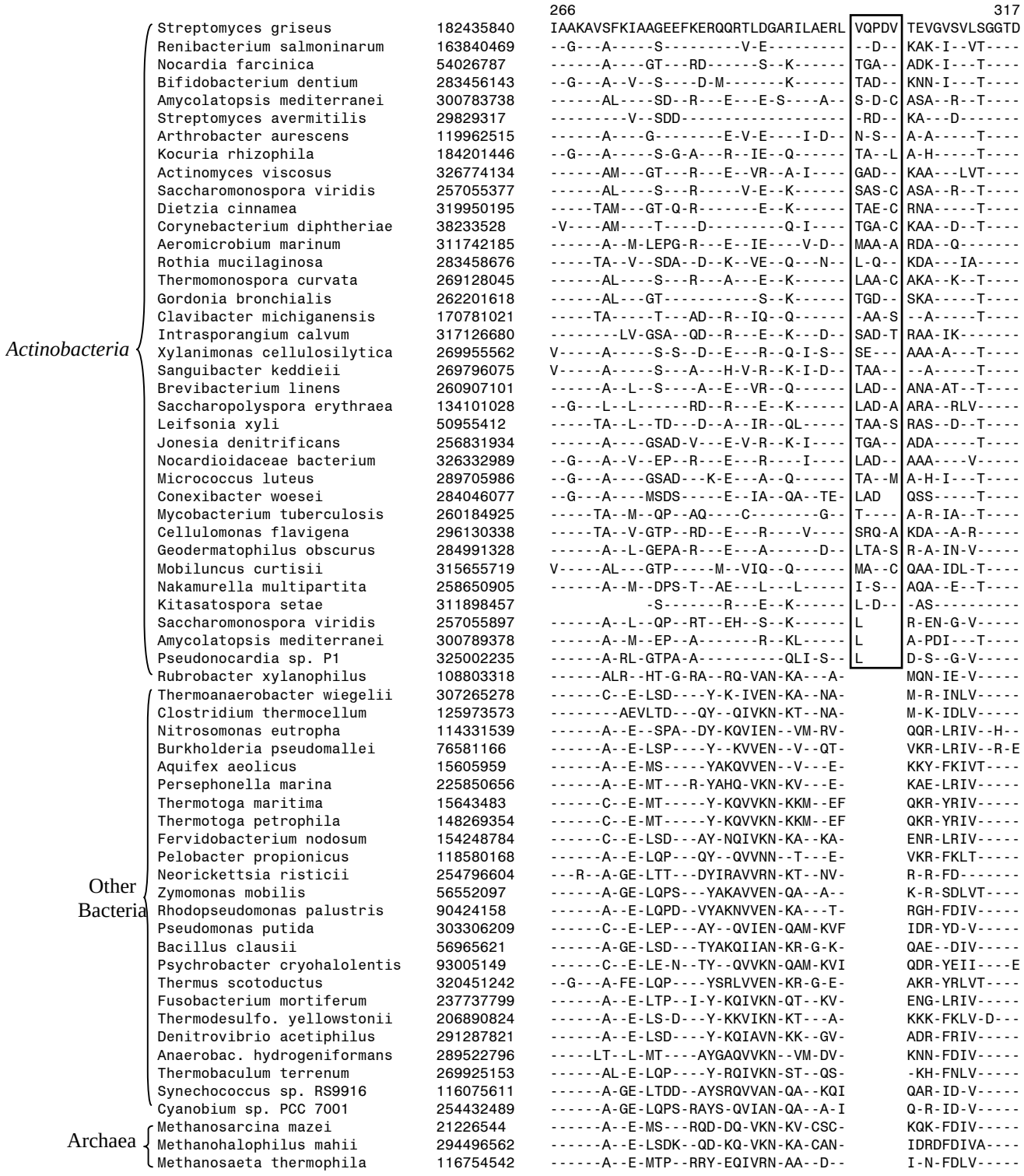
Supplemental File 5

A conserved insert in S-adenosylhomocysteine hydrolase protein that is largely specific for *Actinobacteria*. Sequence information for only limited numbers of actinobacteria is shown. The information regarding distribution of this CSI in other *Actinobacteria* genera is provided in the Supplemental File 7.



Supplemental File 6

A conserved insert in the enzyme Serine hydroxymethyl transferase that is largely specific for *Actinobacteria*. Sequence information for only limited numbers of actinobacteria is shown. The information regarding distribution of this CSI in other *Actinobacteria* genera is provided in the Supplemental File 7.



Supplemental File 7: Presence or absence of various *Actinobacteria*-specific CSPs in different genera

Genus	ML0642	ML1009	ML1029	ML1306	ML0760	ML0804	ML0857	ML0869	ML1016	ML1026	ML2137	ML2204	ML0013
Mycobacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Tsukamurella	+	+	+	+	+	+	+	+	+	+	+	+	+
Gordonia	+	+	+	+	+	+	+	+	+	+	+	-	+
Nocardia	+	+	+	+	+	+	+	+	+	+	+	-	+
Rhodococcus	+	+	+	+	+	+	+	+	+	+	+	+	+
Segniliparus	+	+	+	+	+	+	+	+	+	+	+	+	+
Amycolicicoccus	+	+	+	+	+	+	+	+	+	+	+	+	+
Corynebacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Nakamurella	+	+	-	+	+	+	+	+	+	+	+	+	+
Pseudonocardia	+	+	+	+	+	+	+	+	+	+	+	+	+
Saccharopolyspora	+	+	+	+	+	+	+	+	+	+	+	-	+
Actinosynnema	+	+	+	+	+	+	+	+	+	+	+	+	+
Saccharomonospora	+	+	+	+	+	+	+	+	+	+	+	+	+
Amycolatopsis	+	+	+	+	+	+	+	+	+	+	+	+	+
Geodermatophilus	+	+	+	+	+	+	+	+	+	+	+	+	+
Stackebrandtia	+	+	+	+	+	+	+	+	+	+	+	+	+
Verrucosispora	+	+	+	+	+	-	+	+	+	-	+	-	+
Micromonospora	+	+	+	+	+	+	+	+	+	+	+	+	+
Salinispora	+	+	+	+	+	+	+	+	+	+	+	+	+
Frankia	+	+	+	+	+	+	+	+	+	+	+	+	+
Acidothermus	+	+	+	+	+	+	+	+	+	+	+	+	+
Streptosporangium	+	+	+	+	+	+	+	+	+	+	+	+	+
Thermomonospora	+	+	+	+	+	+	+	+	+	+	+	-	+
Thermobifida	+	+	+	+	+	+	+	+	+	+	+	+	-
Nocardiopsis	+	+	+	+	+	+	+	+	+	+	+	+	+
Catenulispora	+	+	+	+	+	+	+	+	+	+	+	+	+
Streptomyces	+	+	+	+	+	+	+	+	+	+	+	+	+
Propionibacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Kribbella	+	+	+	+	+	+	+	+	+	+	+	+	+
Nocardioides	+	+	+	+	+	+	+	+	+	+	+	+	+
Kineococcus	+	+	+	+	+	+	+	+	+	+	+	+	+
Janibacter	+	+	+	+	+	+	+	+	+	+	+	+	+
Kytococcus	+	+	+	+	+	+	+	+	+	+	+	+	+
Brachybacterium	+	+	+	+	+	+	+	+	+	+	+	+	-
Brevibacterium	+	+	+	+	+	+	+	+	+	-	+	+	+
Intrasporangium	+	+	+	+	+	+	+	+	+	+	+	+	+
Isoptericola	+	+	+	+	+	+	+	+	+	+	+	+	+
Microbacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Rothia	+	+	+	+	+	+	+	+	+	+	+	+	-
Kocuria	+	+	+	+	+	+	+	+	+	+	+	+	+
Micrococcus	+	+	+	+	+	+	+	+	+	+	+	+	+
Renibacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Arthrobacter	+	+	+	+	+	+	+	+	+	+	+	+	+
Beutenbergia	+	+	+	+	+	+	+	+	+	+	+	+	+
Cellulomonas	+	+	+	+	+	+	+	+	+	+	-	+	+
Xylanimonas	+	+	+	+	+	+	+	+	+	+	+	+	+
Jonesia	+	+	+	+	+	+	+	+	+	+	+	+	+
Sanguibacter	+	+	+	+	+	+	+	+	+	+	+	+	+
Mobiluncus	+	+	+	+	+	+	+	+	+	+	+	+	+
Actinomyces	+	+	+	+	+	+	+	+	+	+	+	+	+
Arcanobacterium	+	+	+	-	+	+	+	-	+	+	+	+	-
Gardnerella	+	-	+	-	+	+	+	+	+	+	+	+	+
Bifidobacterium	+	+	+	+	+	+	+	+	+	+	+	+	+
Tropheryma	+	+	+	+	+	+	+	+	+	+	+	+	+
Leifsonia	+	+	+	+	+	+	+	+	+	+	+	+	+
Clavibacter	+	+	+	+	+	+	+	+	+	+	+	+	+
*marine actino**	+	+	+	+	+	+	+	+	+	+	+	+	+
Acidimicrobium	-	+	-	+	+	+	-	-	-	-	-	-	-
Conexibacter	+	+	-	+	-	-	-	-	-	-	-	-	-
Rubrobacter	+	+	+	+	-	-	-	-	-	-	-	-	-
Olsenella	-	-	-	-	-	-	-	-	-	-	-	-	-
Slackia	-	-	-	-	-	-	-	-	-	-	-	-	-
Eggerthella	-	-	-	-	-	-	-	-	-	-	-	-	-
Cryptobacterium	-	-	-	-	-	-	-	-	-	-	-	-	-
Coriobacterium	-	-	-	-	-	-	-	-	-	-	-	-	-

Notes: The left column lists all currently genome sequenced actinobacterial genera. +: present; -: absent. *: marine actinobacterium PHSC20C1.

Supplemental File 7 (page 2)

Genus	ML0007	ML0580	ML0921	ML1439	ML1610	ML2207	ML0256	ML0775	ML0761	ML0814	ML1649	ML2142
Mycobacterium	+	+	+	+	+	+	+	+	+	+	+	+
Tsukamurella	+	+	+	+	+	+	+	+	+	+	+	+
Gordonia	+	+	+	+	+	+	+	+	+	+	+	+
Nocardia	+	+	+	+	+	+	+	+	+	+	+	+
Rhodococcus	+	+	+	+	+	+	+	+	+	+	+	+
Segniliparus	+	+	+	+	+	+	+	+	+	+	+	+
Amycolicicoccus	+	-	+	+	+	+	+	+	+	+	+	+
Corynebacterium	+	+	+	+	+	+	+	+	+	+	+	+
Nakamurella	+	+	+	+	+	+	+	+	+	+	+	+
Pseudonocardia	+	+	+	+	+	+	+	+	+	+	+	+
Saccharopolyspora	+	+	+	+	+	+	+	+	+	+	+	+
Actinosynnema	+	+	+	+	+	+	+	+	+	+	+	+
Saccharomonospora	+	+	+	+	+	+	+	+	+	+	+	+
Amycolatopsis	+	+	+	+	+	+	+	+	+	+	+	+
Geodermatophilus	+	+	+	+	+	+	+	+	+	+	+	+
Stackebrandtia	+	+	+	+	+	+	+	+	+	+	+	+
Verrucosispora	+	+	+	-	-	+	+	-	+	+	+	+
Micromonospora	+	+	+	+	-	+	+	-	+	+	+	+
Salinispora	+	+	+	+	+	+	+	+	+	+	+	+
Frankia	+	+	+	+	+	+	+	-	+	+	+	+
Acidothermus	+	+	+	+	+	+	+	+	+	+	+	+
Streptosporangium	+	+	+	+	+	+	+	+	+	+	+	+
Thermomonospora	+	+	+	-	+	+	+	+	+	+	+	+
Thermobifida	+	+	+	+	+	+	+	+	+	-	+	+
Nocardiopsis	+	+	+	+	+	+	+	+	+	-	+	+
Catenulispora	+	+	+	+	+	+	+	-	+	+	+	+
Streptomyces	+	+	+	+	+	+	+	+	+	+	+	+
Propionibacterium	+	+	+	+	+	+	+	+	+	+	+	+
Kribbella	+	+	+	+	+	+	+	+	+	+	+	+
Nocardioides	+	+	+	+	+	+	+	+	+	+	+	+
Kineococcus	+	+	+	+	+	+	+	+	+	+	+	+
Janibacter	+	+	+	+	+	+	+	+	+	+	+	+
Kytococcus	+	+	+	+	+	+	+	+	-	+	+	+
Brachybacterium	+	+	+	+	+	+	+	+	+	+	+	+
Brevibacterium	+	-	+	+	+	+	-	+	-	+	+	+
Intrasporangium	+	+	+	+	+	+	+	+	+	+	+	+
Isoptericola	+	+	+	+	+	+	+	+	+	+	+	+
Microbacterium	-	+	+	+	+	+	+	+	-	+	-	+
Rothia	+	+	+	+	+	+	+	+	-	+	-	+
Kocuria	+	+	+	+	-	+	+	+	-	+	+	+
Micrococcus	+	+	+	+	+	+	+	+	-	+	+	+
Renibacterium	+	+	+	+	+	+	+	+	+	+	+	+
Arthrobacter	+	+	+	+	+	+	+	+	+	+	+	+
Beutenbergia	+	+	+	+	+	+	+	+	+	+	+	+
Cellulomonas	+	+	+	+	+	+	+	+	+	+	+	+
Xylanimonas	+	+	+	+	+	+	+	+	+	+	+	+
Jonesia	+	+	+	+	+	+	+	+	+	+	+	+
Sanguibacter	+	+	+	+	+	+	+	+	+	+	+	+
Mobiluncus	+	-	+	+	+	+	+	+	+	+	+	+
Actinomyces	+	+	+	+	+	+	+	+	+	+	+	+
Arcanobacterium	+	-	+	+	+	+	+	+	+	-	+	+
Gardnerella	+	+	+	+	+	+	+	-	+	+	+	+
Bifidobacterium	+	+	+	+	+	+	+	+	+	+	+	+
Tropheryma	-	-	-	-	-	-	+	-	-	+	-	-
Leifsonia	+	+	+	+	+	+	+	+	-	+	-	+
Clavibacter	+	+	+	+	+	+	+	+	+	+	-	+
*marine actino**	+	+	+	+	+	+	-	+	+	+	-	+
Acidimicrobium	-	-	-	-	+	-	-	-	-	-	-	-
Conexibacter	-	-	-	-	-	-	-	-	-	-	-	-
Rubrobacter	-	-	-	-	-	-	-	-	-	-	-	-
Olsenella	-	-	-	-	-	-	-	-	-	-	-	-
Slackia	-	-	-	-	-	-	-	-	-	-	-	-
Eggerthella	-	-	-	-	-	-	-	-	-	-	-	-
Cryptobacterium	-	-	-	-	-	-	-	-	-	-	-	-
Coriobacterium	-	-	-	-	-	-	-	-	-	-	-	-

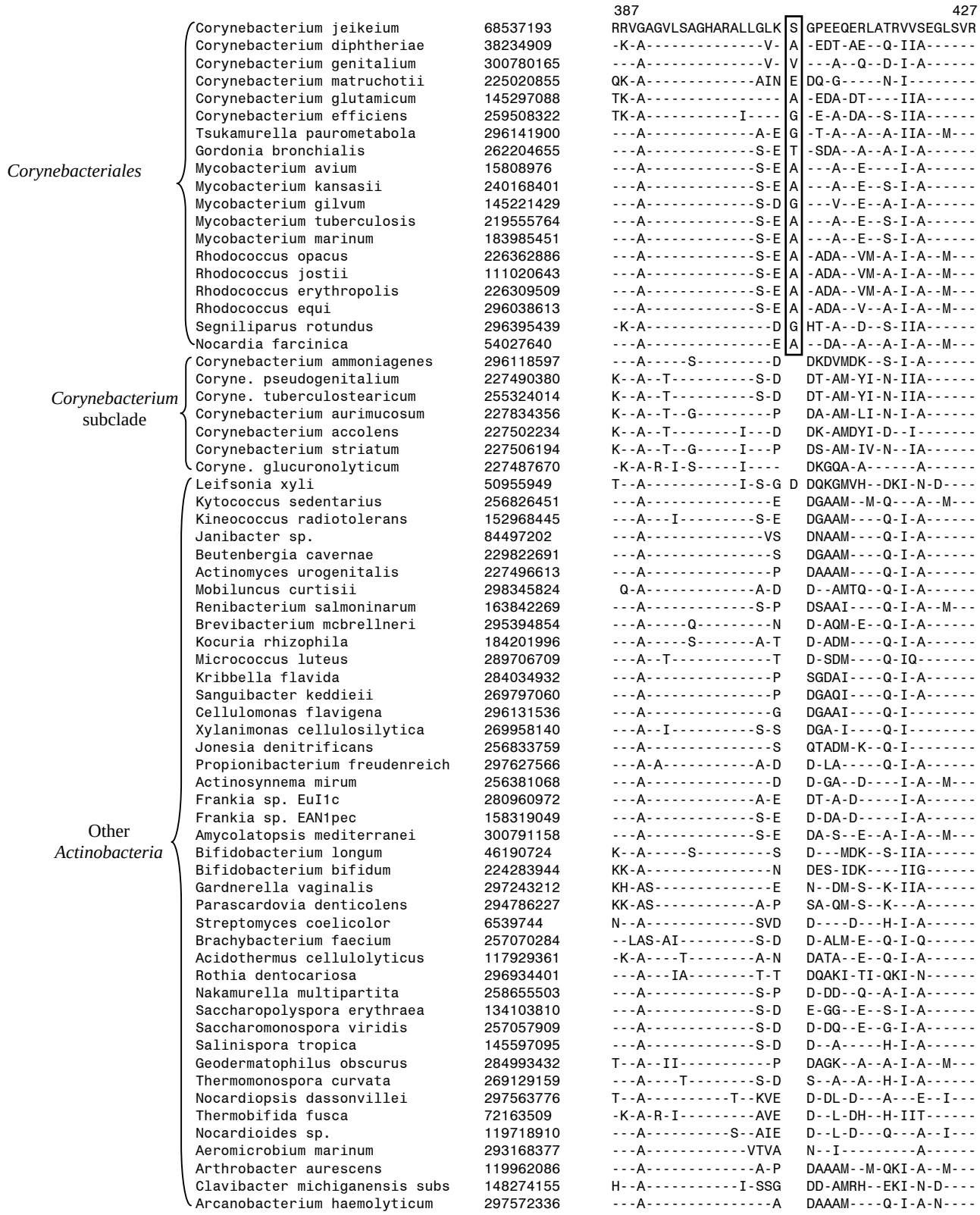
Supplemental File 8

Partial sequence alignment of the protein Alpha keotglutarate decarboxylase showing a 1 aa deletion that is uniquely present in various *Corynebacteriales* species but not found in any other actinobacteria. Sequence information is shown here for a only limited number of species; The dashes (-) in the sequence alignments indicate identity with the amino acid on the top line.

			360		397
	Segniliparus rotundus	296393143	GAESGDFLRTIHELLISDA		FYDEIYAALRIPYEPIRW
	Gordonia bronchialis	262201823	----------LD--		-----FT-FH-----V--
	Tsukamurella paurometabola	296138938	---------YV---TL--E		-W-D-FRT-HV----V--
	Nocardia farcinica	161598437	----------Q-----E		-----FHG-GV----V--
	Mycobacterium avium	41408634	----------M-L--G		-W-D-FRE-S---L----
	Mycobacterium leprae	221230016	----------MVL--S		-W---FRE-S---L-V--
Corynebacteriales	Mycobacterium tuberculosis	215403087	----------L--G		-W--VFRE-S---L-V--
	Rhodococcus opacus	226365480	----------K-----E		-----FHS-H-----V--
	Rhodococcus jostii	111022975	----------N-----E		-----FH--H-----V--
	Rhodococcus equi	296039173	----------N-----E		----VFHS-H-----V--
	Corynebacterium efficiens	25027746	-----E----MSQ--VD--		-W-H-FEEMNV--T-M--
	Corynebacterium jeikeium	260577589	-----E----MSR---N-E		-W---FH-M-V--A-T--
	Corynebacterium diphtheriae	161598675	-----E----MSQ-F-D--		-W---FESIG---A-M--
	Corynebacterium striatum	227503791	-----E----LGQ-IVD-R	G	-W--LFESMGV--Q-F--
	Stackebrandtia nassauensis	291302596	--Q--E--KRV----LGAD	G	-F-DVFTS--V----V--
	Saccharopolyspora erythraea	134102819	----------V-Q--LGEN	G	---D-FTS-------V--
	Nakamurella multipartita	258651662	-----E---R--Q---GED	Q	---DVF-S--L----V--
	Actinosynnema mirum	256380243	-----E--KRV-Q--LGED	G	---DVFTS--L-------
	Micromonospora aurantiaca	270502237	--Q--E--KVM----LGER	G	---D-FTS-------V--
	Salinispora arenicola	159039598	--Q--E--KVM----LGEH	G	---Q-FTS-------V--
	Leifsonia xyli	50955456	--G--E--KIV----TGKR	S	--EN-F-E-----A--Q-
	Clavibacter michiganensis	170782429	--G--E--KIV--R--GQH	G	--ED-F-------D--Q-
Other	Arthrobacter aurescens	119963962	--G--E--KLV-Q--LGAQ	N	-----FE--------V--
Actinobacteria	Micrococcus luteus	289706398	--G--E--KIV-G--LGED	G	----VFHS-------V--
	Propionibacterium acnes	295130820	--Q--E---RM-Q--LGAD	R	--ED-FES-----A-VQ-
	Kytococcus sedentarius	256825652	--A--E--KLV-GY-NGED	G	-----FQS-------V--
	Catenulispora acidiphila	256396666	--Q------I--Q--LGED	G	---KLFES-------V--
	Streptomyces coelicolor	21223647	--A--E---QVAN--LGEN	G	---D-FK--------V--
	Frankia sp. EUN1f	288917901	--Q--EY--R--Q--LGAD	G	-----F-S--V--V----
	Thermobifida fusca	72160970	-----E---R--Q--LGED	G	-----FES-------V--
	Streptosporangium roseum	271969479	--Q------QV-R--LGED	G	-----FE--------V--
	Actinomyces coleocanis	227495004	--G--EL--AVANK-TGQD	G	--ERVFDSMKV--S-YE-
	Acidimicrobium ferrooxidans	256372497	-----Q--A--AA--EGEQ	D	-----FS--QV--P-M--
	Arcanobacterium haemolyticum	297570946	--T--R---LMEHK-LGLD	G	--ER--I--HV--K-FV-

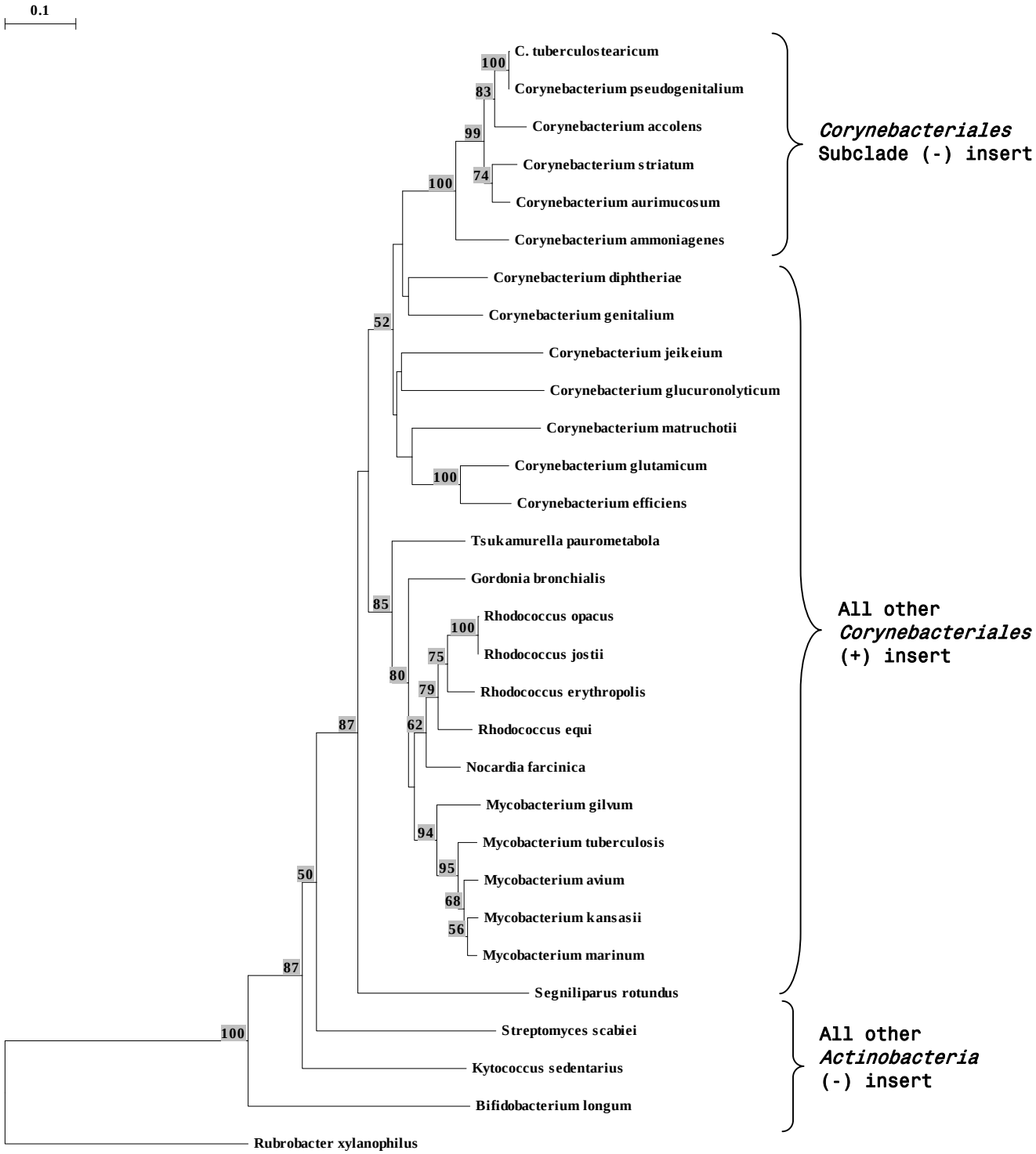
Supplemental File 9

A conserved insert in the Chromosome segregation DNA-binding protein ParB that is largely specific for the *Corynebacteriale*s. This insert is lacking in some of the *Corynebacterium* species, which form a distinct clade within this genus (see Supplemental file 11)



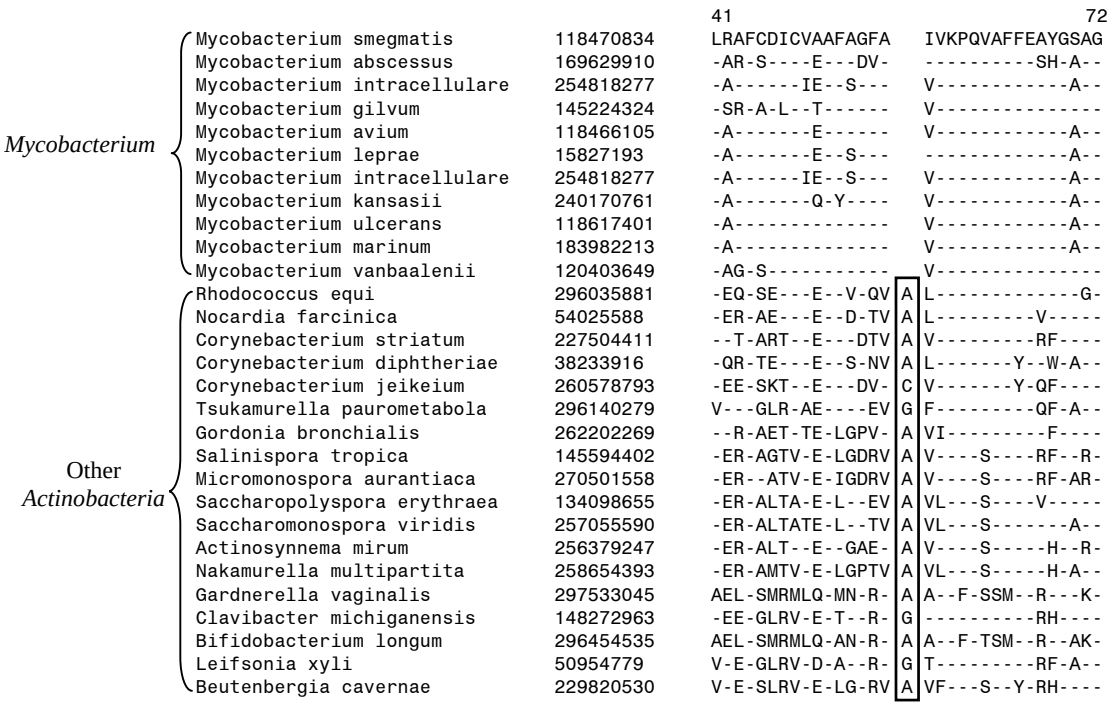
Supplemental File 10

A neighbor-joining distance tree for Corynebacteriales species based upon ParB sequences. Some of the *Corynebacterium* species that lack the insert in this protein (Sup. File 9) form a distinct clade in this tree indicating that the insert-containing gene within this clade was replaced within an insert-lacking gene or this insert was lost from the gene.



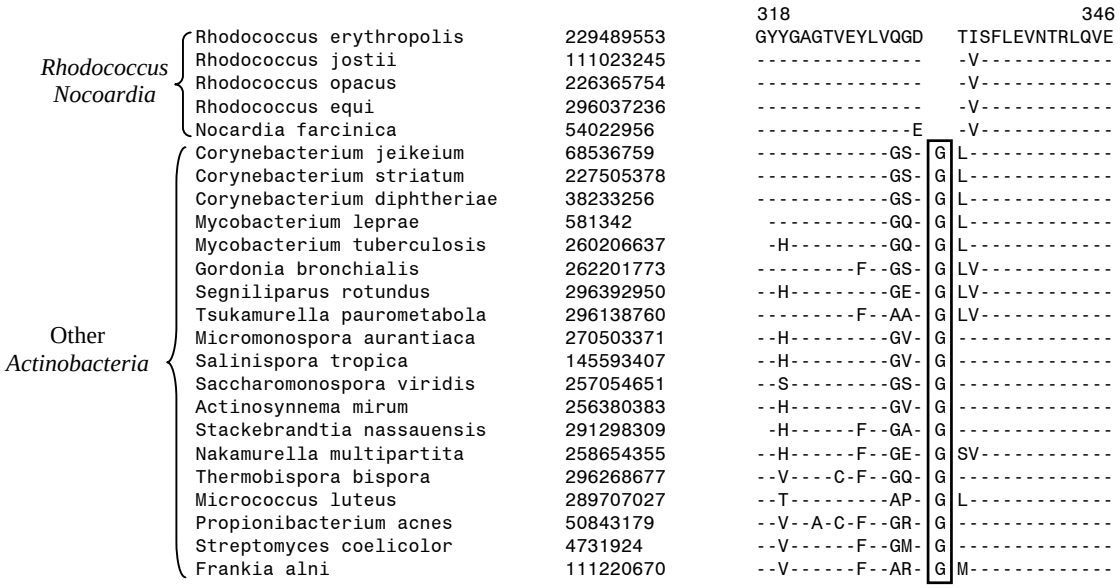
Supplemental File 11

Partial sequence alignment of the protein Orotodine-5'-phosphate decarboxylase showing a 1 aa deletion in conserved regions that is unique to various *Mycobacterium* species but not found in any other actinobacteria.



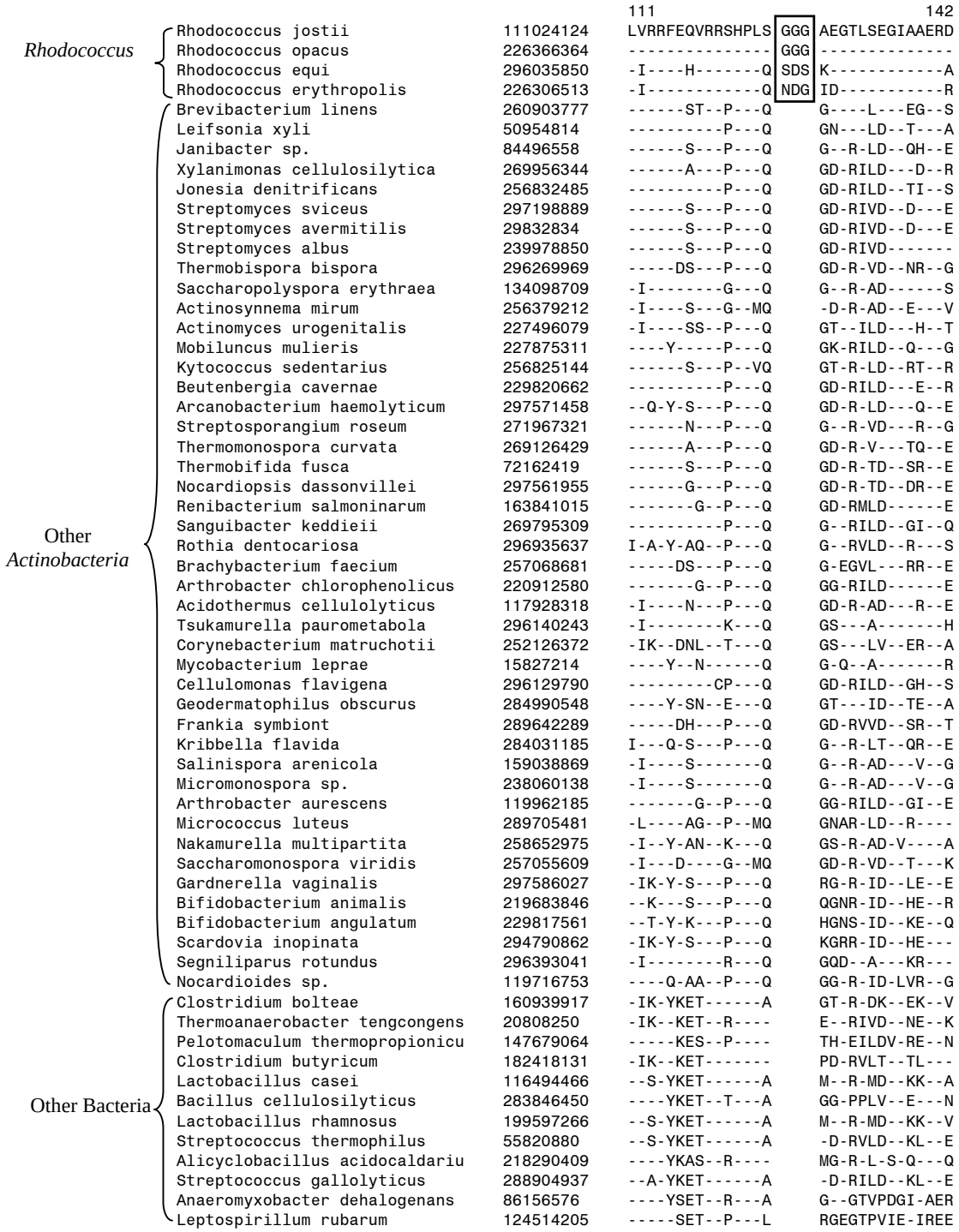
Supplemental File 12

Partial sequence alignment of the protein Acyl-CoA-carboxylase showing a 1 aa deletion in a conserved region that is uniquely found in *Rhodococcus-Nocardia* species.



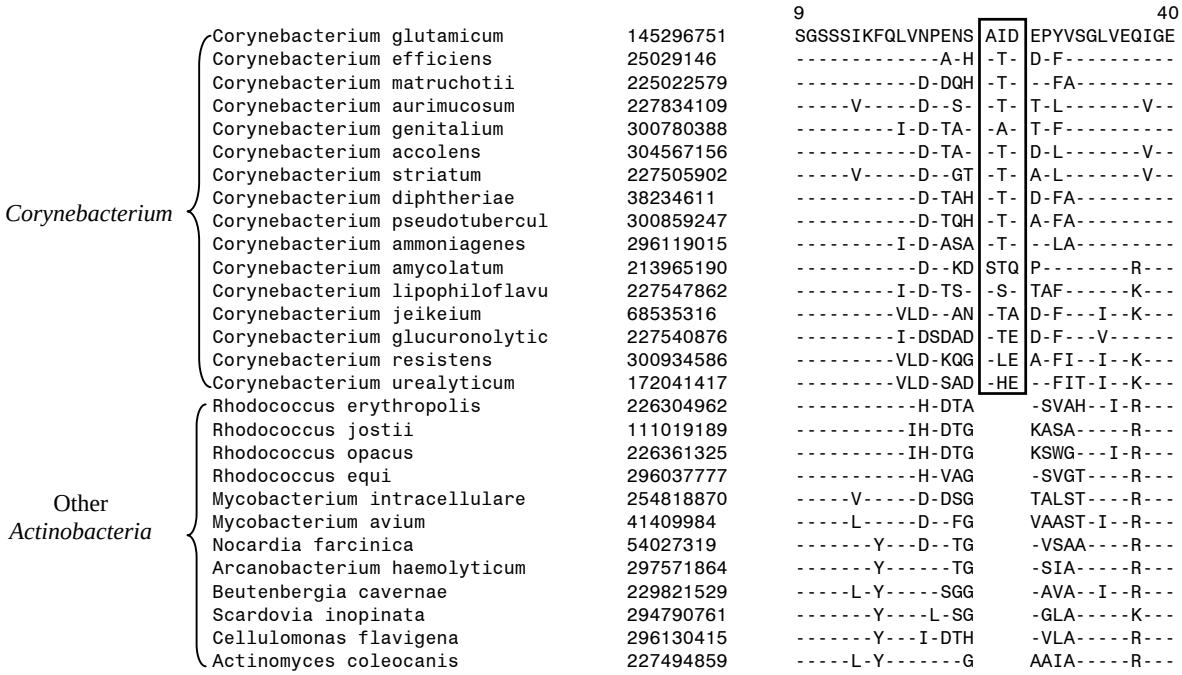
Supplemental File 13

A conserved insert in the protein BlinB_00480 of unknown function that is specific for the genus *Rhodococcus.*



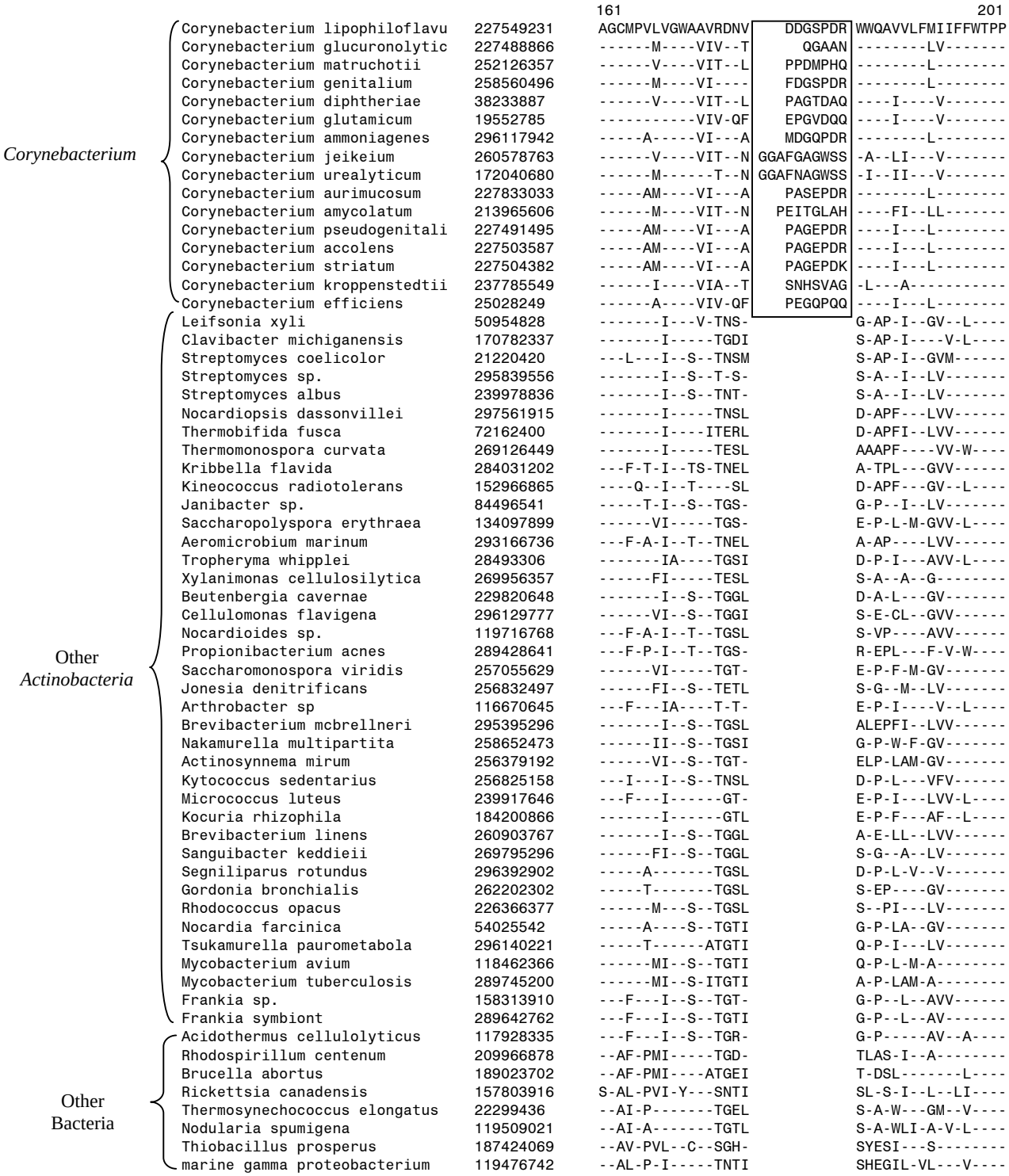
Supplemental File 14

Partial sequence alignments of the protein Acetate Kinase showing a conserved insert that is uniquely found in various *Corynebacterium* species, but not in any other bacteria.



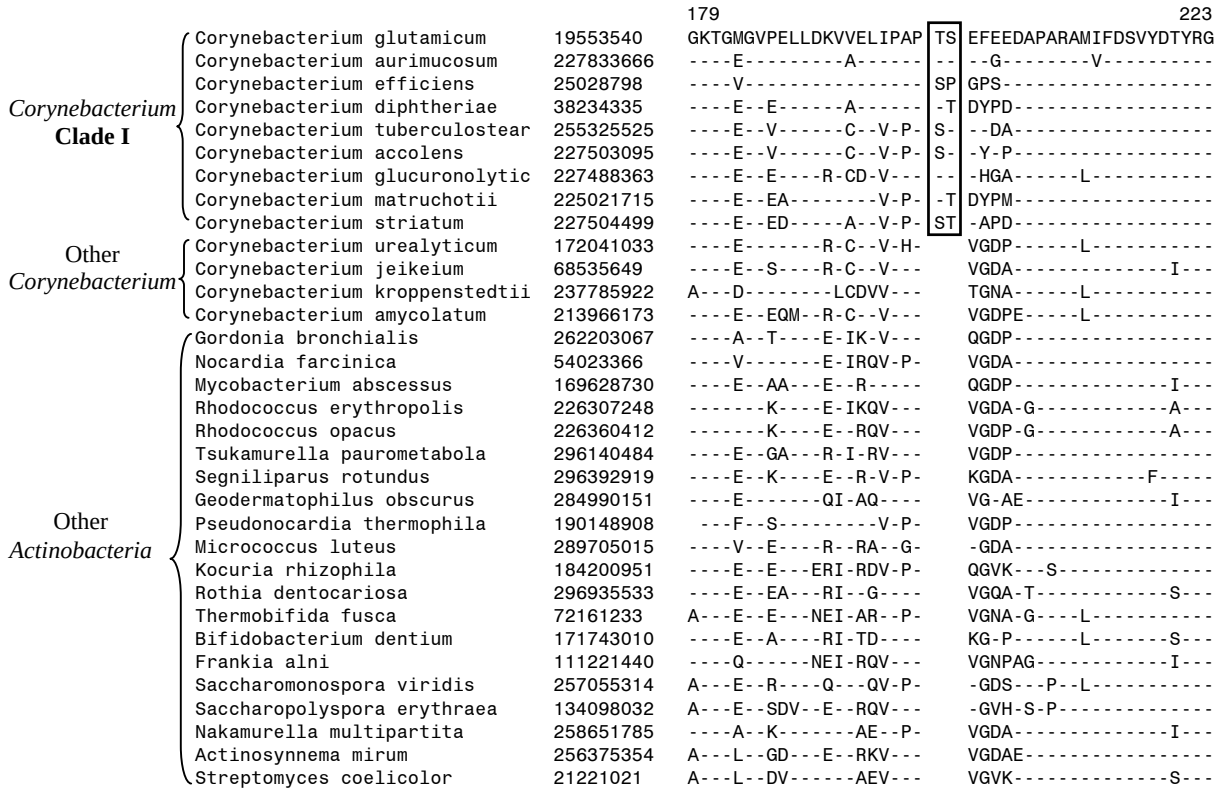
Supplemental File 15

A conserved insert in the protein Protoheme IX farnesyltransferase (CyoE) that is specific for the genus *Corynebacterium.*



Supplemental File 16

A conserved insert in the LepA protein that is specific for the *Corynebacterium* Clade-1 species.



Supplemental File 17

Signature Proteins that are specific for *Corynebacterium* Subclades

Gene/Protein Name	Accession No.	Function	Length (aa)	Species Specificity
NCgl0010	NP_599262.1	Hypothetical	87	*Corynebacterium* Subclade-1
NCgl0032	NP_599285.1	Hypothetical	270	*Corynebacterium* Subclade-1
NCgl0152	NP_599407.1	Hypothetical	247	*Corynebacterium* Subclade-1
NCgl0559	NP_599820.1	Hypothetical	101	*Corynebacterium* Subclade-1
NCgl0750	NP_600012.1	Hypothetical	139	*Corynebacterium* Subclade-1
NCgl1091	NP_600364.1	Hypothetical	165	*Corynebacterium* Subclade-1
NCgl1554	NP_600830.1	Hypothetical	154	*Corynebacterium* Subclade-1
NCgl1838	NP_601119.1	Hypothetical	179	*Corynebacterium* Subclade-1
NCgl1867	NP_601149.1	Hypothetical	179	*Corynebacterium* Subclade-1
NCgl1870	NP_601152.1	Hypothetical	206	*Corynebacterium* Subclade-1
NCgl1914	NP_601195.1	Hypothetical	272	*Corynebacterium* Subclade-1
NCgl1970	NP_601251.1	Hypothetical	126	*Corynebacterium* Subclade-1
NCgl2001	NP_601281.1	Hypothetical	132	*Corynebacterium* Subclade-1
NCgl2018	NP_601298.1	Hypothetical	75	*Corynebacterium* Subclade-1
NCgl2022	NP_601302.1	Hypothetical	400	*Corynebacterium* Subclade-1
NCgl2036	NP_601317.1	Hypothetical	202	*Corynebacterium* Subclade-1
NCgl2177	NP_601457.1	Hypothetical	142	*Corynebacterium* Subclade-1
NCgl2225	NP_601506.1	Hypothetical	188	*Corynebacterium* Subclade-1
NCgl2226	NP_601507.1	Hypothetical	173	*Corynebacterium* Subclade-1
NCgl2043	NP_601325.1	Hypothetical	77	*Corynebacterium* Subclade-1
NCgl2536	NP_601826.1	Hypothetical	162	*Corynebacterium* Subclade-1
NCgl0048	NP_599301.1	Hypothetical	157	*C. glutamicum* and *C. efficiens*
NCgl0077	NP_599330.1	Hypothetical	186	*C. glutamicum* and *C. efficiens*
NCgl0097	NP_599350.1	Hypothetical	183	*C. glutamicum* and *C. efficiens*
NCgl0122	NP_599376.1	Hypothetical	172	*C. glutamicum* and *C. efficiens*
NCgl0138	NP_599393.1	Hypothetical	128	*C. glutamicum* and *C. efficiens*
NCgl0189	NP_599445.1	Hypothetical	163	*C. glutamicum* and *C. efficiens*
NCgl0193	NP_599449.1	Hypothetical	134	*C. glutamicum* and *C. efficiens*
NCgl0366	NP_599625.1	Hypothetical	140	*C. glutamicum* and *C. efficiens*
NCgl0404	NP_599666.1	Hypothetical	94	*C. glutamicum* and *C. efficiens*

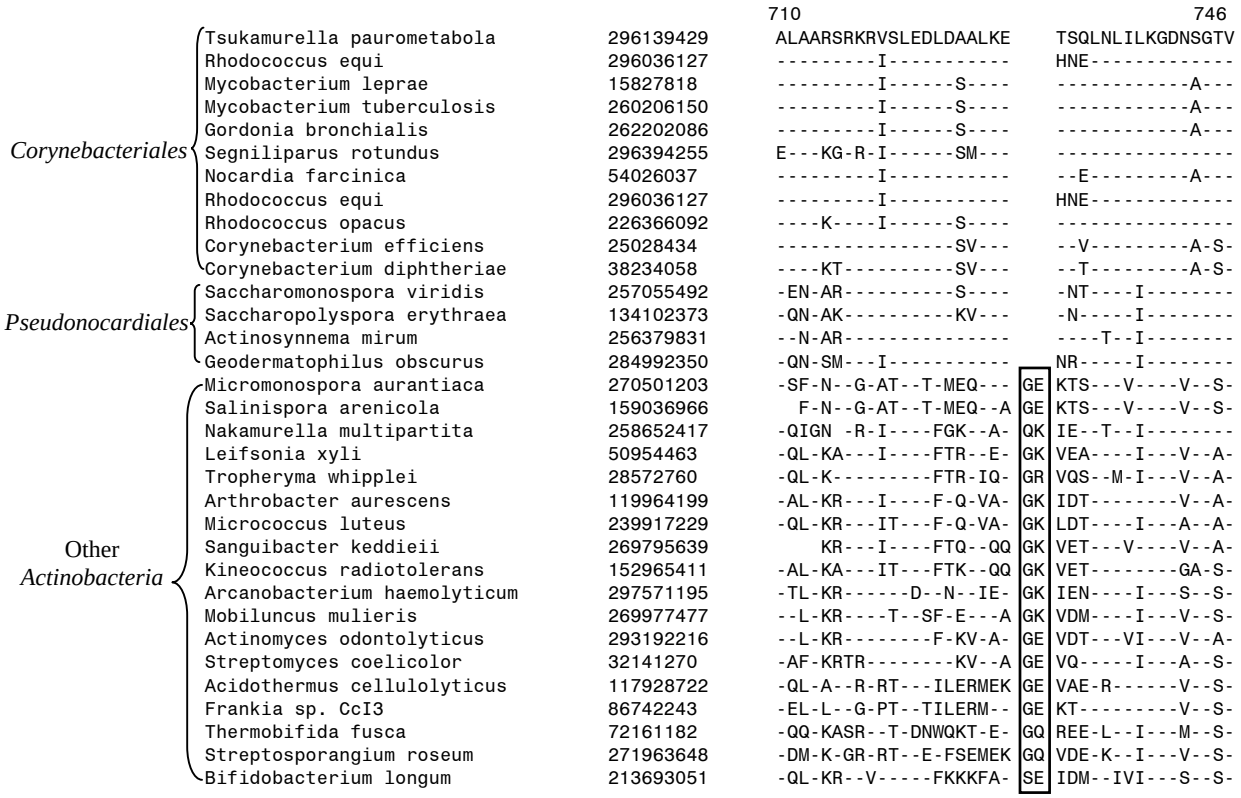
NCgl0473	NP_599735.1	Hypothetical	60	*C. glutamicum* and *C. efficiens*
NCgl0508	NP_599769.1	Hypothetical	209	*C. glutamicum* and *C. efficiens*
NCgl0587	NP_599848.1	Hypothetical	108	*C. glutamicum* and *C. efficiens*
NCgl0633	NP_599895.1	Hypothetical	234	*C. glutamicum* and *C. efficiens*
NCgl0643	NP_599905.1	Hypothetical	63	*C. glutamicum* and *C. efficiens*
NCgl0917	NP_600183.1	Hypothetical	83	*C. glutamicum* and *C. efficiens*
NCgl0932	NP_600198.1	Hypothetical	123	*C. glutamicum* and *C. efficiens*
NCgl0933	NP_600199.1	Hypothetical	126	*C. glutamicum* and *C. efficiens*
NCgl1045	NP_600318.1	Hypothetical	79	*C. glutamicum* and *C. efficiens*
NCgl1047	NP_600320.1	Hypothetical	129	*C. glutamicum* and *C. efficiens*

These CSPs were identified by Blastp searches on different proteins from the genome of *C. glutamicum* ATCC 13032.

Subclade-1 consists of *Corynebacterium glutamicum, C. efficiens, C. aurimucosum, C. diphtheriae, C. pseudotuberculosis, and C. ulcerans.*

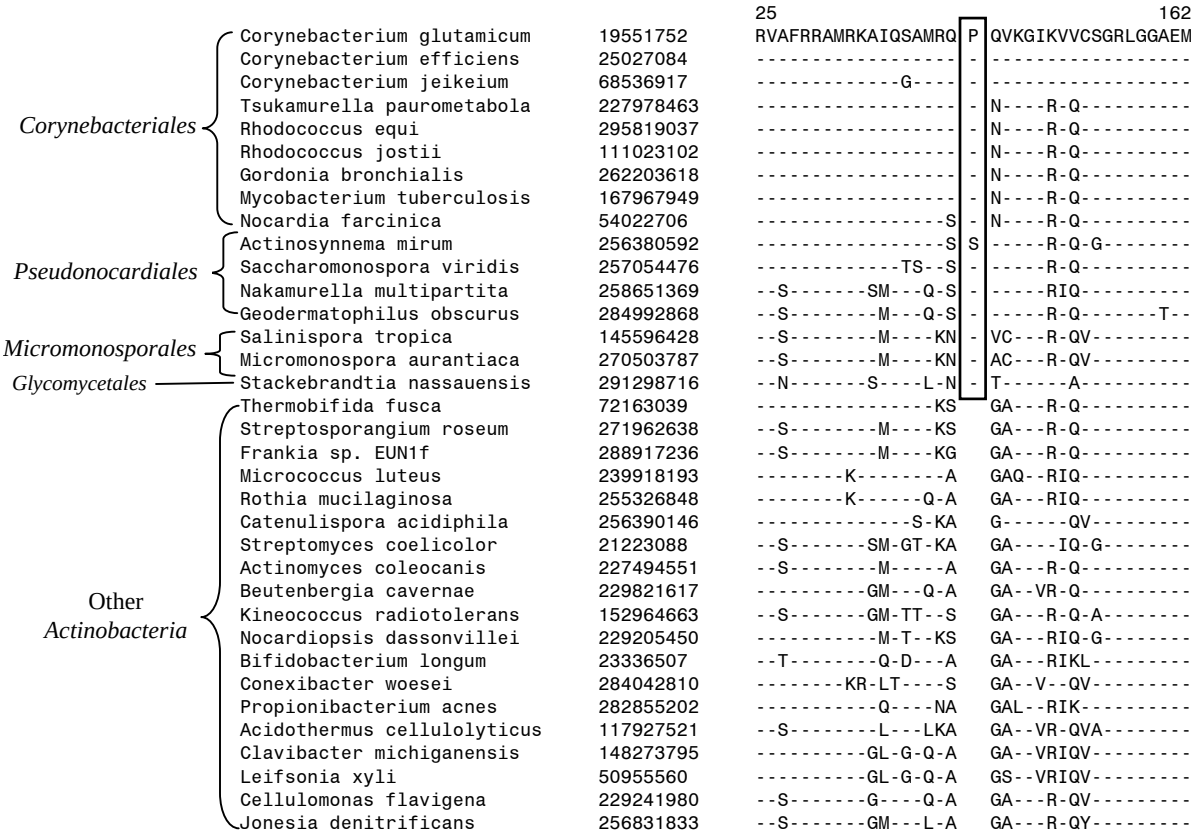
Supplemental File 18

Partial sequence alignments of the protein translation initiation factor-IF2 showing a 2 aa deletion that is uniquely shared by various species from the orders *Corynebacteriales* and *Pseudonocardiales*, but not found in any other actinobacteria.



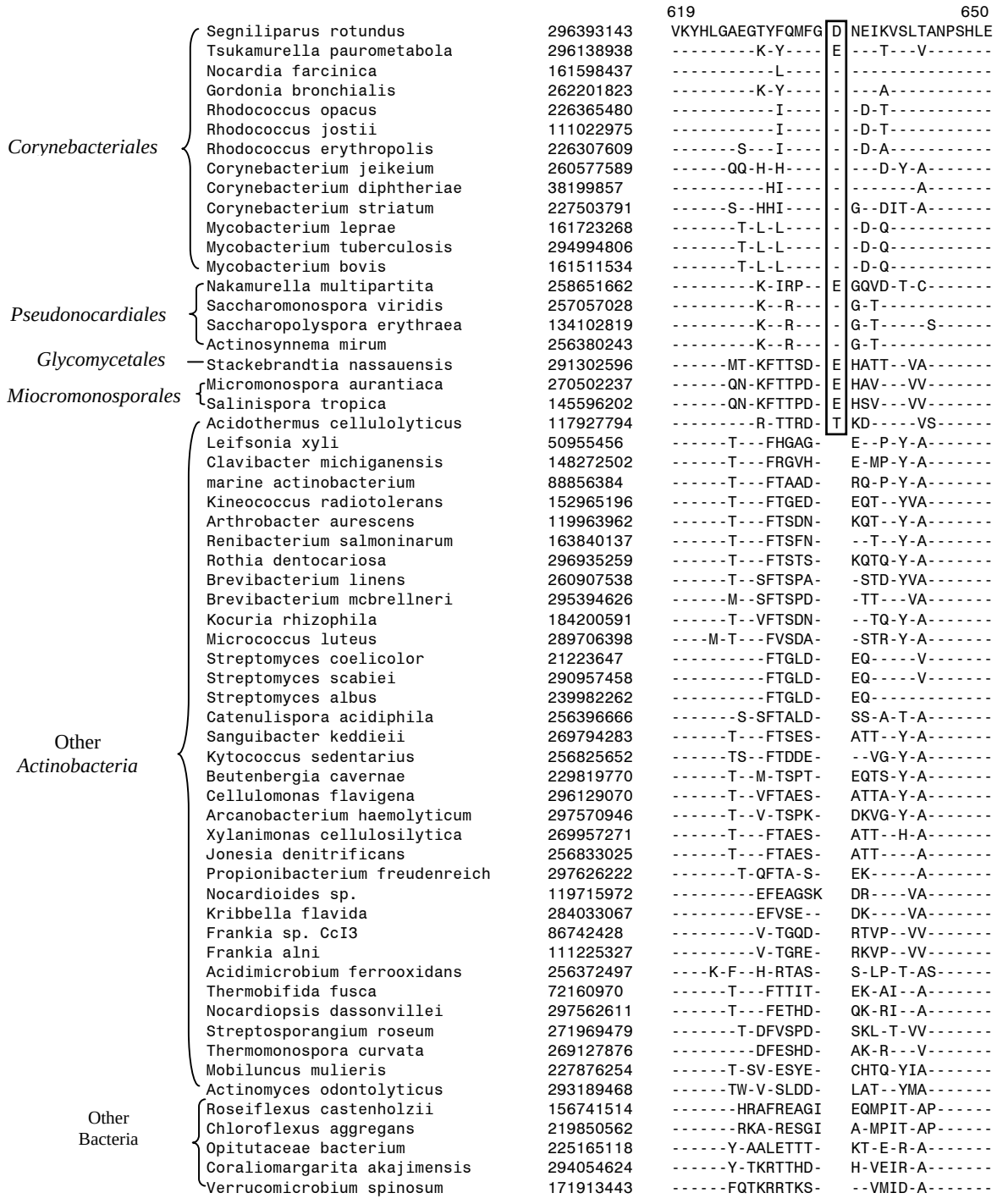
Supplemental File 19

Partial sequence alignment of ribosomal protein S3 showing a CSI that is uniquely shared by various species from the orders *Corynebacteriales, Pseudonocardiales, Micromonosporales* and *Glycomycetales* indicating that species from these groups shared a common ancestor exclusive of other actinobacteria.



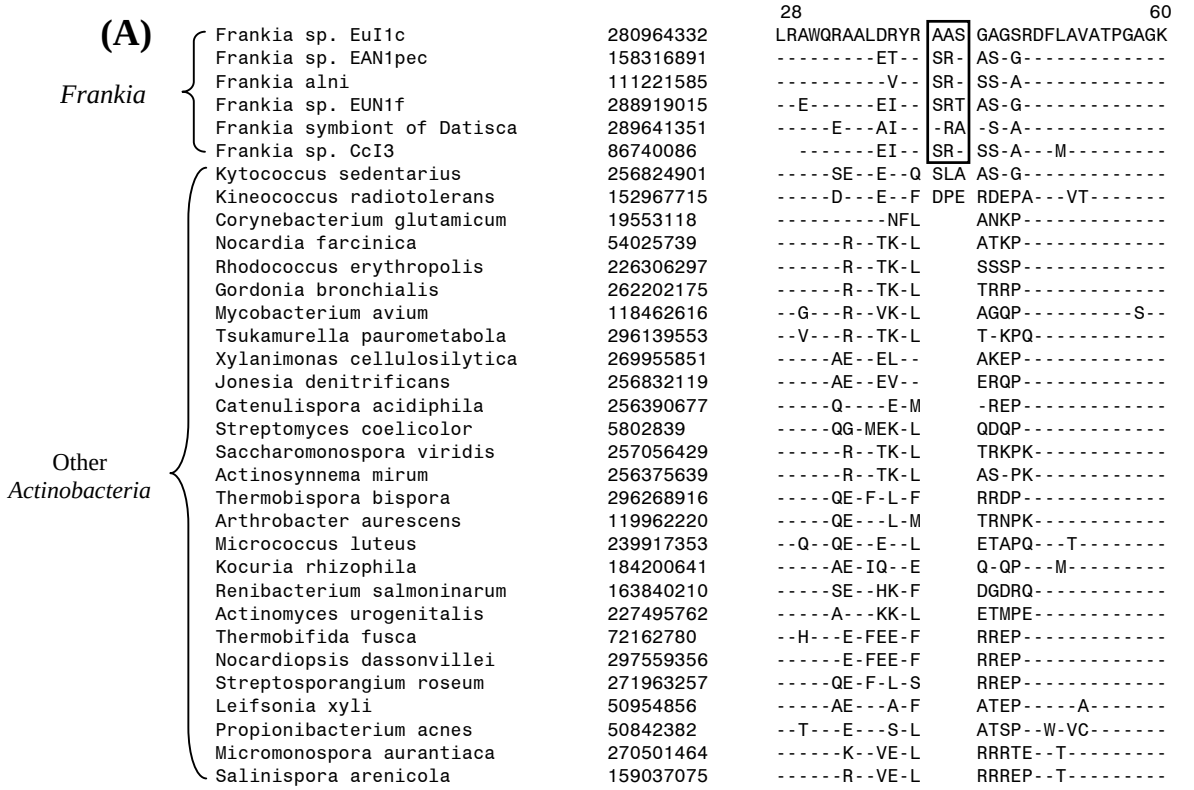
Supplemental File 20

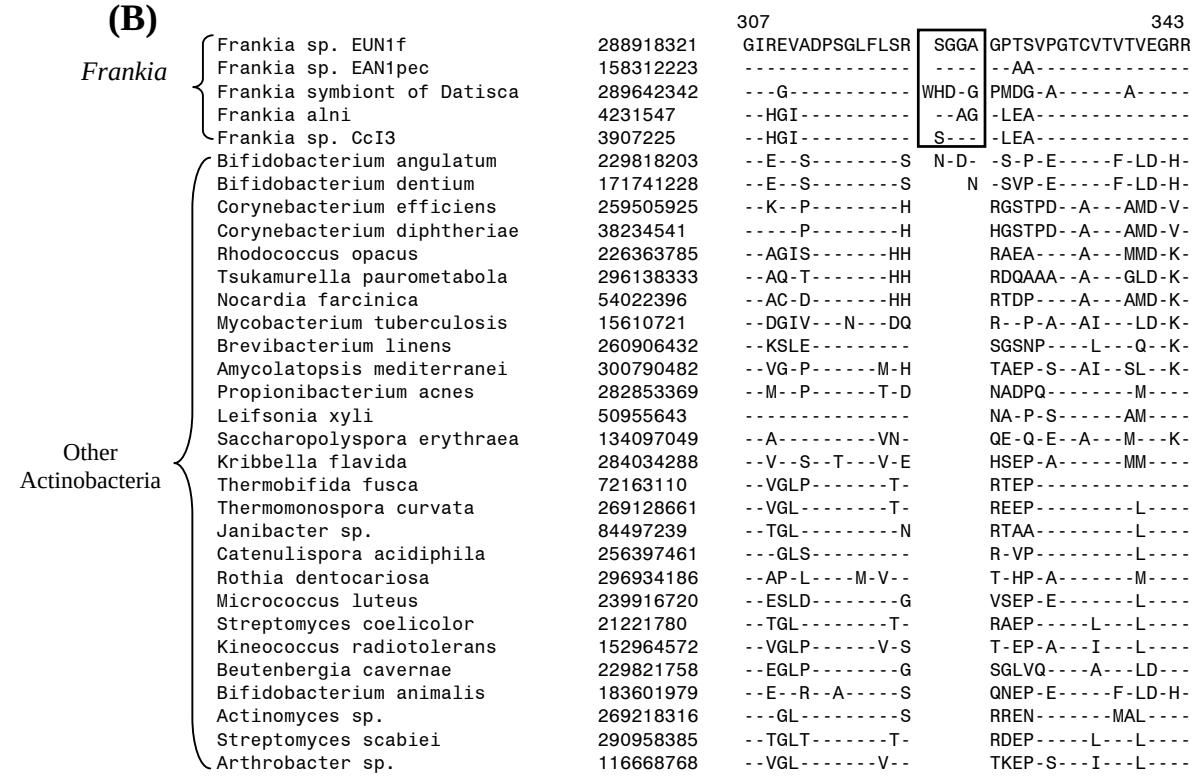
Partial Sequence alignment of the protein Alpha-ketogluterate decarboxylase showing a 1 aa conserved insert that is largely specific for the *Corynebacteriales*, *Pseudonocardiales* and *Micromonosporales* species. This insert is also present in *Glycomyetales* and *A. cellulolyticus*.



Supplemental File 21

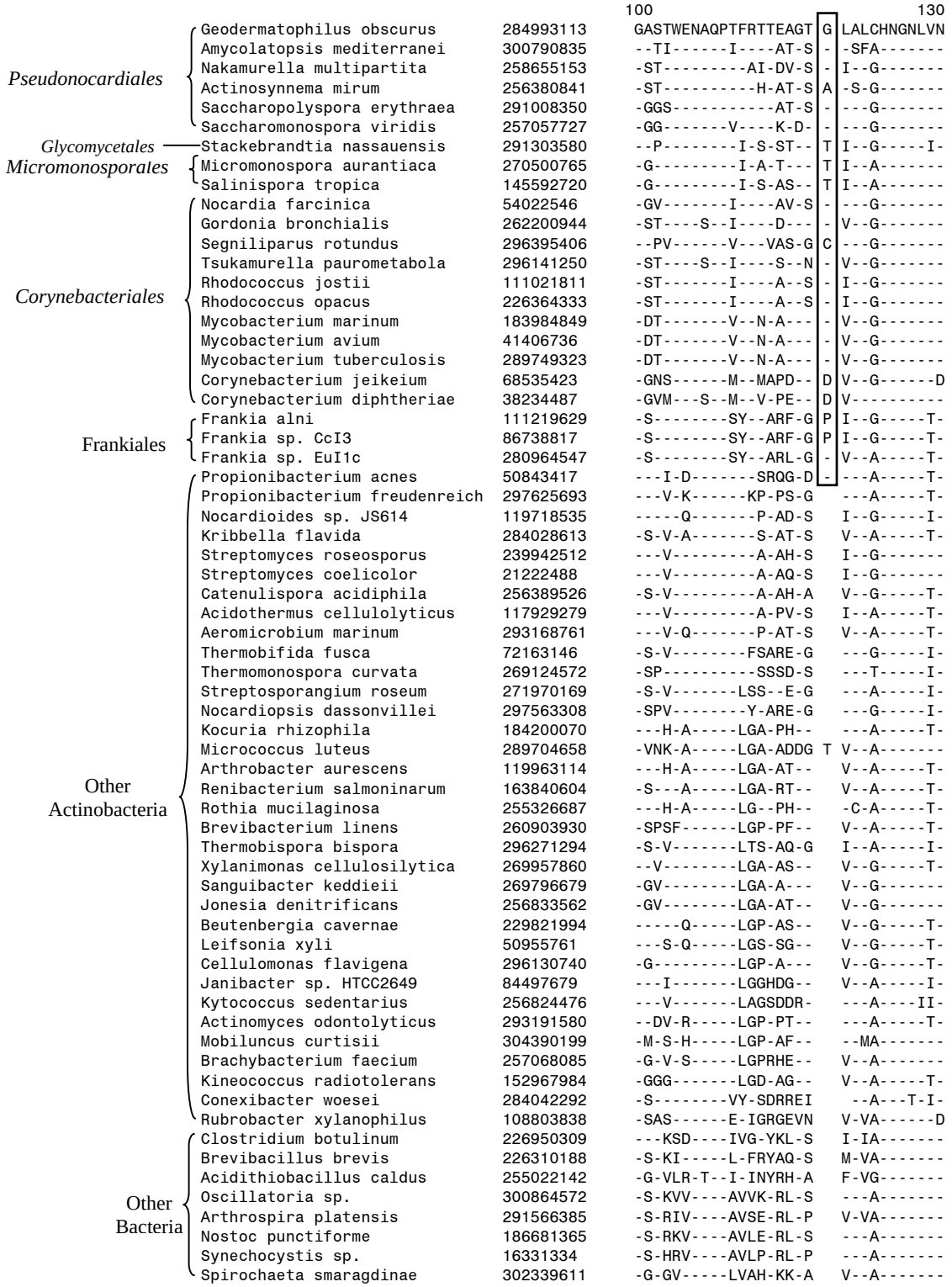
Partial Sequence alignments of (A) the DNA repair RadA protein and (B) hypothetical protein Ncg1 showing two conserved inserts that are largely specific for the *Frankiae* species.





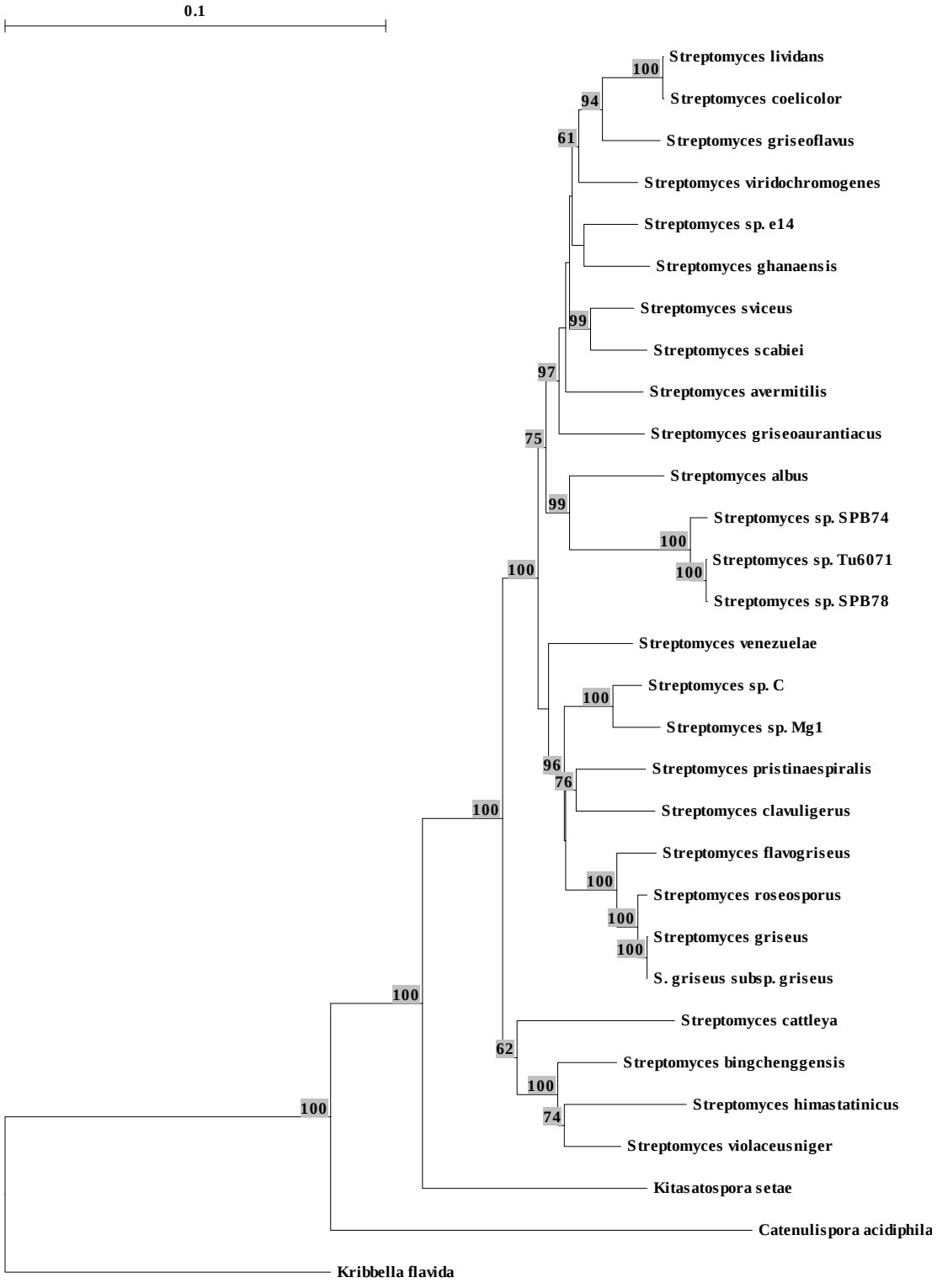
Supplemental File 22

Partial sequence alignments of the protein glutamine phosphoribosyl amidotransferase (GPAT) depicting a 1 aa conserved insert that is commonly shared by various species from the orders Corynebacteriales, Pseudonocardiales and Micromonosporales, Glycomycetales and Frankiales, but not present in any other actinobacteria or bacteria.



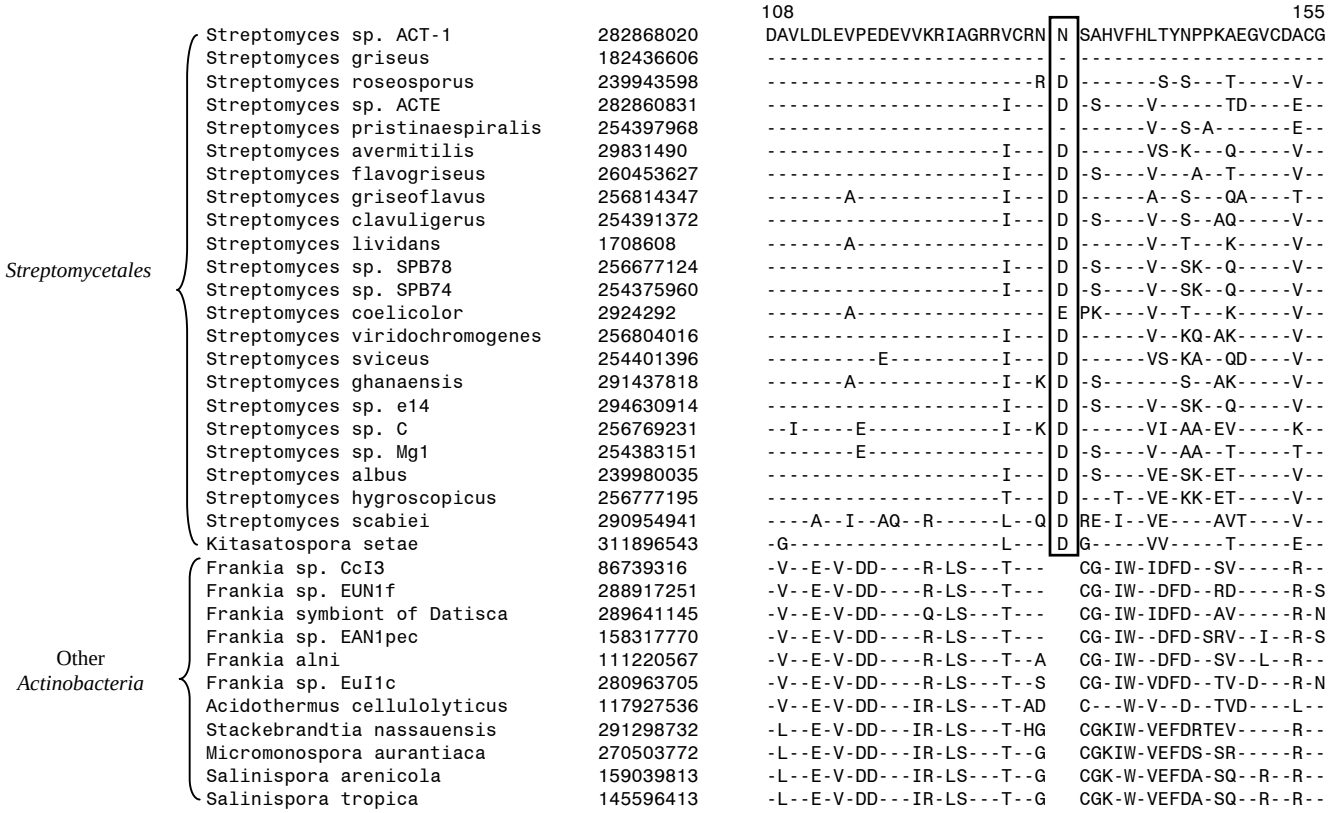
Supplemental File 23

A neighbor-joining distance tree for *Streptomycetales* species based upon concatenated sequences for RpoB, RpoC and Gyrase B proteins.



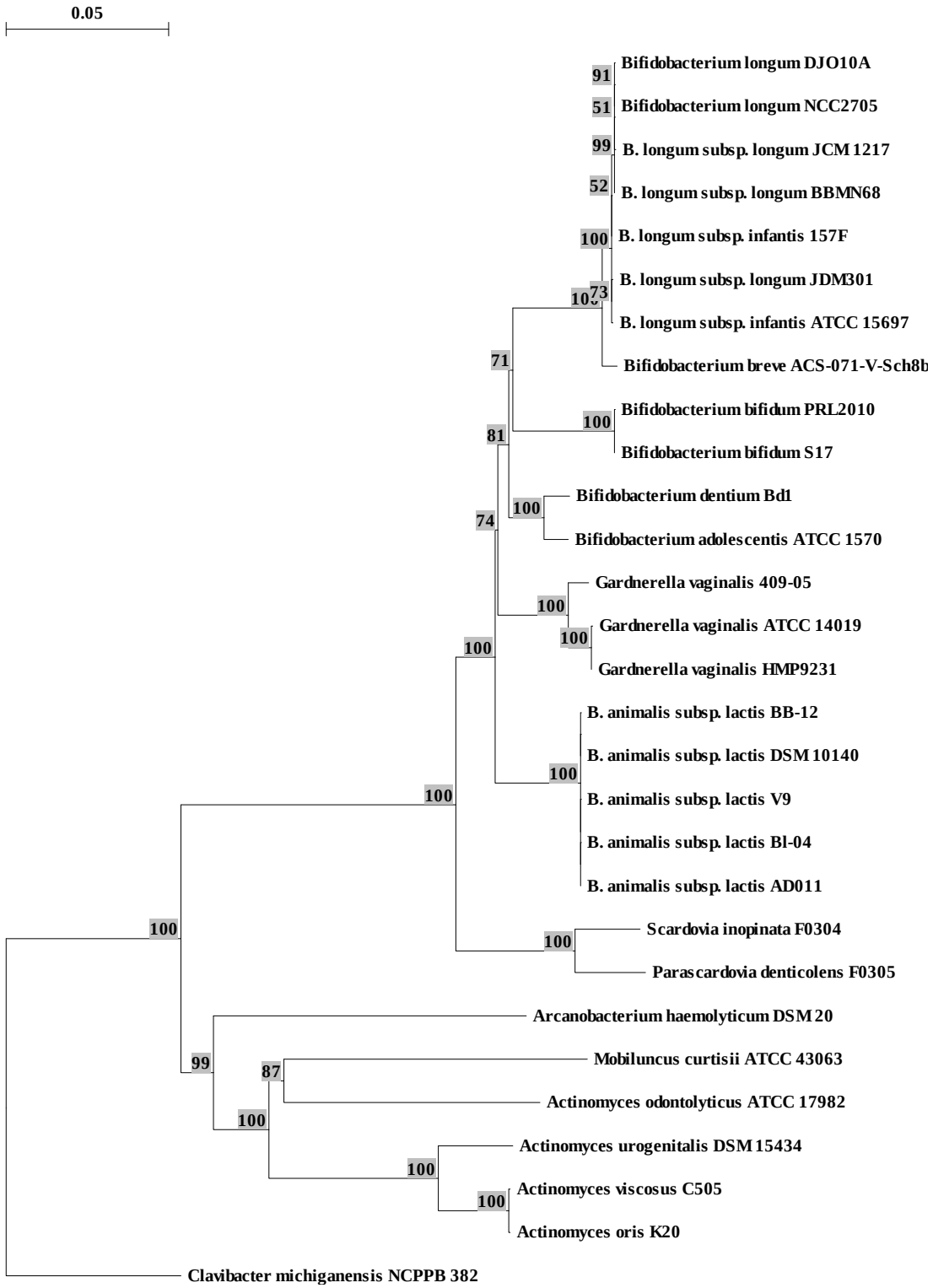
Supplemental File 24

Partial Sequence alignment of the protein Adenylate Kinase showing a 1 aa conserved insert that is specific for the *Streptomycetaceae* species



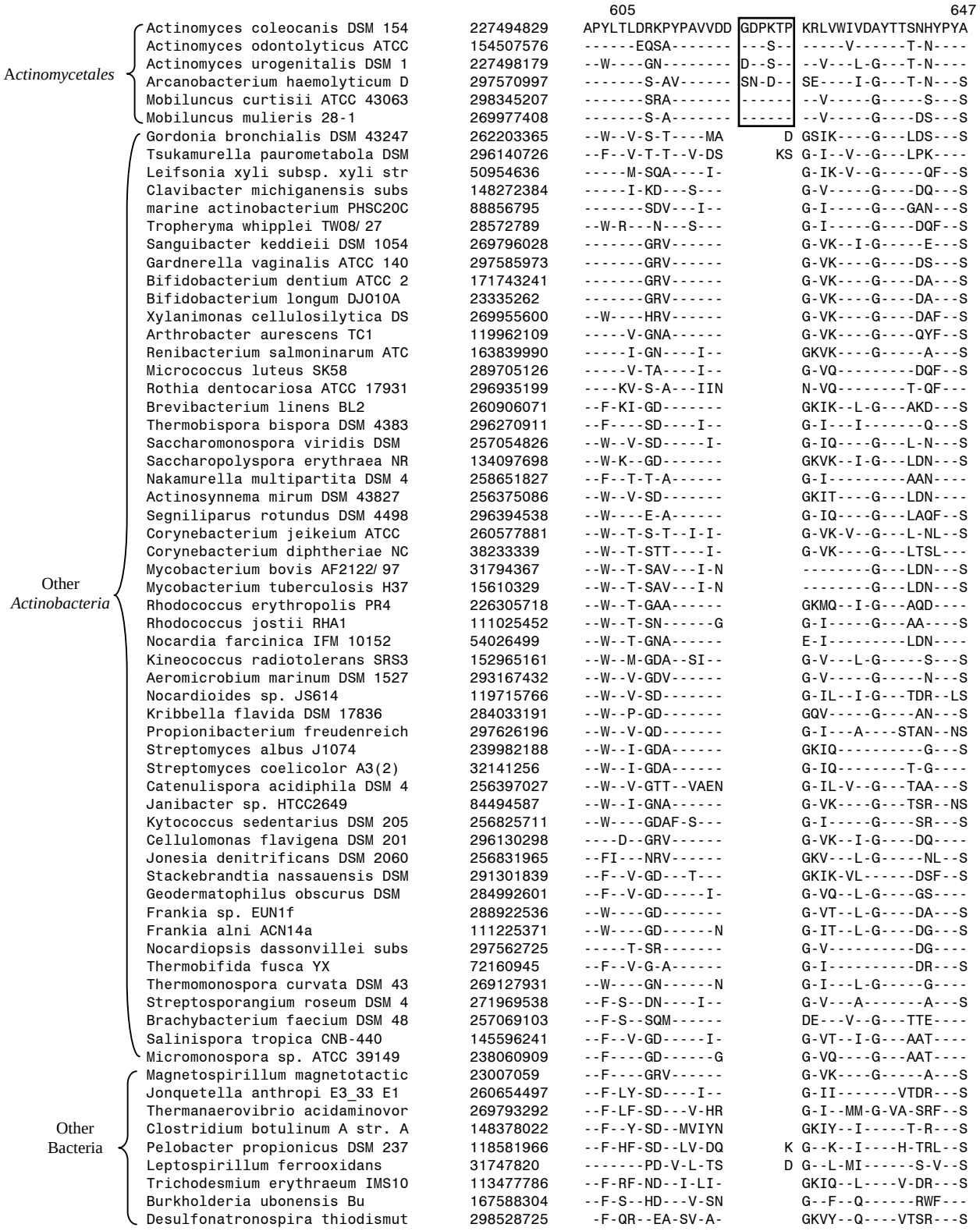
Supplemental File 25

A neighbor-joining tree for sequenced *Bifidobacteriales* and *Actinomycetales* species based upon concatenated sequences for RpoB, RpoC and Gyrase B proteins.



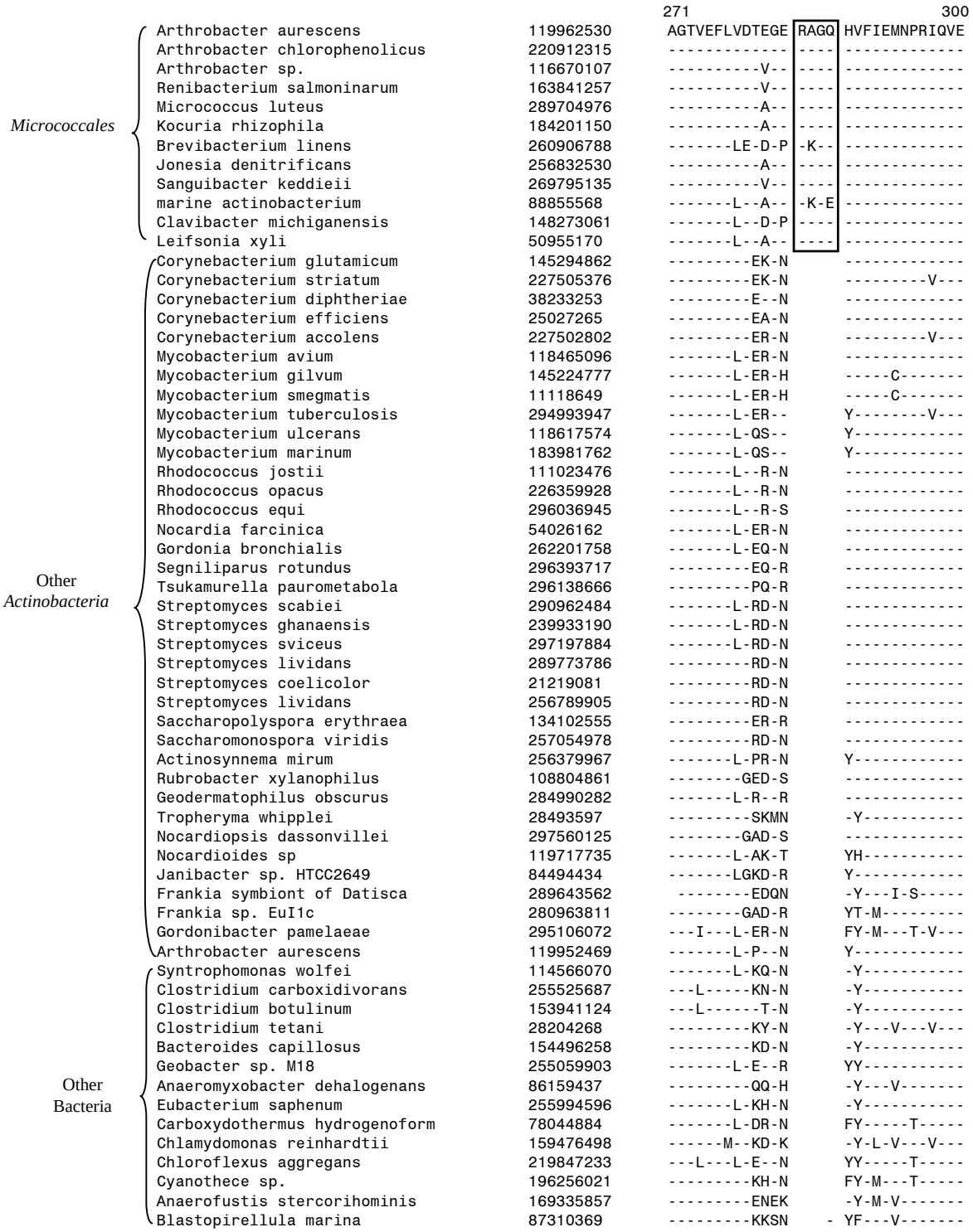
Supplemental File 26

Partial Sequence alignment of the integral membrane protein Lxx09300 showing a 6 aa conserved insert that is specific for the *Actinomycetales* species



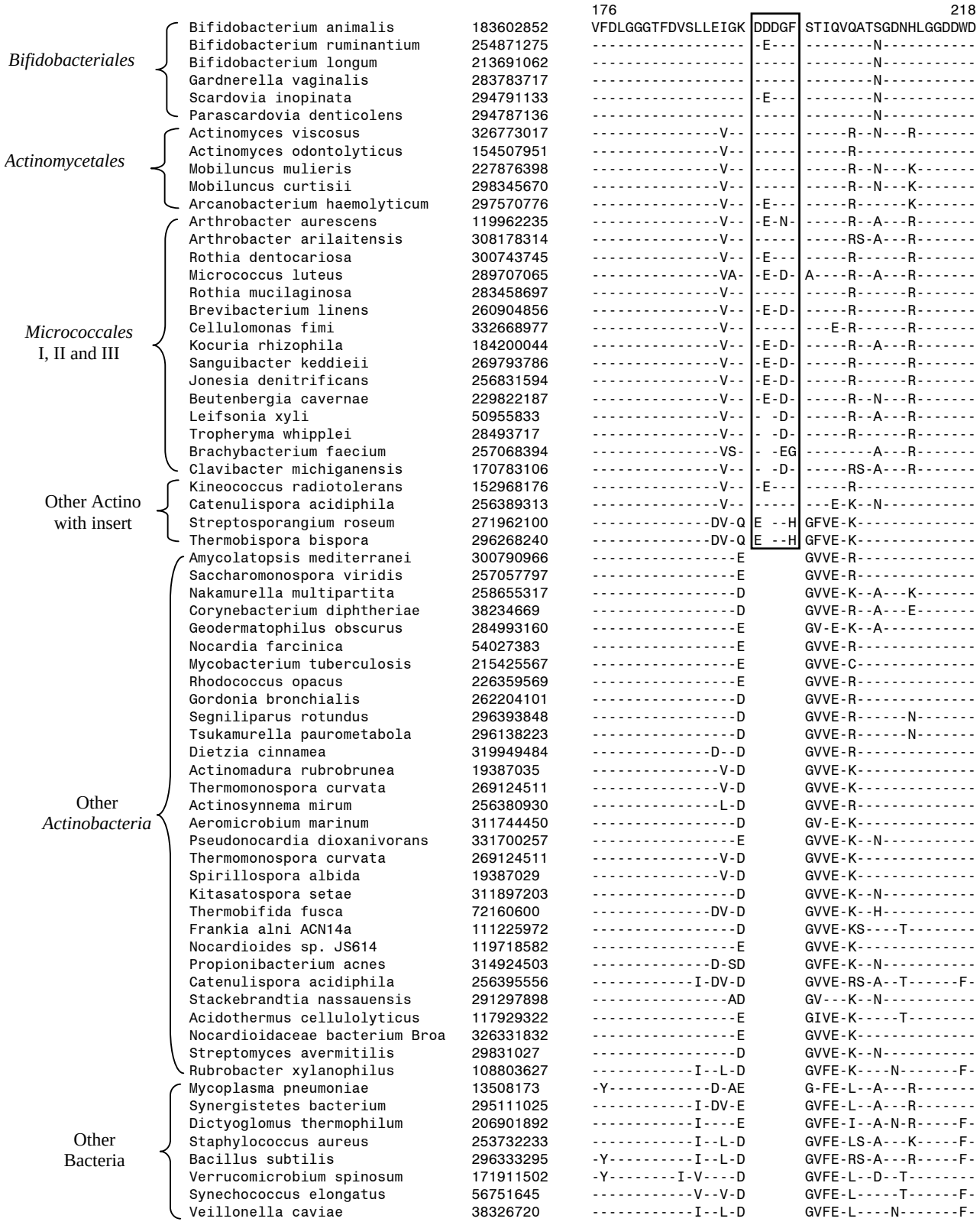
Supplemental File 27

Partial Sequence alignment of the protein puruvate carboxylase showing a 4 aa conserved insert that is specific for various *Micrococcales*.



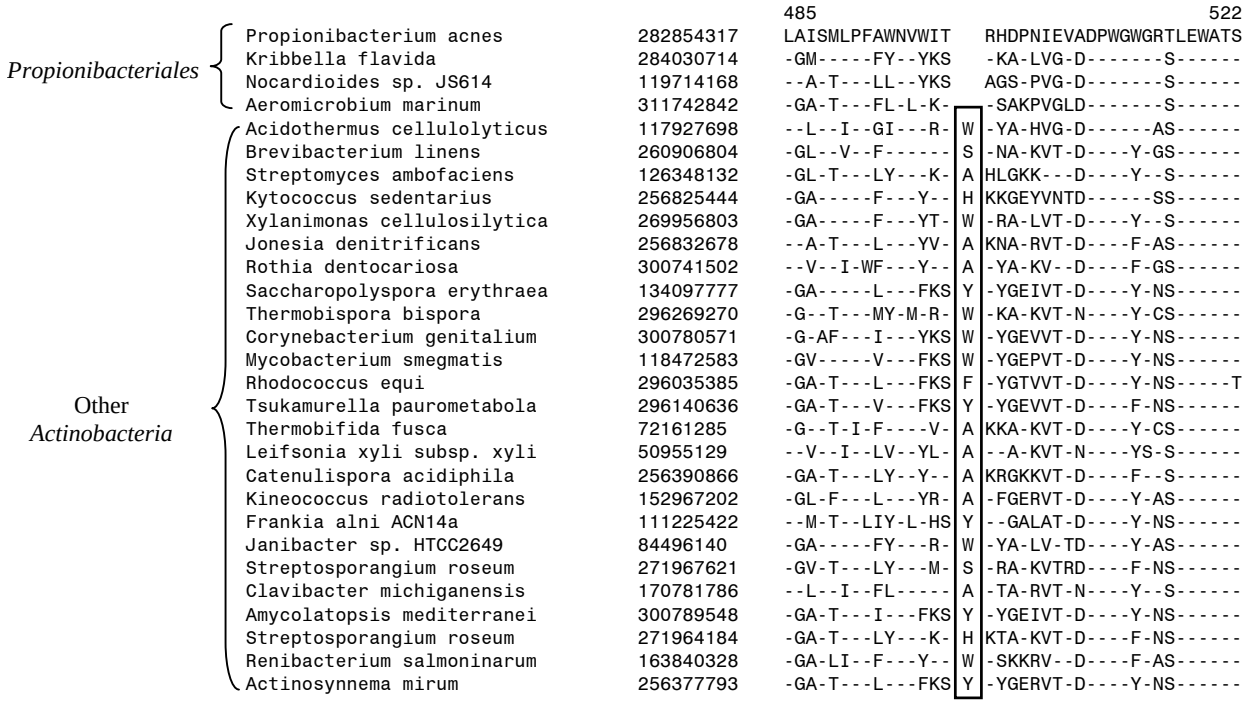
Supplemental File 28

Partial Sequence alignment of the DnaK (Hsp70) protein showing a 5 aa conserved insert that is mainly found in various sequenced *Bifidobacteriales*, *Actinomycetales* and *Micrococcales* species.



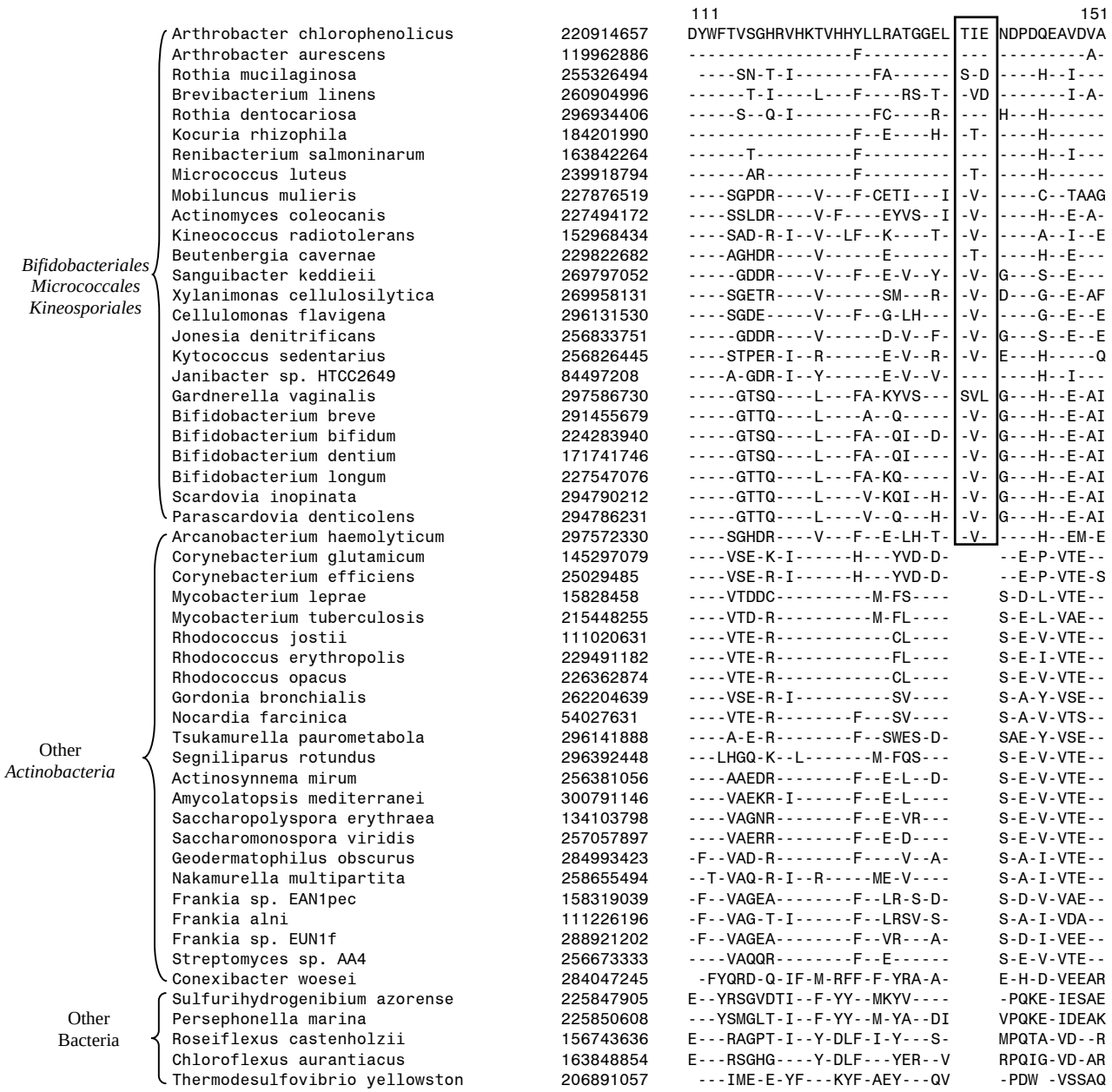
Supplemental File 29

Excerpt from the sequence alignment for Cytochrome c oxidase subunit I showing a 1 aa deletion that is uniquely present in various sequenced *Propionibacteriales* species.



Supplemental File 30

Partial Sequence alignment of the protein cgR_2975 showing a 3 aa conserved insert that is mainly found in various sequenced *Bifidobacteriales*, *Micrococcales* and *Kineosporales* species. The homologs for this protein were not detected in *Actinomycetales.*



Supplemental File 31

Partial Sequence alignment of the protein Cytochrome c oxidase subunit I (Cox1) showing a 1 aa conserved insert that is mainly found in various sequenced *Micrococcales* and *Kineosporales* species. The homologs for this protein were not detected in *Bifidobacteriales* and *Actinomycetales.*