

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 β subunit (RpoB)	1161	Gyrase B	686
RNA polymerase β' 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. Cc13*, *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. magentosprillum*.

		258	GSE	298
<i>Actinobacteria</i>	Thermobifida fusca	72161265	EKLSHYSKRTVDIEYRFNFA	WGEELEGIANRTDYDLSTH
	Corynebacterium glutamicum	19553476	-----A----	-----RV-
	Mobiluncus curtisii	298345360	-----A-G-E	-----GV-
	Nocardioopsis dassonvillei	297559717	-----Q	-----R--
	Jonesia denitrificans	256832756	-----G-T	-----V-----R--
	Nocardioideaceae bacterium	326334002	-----P	-----F--G-
	Clavibacter michiganensis	148272762	-----T-----G--	-----R--
	Actinomyces odontolyticus	154509410	-----L---G-Q	-----TV-
	Streptosporangium roseum	271963873	-----T	-----V-----F--TA-
	Nakamurella multipartita	258652148	-----H-Q	-----F--K--
	Bifidobacterium longum	23464742	-----K-G-Q	-----D-----A-
	Kineococcus radiotolerans	152967313	-----G-Q	-----D-----F--G--
	Propionibacterium acnes	314928065	-----K-G--	-----D-----A-
	Gardnerella vaginalis	298253424	-----D-V	-----A-----F--
	Thermomonospora curvata	269127431	-----K-G-Q	-----D-----A-
	Scardovia inopinata	294790521	-----P	-----M-----F--S-
	Thermobispora bispora	296269165	-----Q-Q	-----V-----A--
	Rhodococcus equi	312140272	-----K-G-S	-----F--G--
	Cellulomonas flavigena	296130022	-----T-----G--	-----GA-----R--
	Clavibacter michiganensis	170782097	-----T-----G-Q	-----F--K--
	Xylanimonas cellulositytica	269956954	-----T-----R-Q	-----V-----K-
	Nocardia farcinica	54023426	-----A-----S-G	-----V-----S-
	Streptomyces coelicolor	21220965	-----E-T	-----Q-----F--T-
	Pseudonocardia dioxanivorans	331698238	-----T-----G-Q	-----F--K--
	Geodermatophilus obscurus	284990200	-----A-----G-Q	-----V-----F--
	Aeromicrobium marinum	311744086	-----T-----G-T	-----F--K--
	Beutenbergia cavernae	229820268	-----A-----Q-G	-----N-----NA-
	Streptomyces scabiei	290960578	-----S---V--K-G-G	-----NP-----V-----
	Gordonia neofelifaecis	326381627	-----A-----G-P	-----L--A-----F--A-
	Acidothermus cellulolyticus	117928006	-----A-----G	-----FS---V-----TV-
	Kitasatospora setae	311895944	-----G-T-----G-Q	-----F--
	Arthrobacter aurescens	119963161	D-----D-----K-G-M	-----NP-----V-----F--
	Mycobacterium tuberculosis	217069510	-----Q-G	-----T--FA-----F--
	Salinispora arenicola	159039341	-----G	-----Q--FS---V-----F--
	Catenulispora acidiphila	256396139	-----K--G	-----K--FD-----F--
	Kribbella flavida	284030420	-----G-----E--	-----F-----F--G--
	Tsukamurella paurometabola	296140437	-----R-G	-----FA-----F--A--
	Micromonospora aurantiaca	302869265	DS-A-----I---K-E--	-----M-V---A-F--K--
	Leifsonia xyli	50954005	D-----D--I-L---G-S	-----F--M-V-----F--V-
	Kytococcus sedentarius	256825028	D--A--A-----G-G	-----K--FD-----F--C-
	Stackebrandtia nassauensis	291299567	-ERA--AG-I-V---G-Q	-----M-V-----GV-
	Rothia mucilaginoso	255327243	A---SG-I-V---G-Q	-----M-V-----G--
	Brevibacterium linens	260903902	DERA--AG-I-F--K-G-S	-----N-----M-V-----TQ-
	Dietzia cinnamea	319947863	DERA--DA-I-L--K-G-Q	-----NP-----M-V-----F--N-
	Brevibacterium mcbrellneri	295394795	DE----S-----F--PWG	-----Y-----F--RA-
Rubrobacter ferrooxidans	256372043	-----N-P--	-----S-----RR-	
Rubrobacter xylanophilus	108803618	-----T-----G-S	-----F--K--	
→ Magnetosprillum magnetotactic	46206189	DE----NA-T---K-P-G	-----W---S-----RQ-	
Other Bacteria	Staphylococcus epidermidis	251810982	-E-----NA-T---H-P-G	-----W---S-----KR-
	Geobacillus kaustophilus	56421965	DE----NA-T---K-P-G	-----W---S---F--RK-
	Staphylococcus aureus	258423993	-E----TA-T---N-P-G	-----W-----F--RQ-
	Thermoanaerobacter ethanolicus	326389845	-E-AF--A-----L-P-G	-----W---D-----KQ-
	Clostridium tetani	28209979	DE----SA-S---L-P-G	-----L-V-----F--KA-
	Mycoplasma agalactiae	291320191	-E----NA-T---FK-P-G	-----W---S---F--KR-
	Bacillus cereus	152977206	-E--F--A-T-V-FL-P-G	-----W---D-----TQ-
	Ruminococcus obeum	153810740	E-AF--G-T---FL-P-G	-----W---D-----GR-
	Eubacterium rectale	238924230	-E-----NG-T-V--E-P-G	-----W---S--N---TQ-
	Fibrobacter succinogenes	261417441	-D-----A-T-L--LYP-G	-----N--W-----KA-
	Thermobaculum terrenum	269926946	---A--ADAA---Q-L-PMG	-Q-V---HS-----RR-
	Rhodothermus marinus	268318023	---A--ADAAS---F--P-G	FK-----HS---F--S-
	Flavobacteria bacterium	225012118	-E--F--EG-S---FKY--G	-----W---S-----NQ-
	Leptospira interrogans	45658196	DE-A--A-EA---Q-E-P-G	-R-----VH-----R-
	Spirochaeta thermophila	307719119	QE----RK---F--F-P-G	-K--C-L-Y-----RR-
	Thermomicrobium roseum	221633039	DE-----F--E-P-G	-K--C-L-Y---F--RR-
	Sphaerobacter thermophilus	269836498	-E-A--A-A---L---PHG	LE-----F--GS-
	Thermus thermophilus	2664208		

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.

		113	145
Actinobacteria	<i>Actinomyces coleocanis</i>	227495317	YEGIDYRVEHYRN AGEIEV FEYSIGDYVLNGG
	<i>Intrasporangium calvum</i>	317125414	-----E--Y-YS-A DLGL-- SVV-L-----
	<i>Beutenbergia cavernae</i>	229821020	-----A--A-R-AG EPGVT- R-V-----
	<i>Mobiluncus mulieris</i>	307700237	-----A--A---E VPNV-- ----L-----
	<i>Actinomyces viscosus</i>	326771774	-----A--P---AS R G--- R-L-----
	<i>Actinomyces odontolyticus</i>	293189835	-----A--A-Y--G - GV-- V-F-----
	<i>Bifidobacterium dentium</i>	306823691	---Y-A-IPQY--A Q GVD- R-----
	<i>Bifido. adolescentis</i>	154486606	---Y-A-IPQY--A Q GVD- R-----
	<i>Gardnerella vaginalis</i>	298253302	---Y-A-IP-Y--A Q G-D- R-----
	<i>Scardovia inopinata</i>	294791384	---Y-A-IPTY--E R GLD- R-----
	<i>Parascardovia denticolens</i>	315227049	---Y-A-LPLY--K R GLN- R-----
	<i>Sanguibacter keddieii</i>	269795665	-----A--A--H-A - GGT- R-L-----
	<i>Xylanimonas cellulosilytica</i>	269955970	-----A--A---A S GRR- S-L-----
	<i>Pseudonocardia</i> sp. P1	324999609	-----E--LLDA-A R GTR- D-V-----V--
	<i>Jonesia denitrificans</i>	256832220	-----A--AD---D R G--- A-L-----
	<i>Corynebacterium glutamicum</i>	62390887	-----Q--IDDA- R YR- R-V-----I--
	<i>Thermobispora bispora</i>	296268966	-----A--DE--S R LR- D-V-----A--
	<i>Salinispora arenicola</i>	159036843	-----Q--LD-AAS R MP- T-V-L----F--
	<i>Nocardioidaceae bacterium</i>	326333672	-----Q--LD-AAT R A-- R-I-L-----
	<i>Kineococcus radiotolerans</i>	152965369	-----A-----T R VR- R-V-L-----
	<i>Cellulomonas flavigena</i>	296129320	-----A-----AA R GR- T-V-L-----
	<i>Kribbella flavida</i>	284032644	-----A--ASYAAE R MR- D-V-----
	<i>Thermobifida fusca</i>	72161069	-----A--AVDAAQ R MP- E-L-----
	<i>Nocardioidaceae bacterium</i>	326333672	-----Q--LD-AAT R A-- R-I-L-----
	<i>Propionibacterium acnes</i>	289426484	-----H--IDYA-S M WT- H-V-L-----
	<i>Kytococcus sedentarius</i>	256824952	-----E--LE-AQ R LDL RIL-L-----
	<i>Brevibacterium linens</i>	260907255	-----Q--IDWSAE H FDL RPM-----
	<i>Renibacterium salmoninarum</i>	163839831	-----E--DWA-E S FT- RPM-L-----
	<i>Mesoplasma florum</i>	50365357	---F-E-ILKYIDV -I-----T--
	<i>Spirochaeta thermophila</i>	315185657	---V-Q--LDF--A R-----SS-
	<i>Synechococcus</i> sp. JA-2-3B	86608693	---F-E-IRQLAT R-V-L--F--T--
	<i>Deinococcus deserti</i>	226355969	---F-A--GLVT R-L---F-MM--
	<i>Paenibacillus</i> sp. JDR-2	251797468	---Y-E-IR-YLVT D-L-----T--
	<i>Borrelia burgdorferi</i>	225548990	-----Q-IIDL-VD --I-----SS-
	<i>Rhodobacter sphaeroides</i>	332559519	F--V-Q--LDA-AV E-V-L--F--T--
	<i>Rickettsia prowazekii</i>	15603988	-----E--IAE-NV T-V-V---I-S--
	<i>Escherichia coli</i>	15803129	-----E--IQTEID E-W-----S--
	<i>Rhizobium leguminosarum</i>	116254278	F--V-Q--I-ARGL E-V-----S--
	<i>Yersinia pestis</i>	294505007	---V-E--IKTEID E-W-----S--
	<i>Legionella pneumophila</i>	148360986	-----E-IIN-HID E-W-L--F--S--
	<i>Pseudomonas fluorescens</i>	70728478	-----E-FIDAHVD E-W-----S--
	<i>Proteus mirabilis</i>	197284285	-----E--IQTEID E-W-V----S--
	<i>Leptospir. ferrodiazotroph</i>	251772984	---V-Q-ILDS-PF E-W-C----S--
	<i>Lentisphaera araneosa</i>	149196240	---V-Q--LDGIID R-I-----I-SN-
	<i>Alkali. metalliredigens</i>	150390504	-----Q-IIDELVT D-I-----T--
<i>Peptostrepto. anaerobius</i>	289423973	---V-Q-IIDLIVT D-V-----T--	
<i>Clostridium difficile</i>	255655267	-----E-IIDLIVT D-I-----T--	
<i>Leptotrichia hofstadii</i>	260889502	---L-Q--IDKFVD E-I-----SS-	
<i>Thermomicrobium roseum</i>	221632260	-----D--RIIL-A R-L-----T--	
<i>Burkholderiales bacterium</i>	303257146	-----Q-FLDT-VD ETL-----S--	
<i>Methylobacillus flagellatus</i>	91776684	---V-E-LLSTLVD E-----S--	
<i>Flavobacteria bacterium</i>	126663570	-K-V-Q--RDMHIT R-I-----S--	
<i>Gramella forsetii</i>	120434979	-K-V-Q--RDQFIT K-I-----S--	
<i>Salinibacter ruber</i>	294507642	-K--Q--RDLLVT R-V---V--S--	
<i>Verrucomicrobium spinosum</i>	171910506	---V-Q-I-DTCVD E-I-L-----TN-	
<i>Parachlamydia acanthamoebae</i>	282889842	-----Q-AIDREVD E-I-----TN-	
<i>Rhodopirellula baltica</i>	327537882	---F-Q--LDILQP E-I----F-----	
<i>Thermosynechococcus elongatus</i>	22299154	---V-E-----LVT Q-I----F--TC-	
<i>Chloroflexus aurantiacus</i>	163845944	-----E--R-TLID R-L-----T--	
<i>Thermobaculum terrenum</i>	269925822	---V-E--R--LAT M-L-----S--	
<i>Campylobacter jejuni</i>	205355555	-----E--L-IFA- EVF---FI-T--	
<i>Geobacter sulfurreducens</i>	298504762	-----E--RDLFVD D-F---F--T--	
<i>Terriglobus saanensis</i>	320107355	---V-E--SQMLCD R-L-----S--	

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.

		180		230
	Mycobacterium tuberculosis	15425192	GIAVGMATNIPPHNLRELADAVFWALENHDA	DEEE TLAAVMGRVKGPDFPT
	Rhodococcus equi	12137524	-----V-E-IY---D-E-	--A--E-E-----
	Nocardia farcinica	4021971	-----T---E-IY---D---	---T---C-E-I-----
	Saccharomonospora viridis	57054096	-----VV-G-I---D-P---	-----L-Q-I-----
	Tsukamurella paurometabola	96137760	-----I-E---C---Y---	-S-A--E-C-E-----
	Segniliparus rugosus	17509460	-----V-----H---	--AT--E-KEI-----
	Segniliparus rotundus	96392547	-----V---I---H---	--AS--E-KEI-----
	Gordonia bronchialis	62200052	-----V---I---HP---	-D-T---C-EAI-----
	Dietzia cinnamea	19950291	-----KM-----V-E-IY-C-D-P---	S-----C-E-I-----
	Actinosynnema mirum	56374167	-----V-EG-V---D-PE-	SD--E-MIA-I-----
	Corynebacterium glutamicum	9551262	-----N---I---L---P-	E-S-A-E-C-KF-----
	Streptosporangium roseum	71961617	-----V-G-R---Q-PE-	SD-L-E-LIA-----
	Saccharopolyspora erythraea	34096627	-----V-G-V---PE-	-D-A-L-E-LIE-I-----
	Gardnerella vaginalis	97243225	-----M-V-G-H---DHP--	SK-L-E-LIQ-I-----
	Amycolatopsis mediterranei	00781944	-----VSEG-V---PE-	TDD-L---LLV-I-----
	Mobiluncus mulieris	27876547	-----R-----I---AL-Y---PE-	SR--L-E-A-Q-I-----
	Kineococcus radiotolerans	52963979	-----V-EGAK-L-D-P--	SD-Q---LLT-I-----
	Thermobifida fusca	2160413	-----V-EG-Y-Y-DHP--	SD-L-D-LIE-I-----
	Cellulomonas flavigena	96127876	-----V-AG-Q-H---HP-	SK-L-E-L-A-I-----
	Stackebrandtia nassauensis	91297545	-----K-----V-A-N-C-D-TE-	----E-LIKL-Q-----
	Thermomonospora curvata	69124284	-----VTEG---Y-D-YHV	SD-L-D-LIE-I-----
	Acidothermus cellulolyticus	17927217	-----V-AGA-Q---HP--	TR-L---LLE-I-----
	Nakamurella multipartita	58650279	-----V-S-V-F---PE-	SP-L-E-S-E-IQ-----
	Xylanimonas cellulolytica	69954817	-----V-EG-R-H-AHPE-	SR-L---L-Q-----
	Bifidobacterium animalis	83602674	-----M-V-G-H---DHPE-	SR-L-D-LIE-I-----
	Kocuria rhizophila	84199652	-----V-G-Q-Y-Q-P-V	-R-T-L-EELIK-----S
	Rothia mucilaginosa	55326476	-----V-EG-E-F-K-PH-	TN--L-N-L-A-I-----
	Streptomyces coelicolor	437481	-----V-AGA-Q-Y---YE-	SH-L-D-LIE-I-----
	Beutenbergia cavernae	29818509	-----V-EG-E-F---HP-	TN--L-G-LLL-I-----
	Kribbella flavida	84028010	-----Q-----V-A-AN-C---P-	SQ-V---C-EHI-----
	Jonesia denitrificans	56831261	-----V-EG-A-Y-A-P-	SR--LRE-LIA-I-----
	Tropheryma whipplei	8492973	-----A-VSAGAL-C---P-	TD--LTE-LLE-I-----
	Salinispora tropica	45592575	-----K-----IGA-Q-C---HPE-	--AS--E-LL-I-----
	Frankia alni	11219513	-----V-TG-Q---DHAE-	TD--L-E-LI-II-----
	Kitasatospora setae	11897313	-----V-SGAL---HPG-	SN--L-E-LVE-I-----
	Propionibac. freudenreich	97625215	-----VNE-Q-F---HPE-	SD--L-E-S-A-----
	Thermobispora bispora	96268005	-----TGE-Q---HPEL	SD-L-EGLIE-I-----
	Renibacterium salomoniarum	63838776	-----V-EGA-Q-Y-A-PT-	SK--L-E-LIQ-I-----
	Nocardioideaceae bacterium	26329137	-----V-SGAK---HP-	TR--LQD-LIE-----
	Catenulispora acidiphila	56389239	-----S---VN-G-Q-F-Q-PE-	TN--L-E-LIE-I-----
	Micrococcus luteus	89705901	-----M-V-AG-Q-Y---HPE-	TR--L-E-LLA-H-----
	Micromonospora aurantiaca	02864518	-----K-----IGA-Q-C---HPE-	--AT--E-LLEI-----
	Actinomyces coleocanis	27494194	-----R-----V-EG-Q-Y-DHPE-	SR--L-N-LI-I-----
	Sanguibacter keddieii	69793365	-----V-EG-Q-Y-DHP--	SP--V-DSLIRII-----
	Leifsonia xyli	0953932	-----V-QGAL-H-AHPE-	SR--L-EELIK-I-----
	Parascardovia denticolens	94786214	-----M-V-QG-H---DHPE-	SR--L-NNLIAII-----
	Brevibacterium linens	60904976	-----V-AGA-Q-L-SHPE-	SK--A-E-LL-I-----
	Arthrobacter aurescens	19961287	-----V-EG-Q---D-PT-	TR--L-E-LLL-I-----
	Kytococcus sedentarius	56823912	-----Q-----V-EGCQ-L-D-PE-	TK--LQEKLLQVI-----
	Clavibacter michiganensis	48271185	-----V-SGAQ-L-AHP-	NR--L-E-LLE-I-----
	Eubacterium rectale	38922438	-----VVN-VRLID-DIE	-K-T--DELIDV-----
	Synergistetes bacterium	95111080	-----VT-VLC-L---EGIN	PA-A-S--DL-E-LP-----
	Campylobacter jejuni	8564562	-----S-N--I-GLLYL-D-K--	S-EEI-QFI-----
	Helicobacter pylori	61839476	-----S---KID-II--LIHV--PN-	E-DEILEF-----
	Thermotoga maritima	645226	-----S-TV--LIYLDHPE-	--VEEL-QFI-----
	Bacillus subtilis	19444321	-----Q-G-II-G-LAVS--P-I	--ISEL-EFIP-----
	Fusobacterium nucleatum	9705415	-----G-V-GILAIID-K-I	EILEL-NYI-----
	Dehalococco. ethenogenes	7233707	-----A--CQ-ICYLID-P-C	GWDDL-QF-S-----
	Pseudomonas aeruginosa	5598364	-----G-VI-GCLALMD-P-L	--VDEL-QYIP-----
	Haemophilus influenzae	09972999	-----N-VLNGCLAYID-NEI	--IDEL-QHIP-----
	Coxiella burnetii	09364105	-----N-IIN-TLALI--P-L	--NVEEL-RHIP-----
	Burkholderia mallei	21598728	-----N-VV--CLYL-N-PQ-	--SVDELIEITPA-----
	Ralstonia eutropha	13866815	-----N-IV-GCLHL-R-PQ-	--VDELIELTIPA-----
	Rickettsia prowazekii	5603946	-----H-C--LlyLD-PQ-	--GINDI-NFI-----
	Dictyoglomus thermophilum	06901654	-----G--I--LlyLI-KPE-	--EIEELNFI-----
	Chloroflexus aggregans	19849213	-----G-C--ISYLDHPE-	--VEEL-EIIP-----
	Leptospirillum rubarum	24516515	-----IT-II-GLLALID-P-L	--SVDDLAI-H-----
	Thermobaculum terrenum	69926182	-----IC-GLIYLDHP--	--IEDLSKI-Q-----
	Deferribacter desulfuricans	91278437	-----T-VI--LVYIID-PSN	--DEEILNFI-----
	Thermomicrobium roseum	21632669	-----G-VV--LVYLD-PE-	--VEELVE-LP-----
	Bacteroides fragilis	3713702	-----M-----S-VI--CEAY-D-K-V	--VEEL-EY--A-----
	Thermovibrio ammonificans	19790113	-----IS-VCK--QYLDHPE-	--TEEL-QFI-----
	Synechococcus sp. BL107	16071053	-----N--I-GLMALIN-PEI	--SDQEL-TLIP-----
	Chlamydomphila pecorum	30444524	-----G--IE-TLLL-S-PT-	--SIEEI-EVMP-----

Actinobacteria

Other Bacteria

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.

		83	83	122
	Streptomyces coelicolor	21221466	NIFSTQDHAAAAIAV	GPNGTPDNP
	Thermobispora bispora	296268748	-----VV-	QGVPVFAWKGETLEEY
	Thermomonospora curvata	269128253	-----VV-	-----
	Kitasatospora setae	311896336	-----VV-	-----ED-
	Thermobifida fusca	72162904	-----VV-	--D-----R-----
	Frankia alni	111226112	-----V-	-----K-D-----
	Streptosporangium roseum	271962934	-----VV-	-----V-D-----
	Gordonia bronchialis	262203513	-----V-	--H-----E-K-----
	Acidothermus cellulolyticus	117927677	-----VV-	--D---Q- R-----I---
	Mycobacterium tuberculosis	215432212	-----VV-	--H---E- K-----
	Rhodococcus jostii	111023285	-----V-	--H---EE- --T-----
	Tsukamurella paurometabola	296138919	-----E---VV-	-----EA-----
	Amycolatopsis mediterranei	300789821	-----V-	--H---EE- K-----S---
	Jonesia denitrificans	256833056	-----E-----	--T-----S-----S---
	Corynebacterium diphtheriae	38233299	-----E---VV-	--H---ED-----
	Kribbella flavida	284029568	-----VV-	--D--A-A- A-----
	Saccharomonospora viridis	257056680	--Y---E---VV-	-R---K- E-----
	Actinosynnema mirum	256380309	-----VV-	--H---EE- R-----S-A--
	Dietzia cinnamea	319949854	-----E---VV-	--E---EE- R-----
	Saccharopolyspora erythraea	134102884	-----D-S---VV-	--H---E- R-----
	Salinispora arenicola	159036558	-----VV-	--E-S-QA- A-----S---
	Micromonospora aurantiaca	302865499	-----V-	--E---EA- A---Y-----P--
	Nocardia farcinica	54026586	-----VV-	--H--VEE- K--S-----
	Geodermatophilus obscurus	284992665	-----V-	-D---EQ- A-----
	Salinispora tropica	145593523	-----VV-	--E-S-EAL A-----S---
	Arthrobacter aurescens	119961459	-----E---VV-	-K---E- L-----
	Aeromicrobium marinum	311743632	-----VV-	-RD--AAD- K-T-----A--
	Janibacter sp. HTCC2649	84498166	--Y---E---VV-	--E--VED- ---Y-----P--
	Amycolatopsis mediterranei	300787136	-----E---VV-	--E--V-A- A-TS-----
	Nakamurella multipartita	258654379	-----VV-	-KY--VEE- K-T-----P--
	Micrococcus luteus	289704918	-----E---VV-	-S---ED- ---N-S--D-
	Stackebrandtia nassauensis	291300565	-----E---VV-	-RE--VAA- K-T-----
	Kineococcus radiotolerans	152967781	-----Q---VV-	-S--VEE- A-----
	Renibacterium salmoninarum	163841741	-----E---TV-	-S--I-A- A-----
	Brevibacterium linens	260906683	-----E---VV-	-T-SVGA- A-----
	Saccharopolyspora erythraea	134100432	-----D-A---VV-	-----EK- E-TS-----
	Verrucospora maris	330467562	--Y---E---VV-	-Q--VGQ- A---Y-----S-R-
	Catenulispora acidiphila	256396703	-----V-	-AG--EA- A-----
	Nocardioideaceae bacterium	326334180	-----VV-	-K---ED- --T-----A--
	Catenulispora acidiphila	256395873	-----E---VV-	--D---GK- T-TS-----
	Sanguibacter keddiei	269794250	-----E---VV-	--H--VED- ---S---
	Geodermatophilus obscurus	284992665	-----V-	-D---EQ- A-----
	Segniliparus rugosus	317508026	-----VV-	--H--V-E- K-----D--
	Pseudonocardia dioxanivorans	331698938	-----TV-	--T---EA- A-----
	Verrucospora maris	330465911	-----V-	--D---A- A---Y-----Q--
	Intrasporangium calvum	317125631	-----E---VV-	--H---ED- -----
	Conexibacter woesei DSM	284046058	-----V-	-----E- -----
	Clavibacter michiganensis	170781092	-----E---V-	-AG--EA- A-----
	Streptomyces pristinaespiralis	297196593	-----A	A-I-----
	Genmatimonas aurantiaca	226229258	-----V-	-----V-K- --T-----
Other bacteria with insert	Anaeromyxobacter dehalogenans	220915295	-----V-	-RE---EA- R-----
	Fibrobacter succinogenes	261415685	-----N---VV-	-K--SVS-- ---S---
	Desulfuromonas acetoxidans	95928711	-----E	T-----Y--S---
	Xylella fastidiosa	9105979	-----T	S-T-----
	Coxiella burnetii	164685840	-----L-Q	K-I-I-----E---
	Waddlia chondrophila	297621233	--Y-----H	-----M-E---
	Synechococcus sp. RCC307	148241205	-----E	R-I---Y---D--
	Legionella pneumophila	54294919	-----S--A	K-I---Y---E---
	Oceanicaulis alexandrii	83945799	-----D	K-I---Y-----
Other Bacteria	Opiritatus terrae	182415662	-----K	A-----
	Chlorobium ferrooxidans	110598853	-----K	A-----D--
	Acidobacterium sp. MP5ACTX8	299135590	-----A	A-----
	Parabacteroides johnsonii	218259482	-----A	D-----
	Leptothrix cholodnii	171056810	-----L--	--T---Y---AD-
	Victivallis vadensis	281357863	-----A	D-----
	Denitrovibrio acetiphilus	291287351	-----A	T-----S---

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.

		266	317
Actinobacteria	Streptomyces griseus	182435840	IAAKAVSFKIAAGEEFKERQQRRLD GARILAERL VQPDV TEVGVSVLSSGGTD
	Renibacterium salmoninarum	163840469	--G--A-----S-----V-E----- --D-- KAK-I--VT----
	Nocardia farcinica	54026787	-----A--GT--RD-----S--K----- TGA-- ADK-I---T----
	Bifidobacterium dentium	283456143	--G--A--V--S---D-M-----K----- TAD-- KNN-I---T----
	Amycolatopsis mediterranei	300783738	-----AL---SD--R--E--E-S---A-- S-D-C ASA--R--T----
	Streptomyces avermitilis	29829317	-----V--SDD-----E-V-E-----I-D-- --RD-- KA---D-----
	Arthrobacter aurescens	119962515	-----A---G-----E-V-E-----I-D-- N-S-- A-A---T-----
	Kocuria rhizophila	184201446	--G--A-----S-G-A---R--IE--Q----- TA--L A-H-----T----
	Actinomyces viscosus	326774134	-----AM--GT--R--E--VR--A-I----- GAD-- KAA---LVT----
	Saccharomonospora viridis	257055377	-----AL---S---R-----V-E--K----- SAS-C ASA--R--T----
	Dietzia cinnamea	319950195	-----TAM--GT-Q-R-----E-K----- TAE-C RNA-----T----
	Corynebacterium diphtheriae	38233528	-V---AM-----T---D-----Q-I----- TGA-C KAA--D--T----
	Aeromicrobium marinum	311742185	-----A-M-LEPG-R--E--IE---V-D----- MAA-A RDA--Q-----
	Rothia mucilaginosa	283458676	-----TA--V--SDA--D--K--VE--Q--N-- L-Q-- KDA---IA----
	Thermonospora curvata	269128045	-----AL---S---R---A---E--K----- LAA-C AKA--K--T----
	Gordonia bronchialis	262201618	-----AL--GT-----S--K----- TGD-- SKA----T----
	Clavibacter michiganensis	170781021	-----TA---T---AD--R--IQ--Q----- --AA-S --A---T-----
	Intrasporangium calvum	317126680	-----LV-GSA--QD--R--E--K--D-- SAD-T RAA-IK-----
	Xylanimonas cellulosilytica	269955562	V-----A-----S-S--D--E--R--Q-I-S-- SE--- AAA-A---T----
	Sanguibacter keddieii	269796075	V-----A-----S---A---H-V-R--K-I-D-- TAA-- --A---T-----
	Brevibacterium linens	260907101	-----A--L--S---A--E--VR--Q----- LAD-- ANA-AT--T----
	Saccharopolyspora erythraea	134101028	--G--L--L-----RD--R--E--K----- LAD-A ARA--RLV-----
	Leifsonia xyli	50955412	-----TA--L--TD---D--A--IR--QL----- TAA-S RAS--D--T----
	Jonesia denitrificans	256831934	-----A---GSAD-V---E-V-R--K-I----- TGA-- ADA----T----
	Nocardioidaceae bacterium	326332989	--G--A--V--EP--R--E--R---I----- LAD-- AAA---V-----
	Micrococcus luteus	289705986	--G--A---GSAD---K-E--A--Q----- TAD-M A-H-I---T----
	Conexibacter woesei	284046077	--G--A---MSDS-----E--IA--QA--TE- LAD QSS-----T----
	Mycobacterium tuberculosis	260184925	-----TA--M--QP--AQ---C-----G-- T----- A-R-IA--T----
	Cellulomonas flavigena	296130338	-----TA--V-GTP--RD--E--R---V----- SRQ-A KDA--A-R-----
	Geodermatophilus obscurus	284991328	-----A--L--GEPA-R--E--A-----D-- LTA-S R-A-IN-V-----
	Mobiluncus curtisii	315655719	V-----AL--GTP-----M--VIQ--Q----- MA--C QAA-IDL-T----
	Nakamurella multipartita	258650905	-----A--M--DPS-T--AE--L--L----- I-S-- AQA--E--T----
	Kitasatospora setae	311898457	-----S-----R--E--K----- L-D-- --AS-----
	Saccharomonospora viridis	257055897	-----A--L--QP--RT--EH--S--K----- L R-EN-G-V-----
	Amycolatopsis mediterranei	300789378	-----A--M--EP--A-----R--KL----- L A-PDI---T----
	Pseudonocardia sp. P1	325002235	-----A-RL-GTPA-A-----QLI-S-- L D-S--G-V-----
	Rubrobacter xylanophilus	108803318	-----ALR--HT-G-RA--RQ-VAN-KA---A- MQN-IE-V-----
	Thermoanaerobacter wiegelii	307265278	-----C--E-LSD---Y-K-IVEN-KA--NA- M-R-INLV-----
	Clostridium thermocellum	125973573	-----AEVLTDD---QY--QIVKN-KT--NA- M-K-IDLV-----
	Nitrosomonas eutropha	114331539	-----A--E--SPA--DY-QQVIEN--VM-RV- QQR-LRIV--H--
	Burkholderia pseudomallei	76581166	-----A--E-LSP---Y--KVVEN--V--QT- VKR-LRIV--R-E
	Aquifex aeolicus	15605959	-----A--E-MS-----YAKQVVEN--V--E- KKY-FKIVT----
	Persephonella marina	225850656	-----A--E-MT---R-YAHQ-VKN-KV--E- KAE-LRIV-----
	Thermotoga maritima	15643483	-----C--E-MT-----Y-QQVVKV-KKM--EF QKR-YRIV-----
	Thermotoga petrophila	148269354	-----C--E-MT-----Y-QQVVKV-KKM--EF QKR-YRIV-----
	Fervidobacterium nodosum	154248784	-----C--E-LSD---AY-NQIVKN-KA--KA- ENR-LRIV-----
	Pelobacter propionicus	118580168	-----A--E-LQP---QY--QVVNN--T---E- VKR-FKLT-----
	Neorickettsia risticii	254796604	---R--A-GE-LTT---DYIRAVVRN-KT--NV- R-R-FD-----
	Zymomonas mobilis	56552097	-----A-GE-LQPS---YAKAVVEN-QA--A-- K-R-SDLVT----
	Rhodopseudomonas palustris	90424158	-----A--E-LQPD--VYAKNVVEN-KA---T- RGH-FDIV-----
	Pseudomonas putida	303306209	-----C--E-LEP---AY--QVIEN-QAM-KVF IDR-YD-V-----
	Bacillus clausii	56965621	-----A-GE-LSD---TYAKQIIAN-KR-G-K- QAE--DIV-----
	Psychrobacter cryohalolentis	93005149	-----C--E-LE-N---TY--QVVKN-QAM-KVI QDR-YEII-----E
	Thermus scotoductus	320451242	--G--A-FE-LQP---YSRLVVEN-KR-G-E- AKR-YRLVT----
	Fusobacterium mortiferum	237737799	-----A--E-LTP--I-Y-QQIVKN-QT--KV- ENG-LRIV-----
Thermodesulfo. yellowstonii	206890824	-----A--E-LS-D---Y-KKVIKN-KT--A- KKK-FKLV-D---	
Denitrovibrio acetiphilus	291287821	-----A--E-LSD---Y-QQIVKN-KT--KV- ADR-FRIV-----	
Anaerobac. hydrogeniformans	289522796	-----LT--L-MT---AYGAQVVKN--VM-DV- KNN-FDIV-----	
Thermobaculum terrenum	269925153	-----AL-E-LQP---Y-RQIVKN-ST--QS- -KH-FNLV-----	
Synechococcus sp. RS9916	116075611	-----A-GE-LTDD--AYSQVVAN-QA--KQI QAR-ID-V-----	
Cyanobium sp. PCC 7001	254432489	-----A-GE-LQPS--RAYS-QVIAN-QA--A-I Q-R-ID-V-----	
Archaea	Methanosarcina mazei	21226544	-----A--E-MS---RQD-DQ-VKN-KV-CSC- KQK-FDIV-----
	Methanohalophilus mahii	294496562	-----A--E-LSDK--QD-KQ-VKN-KA-CAN- IDRFDIVA----
	Methanosaeta thermophila	116754542	-----A--E-MTP--RRY-EQIVRN-AA--D-- I-N-FDLV-----

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
<i>Mycobacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Tsukamurella</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Gordonia</i>	+	+	+	+	+	+	+	+	+	+	+	-	+
<i>Nocardia</i>	+	+	+	+	+	+	+	+	+	+	+	-	+
<i>Rhodococcus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Segniliparus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Amycolicococcus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Corynebacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Nakamurella</i>	+	+	-	+	+	+	+	+	+	+	+	+	+
<i>Pseudonocardia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Saccharopolyspora</i>	+	+	+	+	+	+	+	+	+	+	+	-	+
<i>Actinosynnema</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Saccharomonospora</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Amycolatopsis</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Geodermatophilus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Stackebrandtia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Verrucosispora</i>	+	+	+	+	+	-	+	+	+	-	+	-	+
<i>Micromonospora</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Salinispora</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Frankia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Acidothermus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Streptosporangium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Thermomonospora</i>	+	+	+	+	+	+	+	+	+	+	+	-	+
<i>Thermobifida</i>	+	+	+	+	+	+	+	+	+	+	+	+	-
<i>Nocardiopsis</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Catenulispora</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Streptomyces</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Propionibacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Kribbella</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Nocardioides</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Kineococcus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Janibacter</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Kytococcus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Brachybacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	-
<i>Brevibacterium</i>	+	+	+	+	+	+	+	+	+	-	+	+	+
<i>Intrasporangium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Isoperitcola</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Microbacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Rothia</i>	+	+	+	+	+	+	+	+	+	+	+	+	-
<i>Kocuria</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Micrococcus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Renibacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Arthrobacter</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Beutenbergia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Cellulomonas</i>	+	+	+	+	+	+	+	+	+	+	-	+	+
<i>Xylanimonas</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Jonesia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Sanguibacter</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Mobiluncus</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Actinomyces</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Arcanobacterium</i>	+	+	+	-	+	+	+	-	+	+	+	+	-
<i>Gardnerella</i>	+	-	+	-	+	+	+	+	+	+	+	+	+
<i>Bifidobacterium</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Tropheryma</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Leifsonia</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Clavibacter</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>marine actino*</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Acidimicrobium</i>	-	+	-	+	+	+	-	-	-	-	-	-	-
<i>Conexibacter</i>	+	+	-	+	-	-	-	-	-	-	-	-	-
<i>Rubroacter</i>	+	+	+	+	-	-	-	-	-	-	-	-	-
<i>Olsenella</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Slackia</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Eggerthella</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Cryptobacterium</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Coriobacterium</i>	-	-	-	-	-	-	-	-	-	-	-	-	-

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

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--
	Tsakamurella 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--
	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	-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-
	Arthrobacter aurescens	119963962	--G--E--KLV-Q--LGAQ	N----FE-----V--
	Micrococcus luteus	289706398	--G--E--KIV-G--LGED	G----VFHS-----V--
	Propionibacterium acnes	295130820	--Q--E--RM-Q--LGAQ	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--LGAQ	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--ERVFDMSKV--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-

Corynebacteriales

Other
Actinobacteria

Supplemental File 9

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

		387	427
<i>Corynebacteriales</i>	<i>Corynebacterium jeikeium</i>	68537193	RRVGAGVLSAGHARALLGLK S
	<i>Corynebacterium diphtheriae</i>	38234909	-K-A-----V- A
	<i>Corynebacterium genitalium</i>	300780165	--A-----V- V
	<i>Corynebacterium matruchotii</i>	225020855	QK-A-----AIN E
	<i>Corynebacterium glutamicum</i>	145297088	TK-A-----A A
	<i>Corynebacterium efficiens</i>	259508322	TK-A-----I- G
	<i>Tsukamurella paurometabola</i>	296141900	--A-----A-E G
	<i>Gordonia bronchialis</i>	262204655	--A-----S-E T
	<i>Mycobacterium avium</i>	15808976	--A-----S-E A
	<i>Mycobacterium kansasii</i>	240168401	--A-----S-E A
	<i>Mycobacterium gilvum</i>	145221429	--A-----S-D G
	<i>Mycobacterium tuberculosis</i>	219555764	--A-----S-E A
	<i>Mycobacterium marinum</i>	183985451	--A-----S-E A
	<i>Rhodococcus opacus</i>	226362886	--A-----S-E A
	<i>Rhodococcus jostii</i>	111020643	--A-----S-E A
	<i>Rhodococcus erythropolis</i>	226309509	--A-----S-E A
	<i>Rhodococcus equi</i>	296038613	--A-----S-E A
	<i>Segniliparus rotundus</i>	296395439	-K-A-----D G
	<i>Nocardia farcinica</i>	54027640	--A-----E A
	<i>Corynebacterium</i> subclade	<i>Corynebacterium ammoniagenes</i>	296118597
<i>Coryne. pseudogenitalium</i>		227490380	K-A--T-----S-D
<i>Coryne. tuberculostearicum</i>		255324014	K-A--T-----S-D
<i>Corynebacterium aurimucosum</i>		227834356	K-A--T--G-----P
<i>Corynebacterium accolens</i>		227502234	K-A--T-----I--D
<i>Corynebacterium striatum</i>		227506194	K-A--T--G-----I--P
<i>Coryne. glucuronolyticum</i>		227487670	-K-A-R-I-S----I--A
<i>Leifsonia xyli</i>		50955949	T-A-----I-S-G D
<i>Kytococcus sedentarius</i>		256826451	--A-----E
<i>Kineococcus radiotolerans</i>		152968445	--A--I-----S-E
<i>Janibacter sp.</i>		84497202	--A-----VS
<i>Beutenbergia cavernae</i>		229822691	--A-----S
<i>Actinomyces urogenitalis</i>		227496613	--A-----P
<i>Mobiluncus curtisii</i>		298345824	Q-A-----A-D
<i>Renibacterium salmoninarum</i>		163842269	--A-----S-P
Other <i>Actinobacteria</i>	<i>Brevibacterium mcbrellneri</i>	295394854	--A----Q-----N
	<i>Kocuria rhizophila</i>	184201996	--A-----S-----A-T
	<i>Micrococcus luteus</i>	289706709	--A--T-----T
	<i>Kribbella flavida</i>	284034932	--A-----P
	<i>Sanguibacter keddieii</i>	269797060	--A-----P
	<i>Cellulomonas flavigena</i>	296131536	--A-----G
	<i>Xylanimonas cellulolytica</i>	269958140	--A--I-----S-S
	<i>Jonesia denitrificans</i>	256833759	--A-----S
	<i>Propionibacterium freudenreich</i>	297627566	--A-A-----A-D
	<i>Actinosynnema mirum</i>	256381068	--A-----D
	<i>Frankia sp. Eu11c</i>	280960972	--A-----A-E
	<i>Frankia sp. EAN1pec</i>	158319049	--A-----S-E
	<i>Amycolatopsis mediterranei</i>	300791158	--A-----S-E
	<i>Bifidobacterium longum</i>	46190724	K-A-----S-----S
	<i>Bifidobacterium bifidum</i>	224283944	KK-A-----N
	<i>Gardnerella vaginalis</i>	297243212	KH-AS-----E
	<i>Parascardovia denticolor</i>	294786227	KK-AS-----A-P
	<i>Streptomyces coelicolor</i>	6539744	N-A-----SVD
	<i>Brachybacterium faecium</i>	257070284	--LAS-AI-----S-D
	<i>Acidothermus cellulolyticus</i>	117929361	-K-A--T-----A-N
	<i>Rothia dentocariosa</i>	296934401	--A--IA-----T-T
	<i>Nakamurella multipartita</i>	258655503	--A-----S-P
	<i>Saccharopolyspora erythraea</i>	134103810	--A-----S-D
	<i>Saccharomonospora viridis</i>	257057909	--A-----S-D
	<i>Salinispora tropica</i>	145597095	--A-----S-D
<i>Geodermatophilus obscurus</i>	284993432	T-A--II-----P	
<i>Thermomonospora curvata</i>	269129159	--A--T-----S-D	
<i>Nocardioopsis dassonvillei</i>	297563776	T-A-----T--KVE	
<i>Thermobifida fusca</i>	72163509	-K-A-R-I-----AVE	
<i>Nocardiooides sp.</i>	119718910	--A-----S--AIE	
<i>Aeromicrobium marinum</i>	293168377	--A-----VTVA	
<i>Arthrobacter aurescens</i>	119962086	--A-----A-P	
<i>Clavibacter michiganensis subs</i>	148274155	H-A-----I-SSG	
<i>Arcanobacterium haemolyticum</i>	297572336	--A-----A	
			GPTEEQLRATRVVSEGLSVR
			-EDT-AE--Q-IIA-----
			--A--Q--D-I-A-----
			DQ-G-----N-I-----
			-EDA-DT-----IIA-----
			-E-A-DA--S-IIA-----
			-T-A--A--A-IIA--M---
			-SDA--A--A-I-A--M---
			--A--E--S-I-A-----
			--A--E--S-I-A-----
			--V--E--A-I-A-----
			--A--E--S-I-A-----
			--A--E--S-I-A-----
			-ADA--VM-A-I-A--M---
			-ADA--VM-A-I-A--M---
			-ADA--VM-A-I-A--M---
			-ADA--V--A-I-A--M---
			HT-A--D--S-IIA-----
			--DA--A--A-I-A--M---
			--A-----S-I-A-----
			DKDVMK--S-I-A-----
			DT-AM-YI-N-IIA-----
			DT-AM-YI-N-IIA-----
			DA-AM-LI-N-I-A-----
			DK-AMDYI-D--I-----
			DS-AM-IV-N--IA-----
			DKGQA-A-----A-----
			DQKGMVH--DKI-N-D----
			DGAAM--M-Q--A--M---
			DGAAM----Q-I-A-----
			DNAAM----Q-I-A-----
			DGAAM----Q-I-A-----
			DAAAM----Q-I-A-----
			D--AMTQ--Q-I-A-----
			DSAAI----Q-I-A--M---
			D-AQM-E--Q-I-A-----
			D-ADM----Q-I-A-----
			D-SDM----Q-I-Q-----
			SGDAI----Q-I-A-----
			DGAQI----Q-I-A-----
			DGAAI----Q-I-----
			DGA-I----Q-I-----
			QTADM-K--Q-I-----
			D-LA-----Q-I-A-----
			D-GA--D--I-A--M---
			DT-A-D-----I-A-----
			D-DA-D-----I-A-----
			DA-S--E--A-I-A--M---
			D--MDK--S-IIA-----
			DES-IDK---IIG-----
			N--DM-S--K-IIA-----
			SA-QM-S--K--A-----
			D--D--H-I-A-----
			D-ALM-E--Q-I-Q-----
			DATA--E--Q-I-A-----
			DQAKI-TI-QKI-N-----
			D-DD--Q--A-I-A-----
			E-GG--E--S-I-A-----
			D-DQ--E--G-I-A-----
			D--A-----H-I-A-----
			DAGK--A--A-I-A--M---
			S--A--A--H-I-A-----
			D-DL-D--A--E--I---
			D--L-DH--H-IIT-----
			D--L-D--Q--A--I---
			N--I-----A-----
			DAAAM--M-QKI-A--M---
			DD-AMRH--EKI-N-D----
			DAAAM----Q-I-A-N----

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 Orotidine-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.

		41	72
<i>Mycobacterium</i>	<i>Mycobacterium smegmatis</i>	118470834	LRAFCDICVAAFAGFA IVKPQVAFFEAYGSAG
	<i>Mycobacterium abscessus</i>	169629910	-AR-S---E---DV- -SH-A--
	<i>Mycobacterium intracellulare</i>	254818277	-A-----IE--S--- V-----A--
	<i>Mycobacterium gilvum</i>	145224324	-SR-A-L--T----- V-----
	<i>Mycobacterium avium</i>	118466105	-A-----E----- V-----A--
	<i>Mycobacterium leprae</i>	15827193	-A-----E--S--- -A--
	<i>Mycobacterium intracellulare</i>	254818277	-A-----IE--S--- V-----A--
	<i>Mycobacterium kansasii</i>	240170761	-A-----Q-Y---- V-----A--
	<i>Mycobacterium ulcerans</i>	118617401	-A----- -A--
	<i>Mycobacterium marinum</i>	183982213	-A----- V-----A--
	<i>Mycobacterium vanbaalenii</i>	120403649	-AG-S----- V-----
	<i>Rhodococcus equi</i>	296035881	-EQ-SE---E--V-QV A L-----G-
	<i>Nocardia farcinica</i>	54025588	-ER-AE---E--D-TV A L-----V----
	Other <i>Actinobacteria</i>	<i>Corynebacterium striatum</i>	227504411
<i>Corynebacterium diphtheriae</i>		38233916	-QR-TE---E--S-NV A L-----Y--W-A--
<i>Corynebacterium jeikeium</i>		260578793	-EE-SKT--E---DV- C V-----Y-QF----
<i>Tsukamurella paurometabola</i>		296140279	V--GLR-AE---EV G F-----QF-A--
<i>Gordonia bronchialis</i>		262202269	--R-AET-TE-LGPV- A VI-----F----
<i>Salinispora tropica</i>		145594402	-ER-AGTV-E-LGDRV A V---S---RF--R-
<i>Micromonospora aurantiaca</i>		270501558	-ER--ATV-E-IGDRV A V---S---RF-AR-
<i>Saccharopolyspora erythraea</i>		134098655	-ER-ALTA-E-L--EV A VL---S---V-----
<i>Saccharomonospora viridis</i>		257055590	-ER-ALTATE-L--TV A VL---S---A--
<i>Actinosynnema mirum</i>		256379247	-ER-ALT--E--GAE- A V---S---H--R-
<i>Nakamurella multipartita</i>		258654393	-ER-AMTV-E-LGPTV A VL---S---H-A--
<i>Gardnerella vaginalis</i>		297533045	AEL-SMRMLQ-MN-R- A A--F-SSM--R--K-
<i>Clavibacter michiganensis</i>		148272963	-EE-GLRV-E-T--R- G -----RH----
<i>Bifidobacterium longum</i>		296454535	AEL-SMRMLQ-AN-R- A A--F-TSM--R--AK-
<i>Leifsonia xyli</i>		50954779	V-E-GLRV-D-A--R- G T-----RF-A--
<i>Beutenbergia cavernae</i>	229820530	V-E-SLRV-E-LG-RV A VF---S--Y-RH----	

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.

		318	346
<i>Rhodococcus</i> <i>Nocardia</i>	Rhodococcus erythropolis	229489553	GYYGAGTVEYLVQGD TISFLEVNTRLQVE
	Rhodococcus jostii	111023245	----- -V-----
	Rhodococcus opacus	226365754	----- -V-----
	Rhodococcus equi	296037236	----- -V-----
	Nocardia farcinica	54022956	-----E -V-----
<i>Other</i> <i>Actinobacteria</i>	Corynebacterium jeikeium	68536759	-----GS- G L-----
	Corynebacterium striatum	227505378	-----GS- G L-----
	Corynebacterium diphtheriae	38233256	-----GS- G L-----
	Mycobacterium leprae	581342	-----GQ- G L-----
	Mycobacterium tuberculosis	260206637	-H-----GQ- G L-----
	Gordonia bronchialis	262201773	-----F--GS- G LV-----
	Segniliparus rotundus	296392950	--H-----GE- G LV-----
	Tsukamurella paurometabola	296138760	-----F--AA- G LV-----
	Micromonospora aurantiaca	270503371	--H-----GV- G -----
	Salinispora tropica	145593407	--H-----GV- G -----
	Saccharomonospora viridis	257054651	--S-----GS- G -----
	Actinosynnema mirum	256380383	--H-----GV- G -----
	Stackebrandtia nassauensis	291298309	-H-----F--GA- G -----
	Nakamurella multipartita	258654355	--H-----F--GE- G SV-----
	Thermobispora bispora	296268677	--V----C-F--GQ- G -----
	Micrococcus luteus	289707027	--T-----AP- G L-----
	Propionibacterium acnes	50843179	--V--A-C-F--GR- G -----
	Streptomyces coelicolor	4731924	--V-----F--GM- G -----
	Frankia alni	111220670	--V-----F--AR- G M-----

Supplemental File 13

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

		111		142
<i>Rhodococcus</i>	<i>Rhodococcus jostii</i>	111024124	LVRRFEEQVRRSHPLS	GGG AEGTLSEGIAAERD
	<i>Rhodococcus opacus</i>	226366364	-----	GGG -----
	<i>Rhodococcus equi</i>	296035850	-I---H-----Q	SDS K-----A
	<i>Rhodococcus erythropolis</i>	226306513	-I-----Q	NDG ID-----R
	<i>Brevibacterium linens</i>	260903777	-----ST--P--Q	G---L---EG--S
	<i>Leifsonia xyli</i>	50954814	-----P--Q	GN---LD--T--A
	<i>Janibacter sp.</i>	84496558	-----S--P--Q	G--R-LD--QH--E
	<i>Xylanimonas cellulositytica</i>	269956344	-----A--P--Q	GD-RILD--D--R
	<i>Jonesia denitrificans</i>	256832485	-----P--Q	GD-RILD--TI--S
	<i>Streptomyces svicens</i>	297198889	-----S--P--Q	GD-RIVD--D--E
	<i>Streptomyces avermitilis</i>	29832834	-----S--P--Q	GD-RIVD--D--E
	<i>Streptomyces albus</i>	239978850	-----S--P--Q	GD-RIVD-----
	<i>Thermobispora bispora</i>	296269969	-----DS--P--Q	GD-R-VD--NR--G
	<i>Saccharopolyspora erythraea</i>	134098709	-I-----G--Q	G--R-LD--QH--S
	<i>Actinosynnema mirum</i>	256379212	-I---S--G--MQ	-D-R-AD--E--V
	<i>Actinomyces urogenitalis</i>	227496079	-I---SS--P--Q	GT--ILD--H--T
	<i>Mobiluncus mulieris</i>	227875311	---Y---P--Q	GK-RILD--Q--G
	<i>Kytococcus sedentarius</i>	256825144	-----S--P--VQ	GT-R-LD--RT--R
	<i>Beutenbergia cavernae</i>	229820662	-----P--Q	GD-RILD--E--R
	<i>Arcanobacterium haemolyticum</i>	297571458	--Q-Y-S--P--Q	GD-R-LD--Q--E
	<i>Streptosporangium roseum</i>	271967321	-----N--P--Q	G--R-VD--R--G
	<i>Thermomonospora curvata</i>	269126429	-----A--P--Q	GD-R-V---TQ--E
	<i>Thermobifida fusca</i>	72162419	-----S--P--Q	GD-R-TD--SR--E
	<i>Nocardiosis dassonvillei</i>	297561955	-----G--P--Q	GD-R-TD--DR--E
	<i>Renibacterium salmoninarum</i>	163841015	-----G--P--Q	GD-RMLD-----E
	<i>Sanguibacter keddieii</i>	269795309	-----P--Q	G--RILD--GI--Q
	<i>Rothia dentocariosa</i>	296935637	I-A-Y-AQ--P--Q	G--RVLD--R--S
	<i>Brachybacterium faecium</i>	257068681	-----DS--P--Q	G--EGVL---RR--E
	<i>Arthrobacter chlorophenicus</i>	220912580	-----G--P--Q	GG-RILD-----E
	<i>Acidothermus cellulolyticus</i>	117928318	-I---N--P--Q	GD-R-AD--R--E
<i>Tsukamurella paurometabola</i>	296140243	-I-----K--Q	GS---A-----H	
<i>Corynebacterium matruchotii</i>	252126372	-IK--DNL--T--Q	GS---LV--ER--A	
<i>Mycobacterium leprae</i>	15827214	---Y--N-----Q	G-Q--A-----R	
<i>Cellulomonas flavigena</i>	296129790	-----CP--Q	GD-RILD--GH--S	
<i>Geodermatophilus obscurus</i>	284990548	---Y-SN--E--Q	GT---ID--TE--A	
<i>Frankia symbiont</i>	289642289	----DH--P--Q	GD-RVVD--SR--T	
<i>Kribbella flavida</i>	284031185	I--Q-S--P--Q	G--R-LT--QR--E	
<i>Salinispora arenicola</i>	159038869	-I---S-----Q	G--R-AD--V--G	
<i>Micromonospora sp.</i>	238060138	-I---S-----Q	G--R-AD--V--G	
<i>Arthrobacter aurescens</i>	119962185	-----G--P--Q	GG-RILD--GI--E	
<i>Micrococcus luteus</i>	289705481	-L---AG--P--MQ	GNAR-LD--R----	
<i>Nakamurella multipartita</i>	258652975	-I--Y-AN--K--Q	GS-R-AD-V----A	
<i>Saccharomonospora viridis</i>	257055609	-I--D--G--MQ	GD-R-VD--T--K	
<i>Gardnerella vaginalis</i>	297586027	-IK-Y-S--P--Q	RG-R-ID--LE--E	
<i>Bifidobacterium animalis</i>	219683846	--K--S--P--Q	QGNR-ID--HE--R	
<i>Bifidobacterium angulatum</i>	229817561	--T-Y-K--P--Q	HGNS-ID--KE--Q	
<i>Scardovia inopinata</i>	294790862	-IK-Y-S--P--Q	KGRR-ID--HE---	
<i>Segniliparus rotundus</i>	296393041	-I-----R--Q	GQD--A--KR---	
<i>Nocardioides sp.</i>	119716753	---Q-AA--P--Q	GG-R-ID-LVR--G	
<i>Clostridium bolteae</i>	160939917	-IK-YKET-----A	GT-R-DK--EK--V	
<i>Thermoanaerobacter tengcongens</i>	20808250	-IK--KET--R----	E--RIVD--NE--K	
<i>Pelotomaculum thermopropionicu</i>	147679064	----KES--P----	TH-EILDV-RE--N	
<i>Clostridium butyricum</i>	182418131	-IK--KET-----	PD-RVLT--TL---	
<i>Lactobacillus casei</i>	116494466	--S-YKET-----A	M--R-MD--KK--A	
<i>Bacillus cellulosityticus</i>	283846450	---YKET--T--A	GG-PPLV--E--N	
<i>Lactobacillus rhamnosus</i>	199597266	--S-YKET-----A	M--R-MD--KK--V	
<i>Streptococcus thermophilus</i>	55820880	--S-YKET-----A	-D-RVLD--KL--E	
<i>Alicyclobacillus acidocaldarii</i>	218290409	----YKAS--R----	MG-R-L-S-Q---Q	
<i>Streptococcus gallolyticus</i>	288904937	--A-YKET-----A	-D-RILD--KL--E	
<i>Anaeromyxobacter dehalogenans</i>	86156576	----YSET--R--A	G--GTVPDGI-AER	
<i>Leptospirillum rubarum</i>	124514205	-----SET--P--L	RGEGTPVIE-IREE	

Other
Actinobacteria

Other Bacteria

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.

		9		40
<i>Corynebacterium</i>	<i>Corynebacterium glutamicum</i>	145296751	SGSSSIKFQLVNPENS	AID EPYVSGLVEQIGE
	<i>Corynebacterium efficiens</i>	25029146	-----A-H	-T- D-F-----
	<i>Corynebacterium matruchotii</i>	225022579	-----D-DQH	-T- --FA-----
	<i>Corynebacterium aurimucosum</i>	227834109	----V----D--S-	-T- T-L-----V--
	<i>Corynebacterium genitalium</i>	300780388	-----I-D-TA-	-A- T-F-----
	<i>Corynebacterium accolens</i>	304567156	-----D-TA-	-T- D-L-----V--
	<i>Corynebacterium striatum</i>	227505902	----V----D--GT	-T- A-L-----V--
	<i>Corynebacterium diphtheriae</i>	38234611	-----D-TAH	-T- D-FA-----
	<i>Corynebacterium pseudotubercul</i>	300859247	-----D-TQH	-T- A-FA-----
	<i>Corynebacterium ammoniagenes</i>	296119015	-----I-D-ASA	-T- --LA-----
	<i>Corynebacterium amycolatum</i>	213965190	-----D--KD	STQ P-----R---
	<i>Corynebacterium lipophiloflavu</i>	227547862	-----I-D-TS-	-S- TAF-----K---
	<i>Corynebacterium jeikeium</i>	68535316	-----VLD--AN	-TA D-F--I--K---
	<i>Corynebacterium glucuronolytic</i>	227540876	-----I-DSDAD	-TE D-F--V-----
	<i>Corynebacterium resistens</i>	300934586	-----VLD-KQG	-LE A-FI--I--K---
	<i>Corynebacterium urealyticum</i>	172041417	-----VLD-SAD	-HE --FIT-I--K---
	Other <i>Actinobacteria</i>	<i>Rhodococcus erythropolis</i>	226304962	-----H-DTA
<i>Rhodococcus jostii</i>		111019189	-----IH-DTG	KASA----R---
<i>Rhodococcus opacus</i>		226361325	-----IH-DTG	KSWG---I-R---
<i>Rhodococcus equi</i>		296037777	-----H-VAG	-SVGT---R---
<i>Mycobacterium intracellulare</i>		254818870	----V----D-DSG	TALST---R---
<i>Mycobacterium avium</i>		41409984	----L----D--FG	VAAST-I--R---
<i>Nocardia farcinica</i>		54027319	-----Y--D--TG	-VSAA---R---
<i>Arcanobacterium haemolyticum</i>		297571864	-----Y-----TG	-SIA---R---
<i>Beutenbergia cavernae</i>		229821529	----L-Y----SGG	-AVA--I--R---
<i>Scardovia inopinata</i>		294790761	-----Y----L-SG	-GLA----K---
<i>Cellulomonas flavigena</i>		296130415	-----Y---I-DTH	-VLA----R---
<i>Actinomyces coleocanis</i>	227494859	----L-Y-----G	AAIA----R---	

Supplemental File 15

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

		161		201	
<i>Corynebacterium</i>	<i>Corynebacterium lipophiloflavu</i>	227549231	AGCMPVLVGWAAVRDNV	DDGSPDR WWQAVVLFMIFFWTPP	
	<i>Corynebacterium glucuronolytic</i>	227488866	-----M----VIV--T	QGAAN -----LV-----	
	<i>Corynebacterium matruchotii</i>	252126357	-----V---VIT--L	PPDMPHQ -----L-----	
	<i>Corynebacterium genitalium</i>	258560496	-----M---VI----	FDGSPDR -----L-----	
	<i>Corynebacterium diphtheriae</i>	38233887	-----V---VIT--L	PAGTDAQ ----I----V-----	
	<i>Corynebacterium glutamicum</i>	19552785	-----V---VIV-QF	EPGVDDQ ----I----V-----	
	<i>Corynebacterium ammoniagenes</i>	296117942	-----A----VI---A	MDGQPDR -----L-----	
	<i>Corynebacterium jeikeium</i>	260578763	-----V---VIT--N	GGAFGAGWSS -A--LI--V-----	
	<i>Corynebacterium urealyticum</i>	172040680	-----M-----T--N	GGAFNAGWSS -I--II--V-----	
	<i>Corynebacterium aurimucosum</i>	227833033	-----AM---VI---A	PASEPDR -----L-----	
	<i>Corynebacterium amycolatum</i>	213965606	-----M---VIT--N	PEITGLAH ----FI--LL-----	
	<i>Corynebacterium pseudogenitali</i>	227491495	-----AM---VI---A	PAGEPDR ----I--L-----	
	<i>Corynebacterium accolens</i>	227503587	-----AM---VI---A	PAGEPDR ----I--L-----	
	<i>Corynebacterium striatum</i>	227504382	-----AM---VI---A	PAGEPDK ----I--L-----	
	<i>Corynebacterium kroppenstedtii</i>	237785549	-----I---VIA--T	SNHSVAG -L---A-----	
	<i>Corynebacterium efficiens</i>	25028249	-----A----VIV-QF	PEGQPQQ ----I-----	
	Other <i>Actinobacteria</i>	<i>Leifsonia xyli</i>	50954828	-----I--V-TNS-	G-AP-I--GV--L----
		<i>Clavibacter michiganensis</i>	170782337	-----I-----TGDI	S-AP-I----V-L----
		<i>Streptomyces coelicolor</i>	21220420	---L---I--S--TNSM	S-AP-I--GVM-----
		<i>Streptomyces sp.</i>	295839556	-----I--S--T-S-	S-A--I--LV-----
		<i>Streptomyces albus</i>	239978836	-----I--S--TNT-	S-A--I--LV-----
		<i>Nocardiopsis dassonvillei</i>	297561915	-----I-----TNSL	D-APF---LVV-----
		<i>Thermobifida fusca</i>	72162400	-----I-----ITERL	D-APFI--LVV-----
		<i>Thermomonospora curvata</i>	269126449	-----I-----TESL	AAAPF---VV-W----
		<i>Kribbella flavida</i>	284031202	---F-T-I--TS-TNEL	A-TPL---GVV-----
		<i>Kineococcus radiotolerans</i>	152966865	---Q--I--T---SL	D-APF---GV--L----
		<i>Janibacter sp.</i>	84496541	-----T-I--S--TGS-	G-P--I--LV-----
		<i>Saccharopolyspora erythraea</i>	134097899	-----VI-----TGS-	E-P-L-M-GVV-L----
		<i>Aeromicrobium marinum</i>	293166736	---F-A-I--T--TNEL	A-AP---LVV-----
		<i>Tropheryma whipplei</i>	28493306	-----IA-----TGSI	D-P-I--AVV-L----
		<i>Xylanimonas cellulositytica</i>	269956357	-----FI-----TESL	S-A--A--G-----
		<i>Beutenbergia cavernae</i>	229820648	-----I--S--TGGL	D-A-L--GV-----
		<i>Cellulomonas flavigena</i>	296129777	-----VI--S--TGGI	S-E-CL--GVV-----
		<i>Nocardioides sp.</i>	119716768	---F-A-I--T--TGSL	S-VP---AVV-----
		<i>Propionibacterium acnes</i>	289428641	---F-P-I--T--TGS-	R-EPL---F-V-W----
<i>Saccharomonospora viridis</i>		257055629	-----VI-----TGT-	E-P-F-M-GV-----	
<i>Jonesia denitrificans</i>		256832497	-----FI--S--TETL	S-G--M--LV-----	
<i>Arthrobacter sp.</i>		116670645	---F---IA---T-T-	E-P-I---V--L-----	
<i>Brevibacterium mcbrellneri</i>		295395296	-----I--S--TGSL	ALEPFI--LVV-----	
<i>Nakamurella multipartita</i>		258652473	-----II--S--TGSI	G-P-W-F-GV-----	
<i>Actinosynnema mirum</i>		256379192	-----VI--S--TGT-	ELP-LAM-GV-----	
<i>Kytococcus sedentarius</i>		256825158	---I---I--S--TNSL	D-P-L---VFV-----	
<i>Micrococcus luteus</i>		239917646	---F---I-----GT-	E-P-I---LVV-L----	
<i>Kocuria rhizophila</i>		184200866	-----I-----GTL	E-P-F---AF--L----	
<i>Brevibacterium linens</i>		260903767	-----I--S--TGGL	A-E-LL--LVV-----	
<i>Sanguibacter keddieii</i>		269795296	-----FI--S--TGGL	S-G--A--LV-----	
<i>Segniliparus rotundus</i>		296392902	-----A-----TGSL	D-P-L-V--V-----	
<i>Gordonia bronchialis</i>		262202302	-----T-----TGSL	S-EP---GV-----	
<i>Rhodococcus opacus</i>		226366377	-----M---S--TGSL	S--PI--LV-----	
<i>Nocardia farcinica</i>		54025542	-----A---S--TGTI	G-P-LA--GV-----	
<i>Tsukamurella paurometabola</i>		296140221	-----T-----ATGTI	Q-P-I---LV-----	
<i>Mycobacterium avium</i>		118462366	-----MI--S--TGTI	Q-P-L-M-A-----	
<i>Mycobacterium tuberculosis</i>		289745200	-----MI--S--ITGTI	A-P-LAM-A-----	
<i>Frankia sp.</i>		158313910	---F---I--S--TGT-	G-P--L--AVV-----	
<i>Frankia symbiont</i>		289642762	---F---I--S--TGTI	G-P--L--AV-----	
<i>Acidothermus cellulolyticus</i>		117928335	---F---I--S--TGR-	G-P-----AV--A----	
<i>Rhodospirillum centenum</i>		209966878	--AF-PMI----TGD-	TLAS-I--A-----	
<i>Brucella abortus</i>		189023702	--AF-PMI----ATGEI	T-DSL-----L----	
<i>Rickettsia canadensis</i>		157803916	S-AL-PVI-Y--SNTI	SL-S-I--L--LI-----	
<i>Thermosynechococcus elongatus</i>		22299436	--AI-P-----TGEL	S-A-W---GM--V-----	
<i>Nodularia spumigena</i>		1219509021	--AI-A-----TGTL	S-A-WLI-A-V-L-----	
<i>Thiobacillus prosperus</i>	187424069	--AV-PVL--C--SGH-	SYESI---S-----		
<i>marine gamma proteobacterium</i>	119476742	--AL-P-I-----TNTI	SHEGIL-VL---V-----		

Supplemental File 16

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

		179		223
<i>Corynebacterium</i> Clade I	<i>Corynebacterium glutamicum</i>	19553540	GKTGMGVPELLDKVVELIPAP	TS EFEEAPARAMIFDSVYDTRG
	<i>Corynebacterium aurimucosum</i>	227833666	---E-----A-----	--G-----V-----
	<i>Corynebacterium efficiens</i>	25028798	---V-----	SP GPS-----
	<i>Corynebacterium diphtheriae</i>	38234335	---E-E-----A-----	-T DYPD-----
	<i>Corynebacterium tuberculostear</i>	255325525	---E-V-----C-V-P-	S--DA-----
	<i>Corynebacterium accolens</i>	227503095	---E-V-----C-V-P-	S--Y-P-----
	<i>Corynebacterium glucuronolytic</i>	227488363	---E-E-----R-CD-V--	--HGA-----L-----
	<i>Corynebacterium matruchotii</i>	225021715	---E-EA-----V-P-	-T DYPM-----
	<i>Corynebacterium striatum</i>	227504499	---E-ED-----A-V-P-	ST -APD-----
	Other <i>Corynebacterium</i>	<i>Corynebacterium urealyticum</i>	172041033	---E-----R-C-V-H-
<i>Corynebacterium jeikeium</i>		68535649	---E-S-----R-C-V--	VGDA-----I---
<i>Corynebacterium kroppenstedtii</i>		237785922	A--D-----LCDVV--	TGNA-----L-----
<i>Corynebacterium amycolatum</i>		213966173	---E-EQM--R-C-V--	VGDP-----L-----
<i>Gordonia bronchialis</i>		262203067	---A-T-----E-IK-V--	QGDP-----
Other <i>Actinobacteria</i>	<i>Nocardia farcinica</i>	54023366	---V-----E-IRQV-P-	VGDA-----
	<i>Mycobacterium abscessus</i>	169628730	---E-AA--E-R----	QGDP-----I---
	<i>Rhodococcus erythropolis</i>	226307248	---K--E-IRQV--	VGDA-G-----A---
	<i>Rhodococcus opacus</i>	226360412	---K--E-RQV--	VGDP-G-----A---
	<i>Tsukamurella paurometabola</i>	296140484	---E-GA--R-I-RV--	VGDP-----
	<i>Segniliparus rotundus</i>	296392919	---E-K--E-R-V-P-	KGDA-----F----
	<i>Geodermatophilus obscurus</i>	284990151	---E-----QI-AQ--	VG-AE-----I---
	<i>Pseudonocardia thermophila</i>	190148908	---F-S-----V-P-	VGDP-----
	<i>Micrococcus luteus</i>	289705015	---V-E--R-RA-G-	-GDA-----
	<i>Kocuria rhizophila</i>	184200951	---E-E--ERI-RDV-P-	QGVK--S-----
	<i>Rothia dentocariosa</i>	296935533	---E-EA--RI-G--	VGQA-T-----S---
	<i>Thermobifida fusca</i>	72161233	A--E-E--NEI-AR--P-	VGNA-G--L-----
	<i>Bifidobacterium dentium</i>	171743010	---E-A--RI-TD----	KG-P--L-----S---
	<i>Frankia alni</i>	111221440	---Q-----NEI-RQV--	VGPNPAG-----I---
	<i>Saccharomonospora viridis</i>	257055314	A--E-R--Q--QV-P-	-GDS--P-L-----
<i>Saccharopolyspora erythraea</i>	134098032	A--E-SDV--E-RQV--	-GVH-S-P-----	
<i>Nakamurella multipartita</i>	258651785	---A-K-----AE-P-	VGDA-----I---	
<i>Actinosynnema mirum</i>	256375354	A--L-GD--E-RKV--	VGDAE-----	
<i>Streptomyces coelicolor</i>	21221021	A--L-DV-----AEV--	VGVK-----S---	

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	<i>Corynebacterium</i> Subclade-1
NCgl0032	NP_599285.1	Hypothetical	270	<i>Corynebacterium</i> Subclade-1
NCgl0152	NP_599407.1	Hypothetical	247	<i>Corynebacterium</i> Subclade-1
NCgl0559	NP_599820.1	Hypothetical	101	<i>Corynebacterium</i> Subclade-1
NCgl0750	NP_600012.1	Hypothetical	139	<i>Corynebacterium</i> Subclade-1
NCgl1091	NP_600364.1	Hypothetical	165	<i>Corynebacterium</i> Subclade-1
NCgl1554	NP_600830.1	Hypothetical	154	<i>Corynebacterium</i> Subclade-1
NCgl1838	NP_601119.1	Hypothetical	179	<i>Corynebacterium</i> Subclade-1
NCgl1867	NP_601149.1	Hypothetical	179	<i>Corynebacterium</i> Subclade-1
NCgl1870	NP_601152.1	Hypothetical	206	<i>Corynebacterium</i> Subclade-1
NCgl1914	NP_601195.1	Hypothetical	272	<i>Corynebacterium</i> Subclade-1
NCgl1970	NP_601251.1	Hypothetical	126	<i>Corynebacterium</i> Subclade-1
NCgl2001	NP_601281.1	Hypothetical	132	<i>Corynebacterium</i> Subclade-1
NCgl2018	NP_601298.1	Hypothetical	75	<i>Corynebacterium</i> Subclade-1
NCgl2022	NP_601302.1	Hypothetical	400	<i>Corynebacterium</i> Subclade-1
NCgl2036	NP_601317.1	Hypothetical	202	<i>Corynebacterium</i> Subclade-1
NCgl2177	NP_601457.1	Hypothetical	142	<i>Corynebacterium</i> Subclade-1
NCgl2225	NP_601506.1	Hypothetical	188	<i>Corynebacterium</i> Subclade-1
NCgl2226	NP_601507.1	Hypothetical	173	<i>Corynebacterium</i> Subclade-1
NCgl2043	NP_601325.1	Hypothetical	77	<i>Corynebacterium</i> Subclade-1
NCgl2536	NP_601826.1	Hypothetical	162	<i>Corynebacterium</i> Subclade-1
NCgl0048	NP_599301.1	Hypothetical	157	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0077	NP_599330.1	Hypothetical	186	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0097	NP_599350.1	Hypothetical	183	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0122	NP_599376.1	Hypothetical	172	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0138	NP_599393.1	Hypothetical	128	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0189	NP_599445.1	Hypothetical	163	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0193	NP_599449.1	Hypothetical	134	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0366	NP_599625.1	Hypothetical	140	<i>C. glutamicum</i> and <i>C. efficiens</i>
NCgl0404	NP_599666.1	Hypothetical	94	<i>C. glutamicum</i> and <i>C. efficiens</i>

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

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.

		710	746	
<i>Corynebacteriales</i>	Tsakamurella paurometabola	296139429	ALAARSRKRVSLLEDLDAALKE	TSQLNLIILKGDNSGTV
	Rhodococcus equi	296036127	-----I-----	HNE-----
	Mycobacterium leprae	15827818	-----I-----S----	-----A----
	Mycobacterium tuberculosis	260206150	-----I-----S----	-----A----
	Gordonia bronchialis	262202086	-----I-----S----	-----A----
	Segniliparus rotundus	296394255	E---KG-R-I-----SM---	-----
	Nocardia farcinica	54026037	-----I-----	--E-----A----
	Rhodococcus equi	296036127	-----I-----	HNE-----
	Rhodococcus opacus	226366092	---K---I-----S----	-----
	Corynebacterium efficiens	25028434	-----S-----SV---	--V-----A-S-
<i>Pseudonocardiales</i>	Corynebacterium diphtheriae	38234058	---KT-----SV---	--T-----A-S-
	Saccharomonospora viridis	257055492	-EN-AR-----S----	-NT---I-----
	Saccharopolyspora erythraea	134102373	-QN-AK-----KV---	-N-----I-----
	Actinosynnema mirum	256379831	-N-AR-----	---T-I-----
	Geodermatophilus obscurus	284992350	-QN-SM---I-----	NR---I-----
	Micromonospora aurantiaca	270501203	-SF-N--G-AT--T-MEQ---	GE KTS---V---V--S-
	Salinispora arenicola	159036966	F-N--G-AT--T-MEQ--A	GE KTS---V---V--S-
	Nakamurella multipartita	258652417	-QIGN -R-I---FGK--A-	QK IE--T--I-----
	Leifsonia xyli	50954463	-QL-KA---I---FTR--E-	GK VEA---I---V--A-
	Tropheryma whipplei	28572760	-QL-K-----FTR-IQ-	GR VQS--M-I---V--A-
Other <i>Actinobacteria</i>	Arthrobacter aurescens	119964199	-AL-KR---I---F-Q-VA-	GK IDT-----V--A-
	Micrococcus luteus	239917229	-QL-KR---IT---F-Q-VA-	GK LDT---I---A--A-
	Sanguibacter keddieii	269795639	KR---I---FTQ--QQ	GK VET---V---V--A-
	Kineococcus radiotolerans	152965411	-AL-KA---IT---FTK--QQ	GK VET-----GA-S-
	Arcanobacterium haemolyticum	297571195	-TL-KR-----D--N--IE-	GK IEN---I---S--S-
	Mobiluncus mulieris	269977477	--L-KR---T--SF-E--A	GK VDM---I---V--S-
	Actinomyces odontolyticus	293192216	--L-KR-----F-KV-A-	GE VDT---VI---V--A-
	Streptomyces coelicolor	32141270	-AF-KRTR-----KV--A	GE VQ-----I--A--S-
	Acidothermus cellulolyticus	117928722	-QL-A--R-RT---ILERMEK	GE VAE-R-----V--S-
	Frankia sp. CcI3	86742243	-EL-L--G-PT--TILERM--	GE KT-----V--S-
	Thermobifida fusca	72161182	-QQ-KASR--T-DNWQKT-E-	GQ REE-L--I---M--S-
	Streptosporangium roseum	271963648	-DM-K-GR-RT--E-FSEMEK	GQ VDE-K--I---V--S-
	Bifidobacterium longum	213693051	-QL-KR--V-----FKKKFA-	SE IDM--IVI---S--S-

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.

		25		162
	<i>Corynebacterium glutamicum</i>	19551752	RVAFRMRKAIQSAMRQ	P QVKGIKVVCSSRLGGAEM
	<i>Corynebacterium efficiens</i>	25027084	-----	-----
	<i>Corynebacterium jeikeium</i>	68536917	-----G-----	-----
<i>Corynebacteriales</i>	<i>Tsakamurella paurometabola</i>	227978463	-----	N---R-Q-----
	<i>Rhodococcus equi</i>	295819037	-----	N---R-Q-----
	<i>Rhodococcus jostii</i>	111023102	-----	N---R-Q-----
	<i>Gordonia bronchialis</i>	262203618	-----	N---R-Q-----
	<i>Mycobacterium tuberculosis</i>	167967949	-----	N---R-Q-----
	<i>Nocardia farcinica</i>	54022706	-----S-----	N---R-Q-----
<i>Pseudonocardiales</i>	<i>Actinosynnema mirum</i>	256380592	-----S-----	S-----R-Q-G-----
	<i>Saccharomonospora viridis</i>	257054476	-----TS-S-----	-----R-Q-----
	<i>Nakamurella multipartita</i>	258651369	--S-----SM---Q-S	-----RIQ-----
	<i>Geodermatophilus obscurus</i>	284992868	--S-----M---Q-S	-----R-Q-----T--
<i>Micromonosporales</i>	<i>Salinispora tropica</i>	145596428	--S-----M---KN	VC---R-QV-----
	<i>Micromonospora aurantiaca</i>	270503787	--S-----M---KN	AC---R-QV-----
<i>Glycomycetales</i>	<i>Stackebrandtia nassauensis</i>	291298716	--N-----S---L-N	-T-----A-----
	<i>Thermobifida fusca</i>	72163039	-----	KS GA---R-Q-----
Other <i>Actinobacteria</i>	<i>Streptosporangium roseum</i>	271962638	--S-----M---KS	GA---R-Q-----
	<i>Frankia sp. EUN1f</i>	288917236	--S-----M---KG	GA---R-Q-----
	<i>Micrococcus luteus</i>	239918193	-----K-----A	GAQ--RIQ-----
	<i>Rothia mucilaginosa</i>	255326848	-----K-----Q-A	GA---RIQ-----
	<i>Catenulispora acidiphila</i>	256390146	-----S-KA	G-----QV-----
	<i>Streptomyces coelicolor</i>	21223088	--S-----SM-GT-KA	GA---IQ-G-----
	<i>Actinomyces coleocanis</i>	227494551	--S-----M---A	GA---R-Q-----
	<i>Beutenbergia cavernae</i>	229821617	-----GM---Q-A	GA--VR-Q-----
	<i>Kineococcus radiotolerans</i>	152964663	--S-----GM-TT-S	GA---R-Q-A-----
	<i>Nocardiopsis dassonvillei</i>	229205450	-----M-T-KS	GA---RIQ-G-----
	<i>Bifidobacterium longum</i>	23336507	--T-----Q-D---A	GA---RIKL-----
	<i>Conexibacter woesei</i>	284042810	-----KR-LT---S	GA--V--QV-----
	<i>Propionibacterium acnes</i>	282855202	-----Q---NA	GAL--RIK-----
	<i>Acidothermus cellulolyticus</i>	117927521	--S-----L---LKA	GA--VR-QVA-----
	<i>Clavibacter michiganensis</i>	148273795	-----GL-G-Q-A	GA--VRIQV-----
	<i>Leifsonia xyli</i>	50955560	-----GL-G-Q-A	GS--VRIQV-----
	<i>Cellulomonas flavigena</i>	229241980	--S-----G---Q-A	GA---R-QV-----
<i>Jonesia denitrificans</i>	256831833	--S-----GM---L-A	GA---R-QY-----	

Supplemental File 20

Partial Sequence alignment of the protein Alpha-ketoglutarate 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*.

		619	650	
Corynebacteriales	Segniliparus rotundus	296393143	VKYHLGAEGTYFQMFGD NEIKVSLTANPSHLE	
	Tsakamurella paurometabola	296138938	-----K-Y----- E ---T---V-----	
	Nocardia farcinica	161598437	-----L----- - - - - -	
	Gordonia bronchialis	262201823	-----K-Y----- - - - - -	
	Rhodococcus opacus	226365480	-----I----- - -D-T-----	
	Rhodococcus jostii	111022975	-----I----- - -D-T-----	
	Rhodococcus erythropolis	226307609	-----S---I----- - -D-A-----	
	Corynebacterium jeikeium	260577589	-----QQ-H-H----- - -D-Y-A-----	
	Corynebacterium diphtheriae	38199857	-----HI----- - - - - -	
	Corynebacterium striatum	227503791	-----S---HHI----- G-DIT-A-----	
	Mycobacterium leprae	161723268	-----T-L-L----- - -D-Q-----	
	Mycobacterium tuberculosis	294994806	-----T-L-L----- - -D-Q-----	
	Mycobacterium bovis	161511534	-----T-L-L----- - -D-Q-----	
Pseudonocardiales	Nakamurella multipartita	258651662	-----K-IRP--- E GQVD-T-C-----	
	Saccharomonospora viridis	257057028	-----K-R--- - -G-T-----	
	Saccharopolyspora erythraea	134102819	-----K-R--- - -G-T-----S-----	
	Actinosynnema mirum	256380243	-----K-R--- - -G-T-----	
Glycomyetales	Stackebrandtia nassauensis	291302596	-----MT-KFTTSD E HATT--VA-----	
	Micromonospora aurantiaca	270502237	-----QN-KFTTPD E HAV---VV-----	
Miocromonosporales	Salinispora tropica	145596202	-----QN-KFTTPD E HSV---VV-----	
	Acidothermus cellulolyticus	117927794	-----R-TTRD T KD-----VS-----	
Other Actinobacteria	Leifsonia xyli	50955456	-----T---FHGAG E--P-Y-A-----	
	Clavibacter michiganensis	148272502	-----T---FRGVH E-MP-Y-A-----	
	marine actinobacterium	88856384	-----T---FTAAD RQ-P-Y-A-----	
	Kineococcus radiotolerans	152965196	-----T---FTGED EQT--YVA-----	
	Arthrobacter aurescens	119963962	-----T---FTSDN KQT--Y-A-----	
	Renibacterium salmoninarum	163840137	-----T---FTSFN --T--Y-A-----	
	Rothia dentocariosa	296935259	-----T---FTSTS KQTQ-Y-A-----	
	Brevibacterium linens	260907538	-----T---SFTSPA -STD-YVA-----	
	Brevibacterium mcbrellneri	295394626	-----M---SFTSPD -TT--VA-----	
	Kocuria rhizophila	184200591	-----T---VFSDN --TQ-Y-A-----	
	Micrococcus luteus	289706398	---M-T---FVSDA -STR-Y-A-----	
	Streptomyces coelicolor	21223647	-----FTGLD EQ----V-----	
	Streptomyces scabiei	290957458	-----FTGLD EQ----V-----	
	Streptomyces albus	239982262	-----FTGLD EQ----V-----	
	Catenulispora acidiphila	256396666	-----S-SFTALD SS-A-T-A-----	
	Sanguibacter keddiei	269794283	-----T---FTSES ATT--Y-A-----	
	Kytococcus sedentarius	256825652	-----TS---FTDDE --VG-Y-A-----	
	Beutenbergia cavernae	229819770	-----T---M-TSPT EQTS-Y-A-----	
	Cellulomonas flavigena	296129070	-----T---VFTAES ATTA-Y-A-----	
	Arcanobacterium haemolyticum	297570946	-----T---V-TSPK DKVG-Y-A-----	
	Xylanimonas cellulolytica	269957271	-----T---FTAES ATT--H-A-----	
	Jonesia denitrificans	256833025	-----T---FTAES ATT--A-----	
	Propionibacterium freudenreich	297626222	-----T-QFTA-S EK----A-----	
	Nocardioides sp.	119715972	-----EFEAGSK DR----VA-----	
	Kribbella flavida	284033067	-----EFVSE-- DK----VA-----	
	Frankia sp. CcI3	86742428	-----V-TGQD RTVP--VV-----	
	Frankia alni	111225327	-----V-TGRE RKVP--VV-----	
	Acidimicrobium ferrooxidans	256372497	---K-F--H-RTAS S-LP-T-AS-----	
	Thermobifida fusca	72160970	-----T---FTTIT EK-AI--A-----	
	Nocardiopsis dassonvillei	297562611	-----T---FETHD QK-RI--A-----	
	Streptosporangium roseum	271969479	-----T-DFVSPD SKL-T-VV-----	
	Thermomonospora curvata	269127876	-----DFESH-- AK-R--V-----	
	Mobiluncus mulieris	227876254	-----T-SV-ESYE CHTQ-YIA-----	
	Actinomyces odontolyticus	293189468	-----TW-V-SLDD LAT--YMA-----	
	Other Bacteria	Roseiflexus castenholzii	156741514	-----HRAFREAGI EQMPIT-AP-----
		Chloroflexus aggregans	219850562	-----RKA-RESGI A-MPIT-AP-----
		Opitutaceae bacterium	225165118	-----Y-AALETTT KT-E-R-A-----
		Coralimargarita akajimensis	294054624	-----Y-TKRTTHD H-VEIR-A-----
	Verrucomicrobium spinosum	171913443	-----FQTKRRTKS--VMID-A-----	

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 *Frankia* species.

		28		60		
(A)	<i>Frankia</i>	Frankia sp. Eu11c	280964332	LRAWQRAALDRYR	AAS GAGSRDFLAVATPGAGK	
		Frankia sp. EAN1pec	158316891	-----ET--	SR- AS-G-----	
		Frankia alni	111221585	-----V--	SR- SS-A-----	
		Frankia sp. EUN1f	288919015	--E-----EI--	SRT AS-G-----	
		Frankia symbiont of Datisca	289641351	-----E--AI--	-RA -S-A-----	
		Frankia sp. CcI3	86740086	-----EI--	SR- SS-A--M-----	
	Other <i>Actinobacteria</i>	Kytococcus sedentarius	256824901	----SE--E--Q	SLA AS-G-----	
		Kineococcus radiotolerans	152967715	----D--E--F	DPE RDEPA---VT-----	
		Corynebacterium glutamicum	19553118	-----NFL	ANKP-----	
		Nocardia farcinica	54025739	-----R--TK-L	ATKP-----	
		Rhodococcus erythropolis	226306297	-----R--TK-L	SSSP-----	
		Gordonia bronchialis	262202175	-----R--TK-L	TRRP-----	
		Mycobacterium avium	118462616	--G--R--VK-L	AGQP-----S--	
		Tsukamurella paurometabola	296139553	--V--R--TK-L	T-KPQ-----	
		Xylanimonas cellulositytica	269955851	----AE--EL--	AKEP-----	
		Jonesia denitrificans	256832119	----AE--EV--	ERQP-----	
		Catenulispora acidiphila	256390677	----Q---E-M	-REP-----	
		Streptomyces coelicolor	5802839	----QG-MEK-L	QDQP-----	
		Saccharomonospora viridis	257056429	-----R--TK-L	TRKPK-----	
		Actinosynnema mirum	256375639	-----R--TK-L	AS-PK-----	
		Thermobispora bispora	296268916	----QE-F-L-F	RRDP-----	
		Arthrobacter aurescens	119962220	----QE---L-M	TRNPK-----	
Micrococcus luteus		239917353	--Q--QE--E--L	ETAPQ---T-----		
Kocuria rhizophila		184200641	----AE-IQ--E	Q-QP--M-----		
Renibacterium salmoninarum		163840210	----SE--HK-F	DGDRQ-----		
Actinomyces urogenitalis		227495762	----A--KK-L	ETMPE-----		
Thermobifida fusca		72162780	--H--E-FEE-F	RREP-----		
Nocardioopsis dassonvillei		297559356	-----E-FEE-F	RREP-----		
Streptosporangium roseum		271963257	----QE-F-L-S	RREP-----		
Leifsonia xyli		50954856	----AE---A-F	ATEP-----A		
Propionibacterium acnes		50842382	--T--E---S-L	ATSP--W-VC-----		
Micromonospora aurantiaca		270501464	-----K--VE-L	RRRTE--T-----		
Salinispora arenicola		159037075	-----R--VE-L	RRREP--T-----		
(B)		<i>Frankia</i>	Frankia sp. EUN1f	288918321	GIREVADPSGLFLSR	SGGA GPTSVPGTCVTVTVEGRR
			Frankia sp. EAN1pec	158312223	-----	--AA-----
	Frankia symbiont of Datisca		289642342	--G-----	WHD-G PMDG-A-----A	
	Frankia alni		4231547	--HGI-----	--AG -LEA-----	
	Frankia sp. CcI3		3907225	--HGI-----	S--- -LEA-----	
	Other <i>Actinobacteria</i>	Bifidobacterium angulatum	229818203	--E--S-----S	N-D- -S-P-E-----F-LD-H	
		Bifidobacterium dentium	171741228	--E--S-----S	N -SVP-E-----F-LD-H	
		Corynebacterium efficiens	259505925	--K--P-----H	RGSTPD--A--AMD-V-	
		Corynebacterium diphtheriae	38234541	----P-----H	HGSTPD--A--AMD-V-	
		Rhodococcus opacus	226363785	--AGIS-----HH	RAEA---A--MMD-K-	
		Tsukamurella paurometabola	296138333	--AQ-T-----HH	RDQAAA--A--GLD-K-	
		Nocardia farcinica	54022396	--AC-D-----HH	RTDP---A--AMD-K-	
		Mycobacterium tuberculosis	15610721	--DGIV--N--DQ	R--P-A--AI--LD-K-	
		Brevibacterium linens	260906432	--KSLE-----	SGSNP---L--Q--K-	
		Amycolatopsis mediterranei	300790482	--VG-P-----M-H	TAEP-S--AI--SL--K-	
		Propionibacterium acnes	282853369	--M--P-----T-D	NADPQ-----M---	
		Leifsonia xyli	50955643	-----	NA-P-S-----AM---	
		Saccharopolyspora erythraea	134097049	--A-----VN-	QE-Q-E--A--M--K-	
		Kribbella flavida	284034288	--V--S--T--V-E	HSEP-A-----MM---	
		Thermobifida fusca	72163110	--VGLP-----T-	RTEP-----	
		Thermomonospora curvata	269128661	--VGL-----T-	REEP-----L---	
		Janibacter sp.	84497239	--TGL-----N	RTAA-----L---	
Catenulispora acidiphila		256397461	--GLS-----	R-VP-----L---		
Rothia dentocariosa		296934186	--AP-L----M-V--	T-HP-A-----M---		
Micrococcus luteus		239916720	--ESLD-----G	VSEP-E-----L---		
Streptomyces coelicolor		21221780	--TGL-----T-	RAEP-----L--L---		
Kineococcus radiotolerans		152964572	--VGLP-----V-S	T-EP-A---I--L---		
Beutenbergia cavernae		229821758	--EGLP-----G	SGLVQ---A--LD---		
Bifidobacterium animalis		183601979	--E--R--A-----S	QNEP-E-----F-LD-H-		
Actinomyces sp.		269218316	--GL-----S	RREN-----MAL---		
Streptomyces scabiei	290958385	--TGLT-----T-	RDEP-----L--L---			
Arthrobacter sp.	116668768	--VGL-----V--	TKEP-S---I--L---			

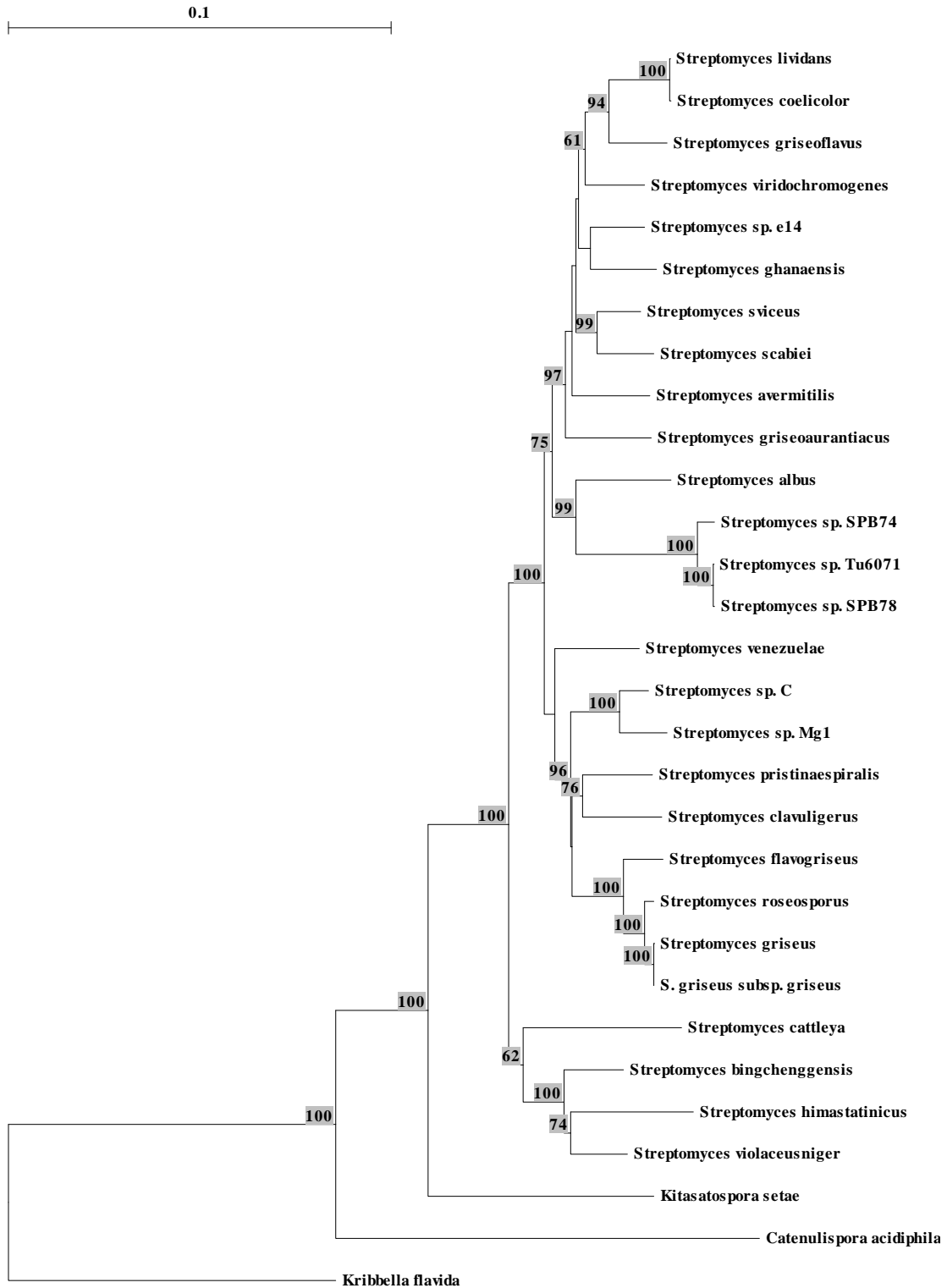
Supplemental File 22

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.

		284993113	100	130
Pseudonocardiales	Geodermatophilus obscurus	284993113	GASTWENAQPTFRTEAGT	G LALCHNGNLVN
	Amycolatopsis mediterranei	300790835	--TI-----I---AT-S	-SFA-----
	Nakamurella multipartita	258655153	-ST-----AI-DV-S	-I--G-----
	Actinosynnema mirum	256380841	-ST-----H-AT-S	A-S-G-----
	Saccharopolyspora erythraea	291008350	-GGS-----AT-S	-G-----
Glycomycetales	Saccharomonospora viridis	257057727	-GG-----V---K-D-	-G-----
	Stackebrandtia nassauensis	291303580	--P-----I-S-ST-	T I--G----I-
Micromonosporales	Micromonospora aurantiaca	270500765	-G-----I-A-T---	T I--A-----
	Salinispora tropica	145592720	-G-----I-S-AS--	T I--A-----
Corynebacteriales	Nocardia farcinica	54022546	-GV-----I---AV-S	-G-----
	Gordonia bronchialis	262200944	-ST----S-I---D---	-V-G-----
	Segniliparus rotundus	296395406	--PV-----V---VAS-G	C--G-----
	Tsukamurella paurometabola	296141250	-ST----S-I---S-N	-V-G-----
	Rhodococcus jostii	111021811	-ST-----I---A-S-	-I--G-----
	Rhodococcus opacus	226364333	-ST-----I---A-S-	-I--G-----
	Mycobacterium marinum	183984849	-DT-----V--N-A---	-V-G-----
	Mycobacterium avium	41406736	-DT-----V--N-A---	-V-G-----
	Mycobacterium tuberculosis	289749323	-DT-----V--N-A---	-V-G-----
	Corynebacterium jeikeium	68535423	-GNS-----M--MAPD-	D V--G-----D
Frankiales	Corynebacterium diphtheriae	38234487	-GVM--S-M--V-PE--	D V-----
	Frankia alni	111219629	-S-----SY--ARF-G	P I--G-----T-
	Frankia sp. CcI3	86738817	-S-----SY--ARF-G	P I--G-----T-
Other Actinobacteria	Frankia sp. EuI1c	280964547	-S-----SY--ARL-G	-V--A-----T-
	Propionibacterium acnes	50843417	---I-D-----SRQG-D	-A-----T-
	Propionibacterium freudenreich	297625693	---V-K-----KP-PS-G	---A-----T-
	Nocardioides sp. JS614	119718535	---Q-----P-AD-S	I--G-----I-
	Kribbella flavida	284028613	-S-V-A-----S-AT-S	V--A-----T-
	Streptomyces roseosporus	239942512	---V-----A-AH-S	I--G-----
	Streptomyces coelicolor	21222488	---V-----A-AQ-S	I--G-----
	Catenulispora acidiphila	256389526	-S-V-----A-AH-A	V--G-----T-
	Acidothermus cellulolyticus	117929279	---V-----A-PV-S	I--A-----T-
	Aeromicrobium marinum	293168761	---V-Q-----P-AT-S	V--A-----T-
	Thermobifida fusca	72163146	-S-V-----FSARE-G	---G-----I-
	Thermomonospora curvata	269124572	-SP-----SSSD-S	---T-----I-
	Streptosporangium roseum	271970169	-S-V-----LSS--E-G	---A-----I-
	Nocardioopsis dassonvillei	297563308	-SPV-----Y-ARE-G	---G-----I-
	Kocuria rhizophila	184200070	---H-A-----LGA-PH--	---A-----T-
	Micrococcus luteus	289704658	-VNK-A-----LGA-ADDG	T V--A-----
	Arthrobacter aurescens	119963114	---H-A-----LGA-AT--	V--A-----T-
	Renibacterium salmoninarum	163840604	-S--A-----LGA-RT--	V--A-----T-
	Rothia mucilaginosa	255326687	---H-A-----LG--PH--	-C-A-----T-
	Brevibacterium linens	260903930	-SPSF-----LGP-PF--	V--A-----T-
	Thermobispora bispora	296271294	-S-V-----LTS-AQ-G	I--A-----I-
	Xylanimonas cellulositytica	269957860	---V-----LGA-AS--	V--G-----T-
	Sanguibacter keddieii	269796679	-GV-----LGA-A---	V--G-----
	Jonesia denitrificans	256833562	-GV-----LGA-AT--	V--G-----
	Beutenbergia cavernae	229821994	---Q-----LGP-AS--	V--G-----T-
	Leifsonia xyli	50955761	---S-Q-----LGS-SG--	V--G-----T-
	Cellulomonas flavigena	296130740	-G-----LGP-A---	V--G-----T-
	Janibacter sp. HTCC2649	84497679	---I-----LGGHDG--	V--A-----I-
	Kytococcus sedentarius	256824476	---V-----LAGSDDR-	---A-----II-
	Actinomyces odontolyticus	293191580	--DV-R-----LGP-PT--	---A-----T-
Mobiluncus curtisii	304390199	-M-S-H-----LGP-AF--	--MA-----	
Brachybacterium faecium	257068085	-G-V-S-----LGPRHE--	V--A-----	
Kineococcus radiotolerans	152967984	-GGG-----LGD-AG--	V--A-----T-	
Conexibacter woesei	284042292	-S-----VY--SDRREI	--A--T-I-	
Rubrobacter xylanophilus	108803838	-SAS-----E-IGRGEVN	V-VA-----D	
Other Bacteria	Clostridium botulinum	226950309	---KSD---IVG-YKL-S	I-IA-----
	Brevibacillus brevis	226310188	-S-KI-----L-FRYAQ-S	M-VA-----
	Acidithiobacillus caldus	255022142	-G-VLR-T--I-INYRH-A	F-VG-----
	Oscillatoria sp.	300864572	-S-KVV---AVVK-RL-S	---A-----
	Arthrospira platensis	291566385	-S-RIV---AVSE-RL-P	V-VA-----
	Nostoc punctiforme	186681365	-S-RKV---AVLE-RL-S	---A-----
	Synechocystis sp.	16331334	-S-HRV---AVLP-RL-P	---A-----
Spirochaeta smaragdinae	302339611	-G-GV----LVAH-KK-A	V--A-----	

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

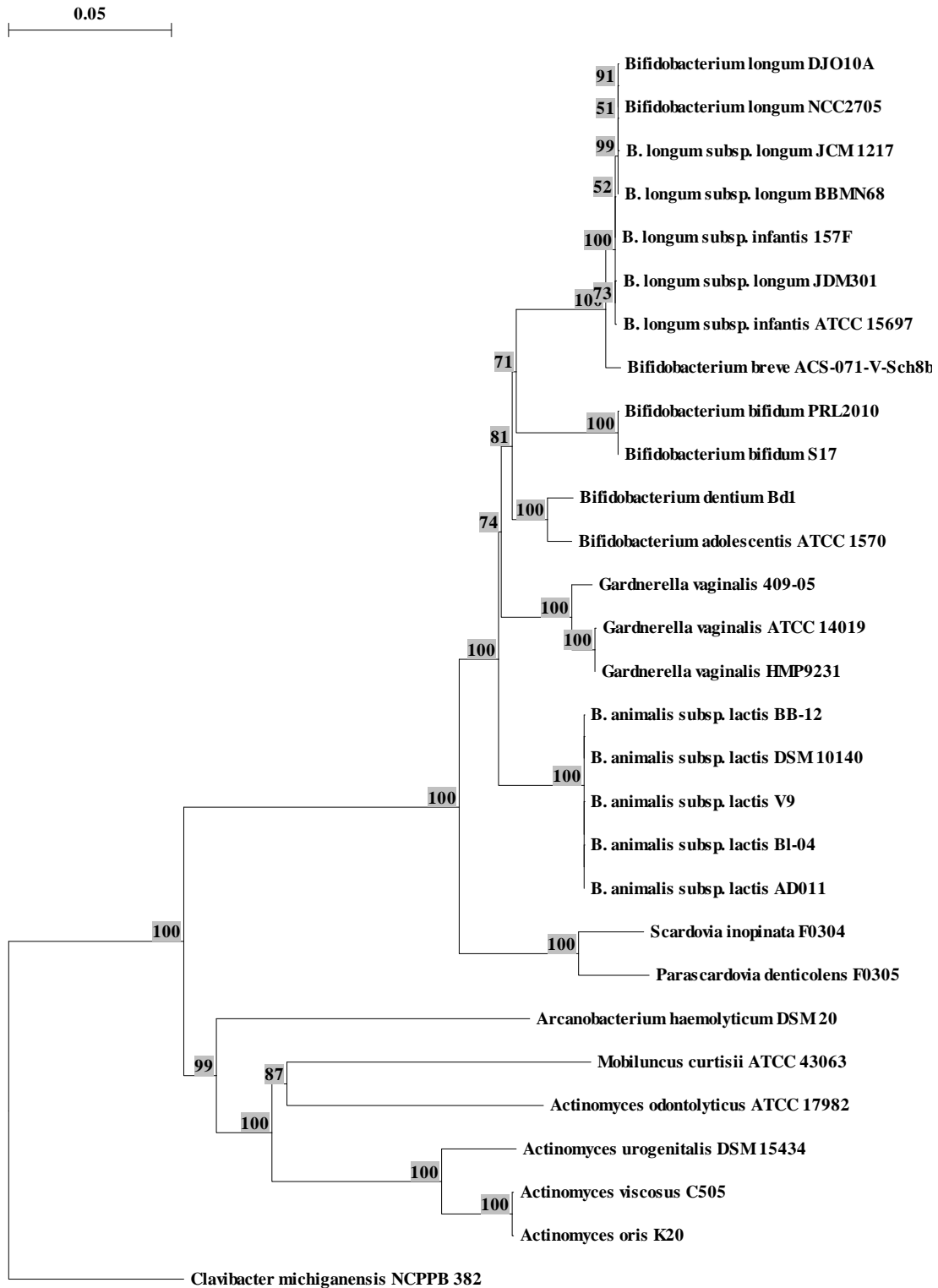
		108		155
	Streptomyces sp. ACT-1	282868020	DAVLDLEVPEDVVKRIAGRRCRN	N SAHVFLHTYNPPKAEGVCDACG
	Streptomyces griseus	182436606	-----	-----
	Streptomyces roseosporus	239943598	-----R	-----S-S---T---V--
	Streptomyces sp. ACTE	282860831	-----I---	D -S---V-----TD---E--
	Streptomyces pristinaespiralis	254397968	-----	-----V-S-A---E--
	Streptomyces avermitilis	29831490	-----I---	D -----VS-K---Q---V--
	Streptomyces flavogriseus	260453627	-----I---	D -S---V---A---T---V--
	Streptomyces griseoflavus	256814347	-----A-----I---	D -----A-S---QA---T--
	Streptomyces clavuligerus	254391372	-----I---	D -S---V-S---AQ---V--
	Streptomyces lividans	1708608	-----A-----	D -----V-T---K---V--
	Streptomyces sp. SPB78	256677124	-----I---	D -S---V---SK-Q---V--
	Streptomyces sp. SPB74	254375960	-----I---	D -S---V---SK-Q---V--
	Streptomyces coelicolor	2924292	-----A-----	E PK---V-T---K---V--
	Streptomyces viridochromogenes	256804016	-----I---	D -----V-KQ-AK---V--
	Streptomyces sviveus	254401396	-----E-----I---	D -----VS-KA--QD---V--
	Streptomyces ghanaensis	291437818	-----A-----I-K	D -S---S---AK---V--
	Streptomyces sp. e14	294630914	-----I---	D -S---V---SK-Q---V--
	Streptomyces sp. C	256769231	--I-----E-----I-K	D -----VI-AA-EV---K--
	Streptomyces sp. Mg1	254383151	-----E-----	D -S---V-AA---T---T--
	Streptomyces albus	239980035	-----I---	D -S---VE-SK-ET---V--
	Streptomyces hygrosopicus	256777195	-----T---	D ---T---VE-KK-ET---V--
	Streptomyces scabiei	290954941	----A--I--AQ--R-----L-Q	D RE-I--VE---AVT---V--
	Kitasatospora setae	311896543	-G-----L---	D G-----VV---T---E--
	Frankia sp. CcI3	86739316	-V--E-V-DD---R-LS---T---	CG-IW-IDFD--SV-----R--
	Frankia sp. EUN1f	288917251	-V--E-V-DD---R-LS---T---	CG-IW--DFD--RD-----R-S
	Frankia symbiont of Datisca	289641145	-V--E-V-DD---Q-LS---T---	CG-IW-IDFD--AV-----R-N
	Frankia sp. EAN1pec	158317770	-V--E-V-DD---R-LS---T---	CG-IW--DFD-SRV--I--R-S
	Frankia alni	111220567	-V--E-V-DD---R-LS---T--A	CG-IW--DFD--SV--L--R--
	Frankia sp. EuI1c	280963705	-V--E-V-DD---R-LS---T--S	CG-IW-VDFD--TV-D---R-N
	Acidothermus cellulolyticus	117927536	-V--E-V-DD---IR-LS---T-AD	C---W-V--D---TVD---L--
	Stackebrandtia nassauensis	291298732	-L--E-V-DD---IR-LS---T-HG	CGKIW-VEFDRTEV-----R--
	Micromonospora aurantiaca	270503772	-L--E-V-DD---IR-LS---T--G	CGKIW-VEFDS-SR-----R--
	Salinispora arenicola	159039813	-L--E-V-DD---IR-LS---T--G	CGK-W-VEFDA-SQ--R--R--
	Salinispora tropica	145596413	-L--E-V-DD---IR-LS---T--G	CGK-W-VEFDA-SQ--R--R--

Streptomycetales

Other Actinobacteria

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

		605	GDPKTP	647
<i>Actinomycetales</i>	Actinomyces coleocanis DSM 154	227494829	APYLTDRKPYPAVVDD	KRLVWIVDAYTTSNHYPYA
	Actinomyces odontolyticus ATCC	154507576	-----EQSA-----	----S-- ----V-----T-N----
	Actinomyces urogenitalis DSM 1	227498179	--W---GN-----	D--S-- --V---L-G---T-N----
	Arcanobacterium haemolyticum D	297570997	-----S-AV-----	SN-D-- SE----I-G---T-N----S
	Mobiluncus curtisii ATCC 43063	298345207	-----SRA-----	----- --V---G---S---S
	Mobiluncus mulieris 28-1	269977408	-----S-A-----	----- --V---G---DS---S
	Gordonia bronchialis DSM 43247	262203365	--W-V-S-T---MA	D GSIK---G---LDS---S
	Tsukamurella paurometabola DSM	296140726	--F-V-T-T--V-DS	KS G-I--V-G---LPK----
	Leifsonia xyli subsp. xyli str	50954636	----M-SQA----I-	G-IK-V-G---QF--S
	Clavibacter michiganensis subs	148272384	----I-KD---S---	G-V---G---DQ---S
	marine actinobacterium PHSC20C	88856795	-----SDV---I--	G-I---G---GAN---S
	Tropheryma whipplei TW08/ 27	28572789	--W-R--N---S---	G-I---G---DQF--S
	Sanguibacter keddieii DSM 1054	269796028	-----GRV-----	G-VK--I-G---E---S
	Gardnerella vaginalis ATCC 140	297585973	-----GRV-----	G-VK---G---DS---S
	Bifidobacterium dentium ATCC 2	171743241	-----GRV-----	G-VK---G---DA---S
Other <i>Actinobacteria</i>	Bifidobacterium longum DJ010A	23335262	-----GRV-----	G-VK---G---DA---S
	Xylanimonas cellulositytica DS	269955600	--W---HRV-----	G-VK---G---DAF--S
	Arthrobacter aurescens TC1	119962109	-----V-GNA-----	G-VK---G---QYF--S
	Renibacterium salmoninarum ATC	163839990	----I-GN---I--	GKVK---G---A---S
	Micrococcus luteus SK58	289705126	----V-TA---I--	G-VQ-----DQF--S
	Rothia dentocariosa ATCC 17931	296935199	---KV-S-A---IIN	N-VQ-----T-QF--
	Brevibacterium linens BL2	260906071	--F-KI-GD-----	GKIK--L-G---AKD---S
	Thermobispora bispora DSM 4383	296270911	--F---SD---I--	G-I---I-----Q---S
	Saccharomonospora viridis DSM	257054826	--W-V-SD---I-	G-IQ---G---L-N---S
	Nakamurella multipartita DSM 4	134097698	--W-K-GD-----	GKVK--I-G---LDN---S
	Actinosynnema mirum DSM 43827	258651827	---F-T-T-A-----	G-I-----AAN---
	Segniliparus rotundus DSM 4498	256375086	--W-V-SD-----	GKIT---G---LDN---
	Corynebacterium jeikeium ATCC	296394538	--W---E-A-----	G-IQ---G---LAQF--S
	Corynebacterium diphtheriae NC	260577881	--W-T-S-T--I-I-	G-VK-V-G---L-NL--S
	Mycobacterium bovis AF2122/ 97	38233339	--W-T-STT---I-	G-VK---G---LTSL---
	Mycobacterium tuberculosis H37	31794367	--W-T-SAV---I-N	-----G---LDN---S
	Rhodococcus erythropolis PR4	15610329	--W-T-SAV---I-N	-----G---LDN---S
	Rhodococcus jostii RHA1	226305718	--W-T-GAA-----	GKMQ--I-G---AQD---
	Nocardia farcinica IFM 10152	111025452	--W-T-SN-----G	G-I---G---AA---S
	Kineococcus radiotolerans SRS3	54026499	--W-T-GNA-----	E-I-----LDN---
	Aeromicrobium marinum DSM 1527	152965161	--W-M-GDA--SI--	G-V---L-G---S---S
	Nocardioides sp. JS614	293167432	--W-V-GDV-----	G-V---G---N---S
	Kribbella flavida DSM 17836	119715766	--W-V-SD-----	G-IL--I-G---TDR--LS
	Propionibacterium freudenreich	284033191	--W-P-GD-----	GQV---G---AN---S
	Streptomyces albus J1074	297626196	--W-V-QD-----	G-I---A---STAN--NS
	Streptomyces coelicolor A3(2)	239982188	--W-I-GDA-----	GKIQ-----G---S
	Catenulispora acidiphila DSM 4	32141256	--W-I-GDA-----	G-IQ-----T-G---
	Janibacter sp. HTCC2649	256397027	--W-V-GTT--VAEN	G-IL-V-G---TAA--S
	Kytococcus sedentarius DSM 205	84494587	--W-I-GNA-----	G-VK---G---TSR--NS
	Cellulomonas flavigena DSM 201	256825711	--W---GDAF-S---	G-I---G---SR---S
	Jonesia denitrificans DSM 2060	296130298	----D--GRV-----	G-VK--I-G---DQ---
	Stackebrandtia nassauensis DSM	256831965	--FI---NRV-----	GKV--L-G---NL--S
	Geodermatophilus obscurus DSM	291301839	--F-V-GD--T---	GKIK-VL-----DSF--S
	Frankia sp. EUN1f	284992601	--F-V-GD-----I-	G-VQ--L-G---GS---
	Frankia alni ACN14a	288922536	--W---GD-----	G-VT--L-G---DA---S
	Nocardioopsis dassonvillei subs	111225371	--W---GD-----N	G-IT--L-G---DG---S
	Thermobifida fusca YX	297562725	----T-SR-----	G-V-----DG---
	Thermomonospora curvata DSM 43	72160945	--F-V-G-A-----	G-I-----DR---S
	Streptosporangium roseum DSM 4	269127931	--W---GN-----N	G-I---L-G---G---
	Brachybacterium faecium DSM 48	271969538	--F-S--DN---I--	G-V---A-----A---S
	Salinispora tropica CNB-440	257069103	--F-S--SQM-----	DE--V-G---TTE----
	Micromonospora sp. ATCC 39149	145596241	--F-V-GD---I-	G-VT--I-G---AAT---
	Magnetospirillum magnetotactic	238060909	--F---GD-----G	G-VQ---G---AAT---
	Jonquetella anthropi E3_33_E1	23007059	--F---GRV-----	G-VK---G---A---S
	Thermanaerovibrio acidaminovor	260654497	--F-LY-SD---I-	G-II-----VTDR---S
Clostridium botulinum A str. A	269793292	--F-LF-SD--V-HR	G-I--MM-G-VA-SRF--S	
Pelobacter propionicus DSM 237	148378022	--F-Y-SD--MVIYN	GKIY--I-----T-R---S	
Leptospirillum ferrooxidans	118581966	--F-HF-SD--LV-DQ	K G--K--I---H-TRL--S	
Trichodesmium erythraeum IMS10	31747820	-----PD-V-L-TS	D G--L-MI-----S-V--S	
Burkholderia ubonensis Bu	113477786	--F-RF-ND--I-LI-	GKIQ--L-G---V-DR--S	
Desulfonatronospira thiodismut	167588304	--F-S--HD--V-SN	G--F--Q---RWF---	
	298528725	-F-QR--EA-SV-A-	GKVV--Q---VTSR--S	

Supplemental File 27

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

		271		300
Micrococcales	Arthrobacter aurescens	119962530	AGTVEFLVDTEGE	RAGQ HVFIEMNPRIQVE
	Arthrobacter chlorophenicus	220912315	-----	-----
	Arthrobacter sp.	116670107	-----V--	-----
	Renibacterium salmoninarum	163841257	-----V--	-----
	Micrococcus luteus	289704976	-----A--	-----
	Kocuria rhizophila	184201150	-----A--	-----
	Brevibacterium linens	260906788	-----LE-D-P	-K--
	Jonesia denitrificans	256832530	-----A--	-----
	Sanguibacter keddieii	269795135	-----V--	-----
	marine actinobacterium	88855568	-----L-A--	-K-E
	Clavibacter michiganensis	148273061	-----L-D-P	-----
	Leifsonia xyli	50955170	-----L-A--	-----
	Corynebacterium glutamicum	145294862	-----EK-N	-----
	Corynebacterium striatum	227505376	-----EK-N	-----V--
	Corynebacterium diphtheriae	38233253	-----E--N	-----
	Corynebacterium efficiens	25027265	-----EA-N	-----
	Corynebacterium accolens	227502802	-----ER-N	-----V--
	Other Actinobacteria	Mycobacterium avium	118465096	-----L-ER-N
Mycobacterium gilvum		145224777	-----L-ER-H	-----C-----
Mycobacterium smegmatis		11118649	-----L-ER-H	-----C-----
Mycobacterium tuberculosis		294993947	-----L-ER--	Y-----V--
Mycobacterium ulcerans		118617574	-----L-QS--	Y-----
Mycobacterium marinum		183981762	-----L-QS--	Y-----
Rhodococcus jostii		111023476	-----L--R-N	-----
Rhodococcus opacus		226359928	-----L--R-N	-----
Rhodococcus equi		296036945	-----L--R-S	-----
Nocardia farcinica		54026162	-----L-ER-N	-----
Gordonia bronchialis		262201758	-----L-EQ-N	-----
Segniliparus rotundus		296393717	-----EQ-R	-----
Tsakamurella paurometabola		296138666	-----PQ-R	-----
Streptomyces scabiei		290962484	-----L-RD-N	-----
Streptomyces ghanaensis		239933190	-----L-RD-N	-----
Streptomyces sviveus		297197884	-----L-RD-N	-----
Streptomyces lividans		289773786	-----RD-N	-----
Streptomyces coelicolor		21219081	-----RD-N	-----
Streptomyces lividans		256789905	-----RD-N	-----
Saccharopolyspora erythraea		134102555	-----ER-R	-----
Saccharomonospora viridis		257054978	-----RD-N	-----
Actinosynnema mirum		256379967	-----L-PR-N	Y-----
Rubrobacter xylanophilus		108804861	-----GED-S	-----
Geodermatophilus obscurus		284990282	-----L-R--R	-----
Tropheryma whipplei		28493597	-----SKMN	-Y-----
Nocardiopsis dassonvillei		297560125	-----GAD-S	-----
Nocardioides sp		119717735	-----L-AK-T	YH-----
Janibacter sp. HTCC2649		84494434	-----LGKD-R	Y-----
Frankia symbiont of Datisca		289643562	-----EDQN	-Y--I-S
Frankia sp. Eu11c		280963811	-----GAD-R	YT-M-----
Gordonibacter pamelaee		295106072	---I--L-ER-N	FY-M---T-V--
Arthrobacter aurescens		119952469	-----L-P--N	Y-----
Other Bacteria	Syntrophomonas wolfei	114566070	-----L-KQ-N	-Y-----
	Clostridium carboxidivorans	255525687	---L-----KN-N	-Y-----
	Clostridium botulinum	153941124	---L-----T-N	-Y-----
	Clostridium tetani	28204268	-----KY-N	-Y--V--V--
	Bacteroides capillosus	154496258	-----KD-N	-Y-----
	Geobacter sp. M18	255059903	-----L-E--R	YY-----
	Anaeromyxobacter dehalogenans	86159437	-----QQ-H	-Y--V-----
	Eubacterium saphenum	255994596	-----L-KH-N	-Y-----
	Carboxydotherrmus hydrogenoform	78044884	-----L-DR-N	FY----T-----
	Chlamydomonas reinhardtii	159476498	-----M--KD-K	-Y-L-V--V--
	Chloroflexus aggregans	219847233	---L--L-E--N	YY----T-----
	Cyanothece sp.	196256021	-----KH-N	FY-M---T-----
	Anaerofustis stercorihominis	169335857	-----ENEK	-Y-M-V-----
	Blastopirellula marina	87310369	-----KKS-N	-YF--V-----

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.

		176	177	178	179	218
		VFDLGGGTFDVSLLLEIGK	DDGGF	STIQVQATSGDNHLGGDDWD		
<i>Bifidobacteriales</i>	<i>Bifidobacterium animalis</i>	183602852	-----	-E---	-----	-N-----
	<i>Bifidobacterium ruminantium</i>	254871275	-----	-----	-----	-N-----
	<i>Bifidobacterium longum</i>	213691062	-----	-----	-----	-N-----
	<i>Gardnerella vaginalis</i>	283783717	-----	-----	-----	-N-----
	<i>Scardovia inopinata</i>	294791133	-----	-E---	-----	-N-----
<i>Actinomycetales</i>	<i>Parascardovia denticolens</i>	294787136	-----	-----	-----	-N-----
	<i>Actinomyces viscosus</i>	326773017	-----	V--	-----	-R-N--R-----
	<i>Actinomyces odontolyticus</i>	154507951	-----	V--	-----	-R-----
	<i>Mobiluncus mulieris</i>	227876398	-----	V--	-----	-R-N--K-----
	<i>Mobiluncus curtisii</i>	298345670	-----	V--	-----	-R-N--K-----
<i>Micrococcales</i> I, II and III	<i>Arcanobacterium haemolyticum</i>	297570776	-----	V--	-E---	-R--K-----
	<i>Arthrobacter aurescens</i>	119962235	-----	V--	-E-N-	-R-A--R-----
	<i>Arthrobacter arilaitensis</i>	308178314	-----	V--	-----	-RS-A--R-----
	<i>Rothia dentocariosa</i>	300743745	-----	V--	-E---	-R--R-----
	<i>Micrococcus luteus</i>	289707065	-----	VA-	-E-D-	A--R-A--R-----
	<i>Rothia mucilaginosa</i>	283458697	-----	V--	-----	-R--R-----
	<i>Brevibacterium linens</i>	260904856	-----	V--	-E-D-	-R--R-----
	<i>Cellulomonas fimi</i>	332668977	-----	V--	-----	-E-R--R-----
	<i>Kocuria rhizophila</i>	184200044	-----	V--	-E-D-	-R-A--R-----
	<i>Sanguibacter keddieii</i>	269793786	-----	V--	-E-D-	-R--R-----
	<i>Jonesia denitrificans</i>	256831594	-----	V--	-E-D-	-R--R-----
	<i>Beutenbergia cavernae</i>	229822187	-----	V--	-E-D-	-R-N--R-----
	<i>Leifsonia xyli</i>	50955833	-----	V--	-D-	-R-A--R-----
	<i>Tropheryma whipplei</i>	28493717	-----	V--	-D-	-R--R-----
	Other Actino with insert	<i>Brachybacterium faecium</i>	257068394	-----	VS-	-----
<i>Clavibacter michiganensis</i>		170783106	-----	V--	-D-	-RS-A--R-----
<i>Kineococcus radiotolerans</i>		152968176	-----	V--	-E---	-R--R-----
<i>Catenulispora acidiphila</i>		256389313	-----	V--	-----	-E-K-N-----
<i>Streptosporangium roseum</i>		271962100	-----	DV-Q	E--H	GFVE-K-----
<i>Thermobispora bispora</i>		296268240	-----	DV-Q	E--H	GFVE-K-----
<i>Amycolatopsis mediterranei</i>		300790966	-----	E	-----	GVVE-R-----
<i>Saccharomonospora viridis</i>		257057797	-----	E	-----	GVVE-R-----
<i>Nakamurella multipartita</i>		258655317	-----	D	-----	GVVE-K-A--K-----
<i>Corynebacterium diphtheriae</i>		38234669	-----	D	-----	GVVE-R-A--E-----
Other <i>Actinobacteria</i>	<i>Geodermatophilus obscurus</i>	284993160	-----	E	-----	GV-E-K-A-----
	<i>Nocardia farcinica</i>	54027383	-----	E	-----	GVVE-R-----
	<i>Mycobacterium tuberculosis</i>	215425567	-----	E	-----	GVVE-C-----
	<i>Rhodococcus opacus</i>	226359569	-----	E	-----	GVVE-R-----
	<i>Gordonia bronchialis</i>	262204101	-----	D	-----	GVVE-R-----
	<i>Segniliparus rotundus</i>	296393848	-----	D	-----	GVVE-R-----N-----
	<i>Tsukamurella paurometabola</i>	296138223	-----	D	-----	GVVE-R-----N-----
	<i>Dietzia cinnamea</i>	319949484	-----	D-D	-----	GVVE-R-----
	<i>Actinomadura rubrobrunea</i>	19387035	-----	V-D	-----	GVVE-K-----
	<i>Thermomonospora curvata</i>	269124511	-----	V-D	-----	GVVE-K-----
	<i>Actinosynnema mirum</i>	256380930	-----	L-D	-----	GVVE-R-----
	<i>Aeromicrobium marinum</i>	311744450	-----	D	-----	GV-E-K-----
	<i>Pseudonocardia dioxanivorans</i>	331700257	-----	E	-----	GVVE-K-N-----
	<i>Thermomonospora curvata</i>	269124511	-----	V-D	-----	GVVE-K-----
	<i>Spirillospora albida</i>	19387029	-----	V-D	-----	GVVE-K-----
	<i>Kitasatospora setae</i>	311897203	-----	D	-----	GVVE-K-N-----
	<i>Thermobifida fusca</i>	72160600	-----	DV-D	-----	GVVE-K-H-----
	<i>Frankia alni ACN14a</i>	111225972	-----	D	-----	GVVE-KS---T-----
	<i>Nocardioides</i> sp. JS614	119718582	-----	E	-----	GVVE-K-----
	<i>Propionibacterium acnes</i>	314924503	-----	D-SD	-----	GVFE-K-N-----
	<i>Catenulispora acidiphila</i>	256395556	-----	I-DV-D	-----	GVVE-RS-A--T-----F-
	<i>Stackebrandtia nassauensis</i>	291297898	-----	AD	-----	GV---K-N-----
	<i>Acidothermus cellulolyticus</i>	117929322	-----	E	-----	GIVE-K-----T-----
	<i>Nocardioidaceae bacterium Broa</i>	326331832	-----	E	-----	GVVE-K-----
	<i>Streptomyces avermitilis</i>	29831027	-----	D	-----	GVVE-K-N-----
<i>Rubrobacter xylanophilus</i>	108803627	-----	I-L-D	-----	GVFE-K---N-----F-	
Other Bacteria	<i>Mycoplasma pneumoniae</i>	13508173	-Y-----	D-AE	-----	G-FE-L-A--R-----
	<i>Synergistetes bacterium</i>	295111025	-----	I-DV-E	-----	GVFE-L-A--R-----
	<i>Dictyoglomus thermophilum</i>	206901892	-----	I---E	-----	GVFE-I--A-N-R-----F-
	<i>Staphylococcus aureus</i>	253732233	-----	I-L-D	-----	GVFE-LS-A--K-----F-
	<i>Bacillus subtilis</i>	296333295	-Y-----	I-L-D	-----	GVFE-RS-A--R-----F-
	<i>Verrucomicrobium spinosum</i>	171911502	-Y-----	I-V---D	-----	GVFE-L-D--T-----
<i>Synechococcus elongatus</i>	56751645	-----	V-V-D	-----	GVFE-L---T-----F-	
<i>Veillonella caviae</i>	38326720	-----	I-L-D	-----	GVFE-L---N-----F-	

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.

		485	522
<i>Propionibacteriales</i>	<i>Propionibacterium acnes</i>	282854317	LAI SMLPFAWNVWIT RHPDNIEVADPWGWRGLEWATS
	<i>Kribbella flavida</i>	284030714	-GM-----FY--YKS -KA-LVG-D-----S-----
	<i>Nocardioides</i> sp. JS614	119714168	--A-T---LL--YKS AGS-PVG-D-----S-----
	<i>Aeromicrobium marinum</i>	311742842	-GA-T---FL-L-K- -SAKPVGLD-----S-----
	<i>Acidothermus cellulolyticus</i>	117927698	--L--I--GI---R- W -YA-HVG-D-----AS-----
	<i>Brevibacterium linens</i>	260906804	-GL--V--F-----S -NA-KVT-D----Y-GS-----
	<i>Streptomyces ambofaciens</i>	126348132	-GL-T---LY---K- A HLGKK---D---Y--S-----
	<i>Kytococcus sedentarius</i>	256825444	-GA-----F---Y-- H KKG EYVNTD-----SS-----
	<i>Xylanimonas cellulositytica</i>	269956803	-GA-----F---YT- W -RA-LVT-D----Y--S-----
	<i>Jonesia denitrificans</i>	256832678	--A-T---L---YV- A KNA-RVT-D----F-AS-----
	<i>Rothia dentocariosa</i>	300741502	--V--I-WF---Y-- A -YA-KV--D----F-GS-----
	<i>Saccharopolyspora erythraea</i>	134097777	-GA-----L---FKS Y -YGEIVT-D----Y-NS-----
	<i>Thermobispora bispora</i>	296269270	-G--T---MY-M-R- W -KA-KVT-N----Y-CS-----
	<i>Corynebacterium genitalium</i>	300780571	-G-AF---I---YKS W -YGEVVT-D----Y-NS-----
	Other <i>Actinobacteria</i>	<i>Mycobacterium smegmatis</i>	118472583
<i>Rhodococcus equi</i>		296035385	-GA-T---L---FKS F -YGTVVT-D----Y-NS-----T
<i>Tsakamurella paurometabola</i>		296140636	-GA-T---V---FKS Y -YGEVVT-D----F-NS-----
<i>Thermobifida fusca</i>		72161285	-G--T-I-F---V- A KKA-KVT-D----Y-CS-----
<i>Leifsonia xyli</i> subsp. <i>xyli</i>		50955129	--V--I--LV--YL- A --A-KVT-N----YS-S-----
<i>Catenulispora acidiphila</i>		256390866	-GA-T---LY--Y-- A KRGKKVT-D----F--S-----
<i>Kineococcus radiotolerans</i>		152967202	-GL-F---L---YR- A -FGERVT-D----Y-AS-----
<i>Frankia alni</i> ACN14a		111225422	--M-T--LIY-L-HS Y --GALAT-D----Y-NS-----
<i>Janibacter</i> sp. HTCC2649		84496140	-GA-----FY---R- W -YA-LV-TD----Y-AS-----
<i>Streptosporangium roseum</i>		271967621	-GV-T---LY---M- S -RA-KVTRD----F-NS-----
<i>Clavibacter michiganensis</i>		170781786	--L--I--FL-----A -TA-RVT-N----Y--S-----
<i>Amycolatopsis mediterranei</i>		300789548	-GA-T---I---FKS Y -YGEIVT-D----Y-NS-----
<i>Streptosporangium roseum</i>		271964184	-GA-T---LY---K- H KTA-KVT-D----F-NS-----
<i>Renibacterium salmoninarum</i>		163840328	-GA-LI--F---Y-- W -SKKRV--D----F-AS-----
<i>Actinosynnema mirum</i>		256377793	-GA-T---L---FKS Y -YGERVT-D----Y-NS-----

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 *Kineosporiales* species. The homologs for this protein were not detected in *Actinomycetales*.

		111	151
Bifidobacteriales Micrococcales Kineosporiales	Arthrobacter chlorophenicus	220914657	DYWFTVSGHRVHKTVHHYLLRATGGEL TIE NNDPQEAVDVA
	Arthrobacter aurescens	119962886	-----F-----
	Rothia mucilaginosa	255326494	----SN-T-I-----FA----- S-D ----H-I---
	Brevibacterium linens	260904996	-----T-I----L----F----RS-T- -VD ----I-A-
	Rothia dentocariosa	296934406	-----S--Q-I-----FC----R- ---H--H-----
	Kocuria rhizophila	184201990	-----F--E--H- -T- ----H-----
	Renibacterium salmoninarum	163842264	-----T-----F-----
	Micrococcus luteus	239918794	-----AR-----F----- -T- ----H-----
	Mobiluncus mulieris	227876519	---SGPDR---V---F-CETI---I -V- ----C--TAAG
	Actinomyces coleocanis	227494172	---SSLDR---V-F---EYVS--I -V- ----H--E-A-
	Kineococcus radiotolerans	152968434	---SAD-R-I--V--LF--K---T- -V- ----A--I--E
	Beutenbergia cavernae	229822682	---AGHDR---V-----E----- -T- ----H--E---
	Sanguibacter keddieii	269797052	-----GDDR---V---F--E-V-Y- -V- G--S--E---
	Xylanimonas cellulositytica	269958131	---SGETR---V-----SM---R- -V- D--G--E-AF
	Cellulomonas flavigena	296131530	---SGDE---V---F--G-LH--- -V- ----G--E--E
	Jonesia denitrificans	256883751	-----GDDR---V-----D-V-F- -V- G--S--E--E
	Kytococcus sedentarius	256826445	---STPER-I--R-----E-V--R- -V- E--H---Q
	Janibacter sp. HCC2649	84497208	---A-GDR-I--Y-----E-V-V- --- --H--I---
	Gardnerella vaginalis	297586730	---GTSQ---L--FA-KYVS--- SVL G--H--E-AI
	Bifidobacterium breve	291455679	---GTTQ---L---A--Q----- -V- G--H--E-AI
	Bifidobacterium bifidum	224283940	---GTSQ---L--FA--QI--D- -V- G--H--E-AI
	Bifidobacterium dentium	171741746	---GTSQ---L--FA--QI--D- -V- G--H--E-AI
	Bifidobacterium longum	227547076	---GTTQ---L--FA-KQ----- -V- G--H--E-AI
	Scardovia inopinata	294790212	---GTTQ---L---V-QKI--H- -V- G--H--E-AI
	Parascardovia denticolens	294786231	---GTTQ---L---V-Q--H- -V- G--H--E-AI
	Arcanobacterium haemolyticum	297572330	---SGHDR---V---F--E-LH-T- -V- ----H--EM-E
	Corynebacterium glutamicum	145297079	---VSE-K-I-----H---YVD-D- --E-P-VTE--
	Corynebacterium efficiens	25029485	---VSE-R-I-----H---YVD-D- --E-P-VTE--
	Mycobacterium leprae	15828458	---VTDDC-----M-FS--- S-D-L-VTE--
	Mycobacterium tuberculosis	215448255	---VTD-R-----M-FL--- S-E-L-VAE--
	Rhodococcus jostii	111020631	---VTE-R-----CL--- S-E-V-VTE--
	Rhodococcus erythropolis	229491182	---VTE-R-----FL--- S-E-I-VTE--
	Rhodococcus opacus	226362874	---VTE-R-----CL--- S-E-V-VTE--
	Gordonia bronchialis	262204639	---VSE-R-I-----SV--- S-A-Y-VSE--
	Nocardia farcinica	54027631	---VTE-R-----F--SV--- S-A-V-VTS--
Tsukamurella paurometabola	296141888	---A-E-R-----F--SWES-D- SAE-Y-VSE--	
Segniliparus rotundus	296392448	---LHGQ-K-L-----M-FQS--- S-E-V-VTE--	
Actinosynnema mirum	256381056	---AAEDR-----F--E-L--D- S-E-V-VTE--	
Amycolatopsis mediterranei	300791146	---VAEKR-I-----F--E-L--- S-E-V-VTE--	
Saccharopolyspora erythraea	134103798	---VAGNR-----F--E-VR--- S-E-V-VTE--	
Saccharomonospora viridis	257057897	---VAERR-----F--E-D--- S-E-V-VTE--	
Geodermatophilus obscurus	284993423	-F--VAD-R-----F---V--A- S-A-I-VTE--	
Nakamurella multipartita	258655494	--T-VAQ-R-I--R-----ME-V--- S-A-I-VTE--	
Frankia sp. EAN1pec	158319039	-F--VAGEA-----F--LR-S-D- S-D-V-VAE--	
Frankia alni	111226196	-F--VAG-T-I-----F--LRSV-S- S-A-I-VDA--	
Frankia sp. EUN1f	288921202	-F--VAGEA-----F--VR--A- S-D-I-VEE--	
Streptomyces sp. AA4	256673333	---VAQQR-----F--E----- S-E-V-VTE--	
Conexibacter woesei	284047245	-FYQRD-Q-IF-M-RFF-F-YRA-A- E-H-D-VEEAR	
Sulfurihydrogenibium azorense	225847905	E--YRSGVDTI--F-YY--MKYV---- -PQKE-IESAE	
Persephonella marina	225850608	---YSMGLT-I--F-YY--M-YA--DI VPQKE-IDEAK	
Roseiflexus castenholzii	156743636	E--RAGPT-I--Y-DLF-I-Y--S- MPQTA-VD--R	
Chloroflexus aurantiacus	163848854	E--RSGHG---Y-DLF--YER--V RPQIG-VD-AR	
Thermodesulfovibrio yellowston	206891057	---IME-E-YF---KYF-AEY---QV -PDW -VSSAQ	

Other
Actinobacteria

Other
Bacteria

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*.

		36	87
<i>Kineosproiineales</i>	Janibacter sp. HTCC2649	84496140	GGVMALIIRAELFEPGIQLV D NPDQFNQLFTMHGTIMLLLF
	Kineococcus radiotolerans	152967202	-----L-----M-VF G TKNEY--A-----V----
	Kytococcus sedentarius	256825444	--L--L-----D--M-VL Q SK-----A-----V----
	Brachybacterium faecium	257068579	--LL--G--T--W--L-V V SK--Y-----M----
<i>Micrococcales</i> II	Jonesia denitrificans	256832678	--L--V-----I- Q SNE-Y-----
	Beutenbergia cavernae	229820378	--L--L-----A--M-V Q TGE----F-----
	Sanguibacter keddieii	269794878	--L--V-----I- Q SNE-Y-----
	Xylanimonas cellulositytica	269956803	--L--V-----I- Q SKE-Y-----
	Cellulomonas flavigena	296129925	--IL--L-----Q--MD-F Q SKE-Y--A-----
<i>Micrococcales</i> I	Rothia dentocariosa	296935578	--T--LM-I-----M-IL Q TKE-----L--M-
	Micrococcus luteus	239917722	-----M-----M-IL Q TKE-Y-----L--M-
	Brevibacterium linens	260906804	-----L--D--M-II E TKE-Y-----L--M-
	Kocuria rhizophila	184200931	-----L-----Q--M-IL E TKE-Y-----M-
	Renibacterium salmoninarum	163840328	-----L-----M-IL Q TKE-Y-----M-
	Arthrobacter aurescens	119964055	-----L-----M-IL Q TKE-Y--M-----V--M-
<i>Micrococcales</i> III	Leifsonia xyli	50955129	-----Q----L-V- A TKE-Y-----M-
	Clavibacter michiganensis	148273022	-----V--Q----LHV- E TKE-Y-----M-
	marine actinobacterium	88856530	-----V--Q----LSI- A TGE-Y-----M-
	Tropheryma whipplei	28493211	A----L--TQ----MH-I Q TRE-Y-----V--F--
<i>Other Actinobacteria</i>	Streptomyces coelicolor	21220632	-----FM---AR--L-IM SNE---A-----M-
	Catenulispora acidiphila	256390866	--LL--LM---AR--L-FL TTE---A-----M-
	Stackebrandtia nassauensis	291299036	--I---LM---AR--M-FL SV-----M-
	Salinispora tropica	145593646	--L---M---AR--L-FL S-E-----F-
	Micromonospora aurantiaca	270498091	--L---M---AR--L-FL S-E-Y-----M-
	Frankia alni	111224506	A-IL-VMM---AR--L-YF SNE-Y-----L--M-
	Acidothermus cellulolyticus	117927698	A-I---LM---AR--L-FL T-EEY-----L--F--
	Geodermatophilus obscurus	284989902	--L---LM-G--AR--L-FL S-E-Y--V----V--F-
	Nocardia farcinica	54026265	--L---LM-G--AR--L-FL S-E-----FY
	Nakamurella multipartita	258653794	--LL--LM---AR--L-FL A-E-Y-----FY
	Tsukamurella paurometabola	296140636	--L---LM---A--L-FL STE-----V---Y
	Segniliparus rotundus	296393777	--L---LM---AQ--L-FL S-E-----A---Y
	Gordonia bronchialis	262203170	--L---LM-G--AH--M-FL STE-----V---MY
	Rhodococcus opacus	226365908	--L---MM---AV--M-FL SNE-----Y
	Mycobacterium avium	41409189	--L---LM-T--AA--L-FL SNE-----Y
	Corynebacterium efficiens	25028974	--L---L-----T--L-FL SNE-----V---Y
	Corynebacterium glutamicum	23308947	--L---L-----T--L-FL SNE-----V---Y
	Streptosporangium roseum	271964184	-----M---AQ--L-IT SNE-----V--M-
	Thermomonospora curvata	269126195	--LL--VV---A--L-FI SNE-Y--A----LV----
	Thermobispora bispora	296269270	--L---VM-L--AR--L-F- SNE-----
	Thermobifida fusca	72161285	A-----L---MY--M--M -NETY-----F--
	Nocardioopsis dassonvillei	297560890	--IL-VL-----F--M-IM SNEEY-----
	Amycolatopsis mediterranei	300784963	--LT--LM-G--AR--L-FL S-E-Y-----Y
	Actinosynnema mirum	256375132	--A--ML--T--AR--M-FL SNE-Y-----V---Y
	Saccharopolyspora erythraea	291005587	--L--ML--G--AA--M-FL SQE-Y-----Y
	Saccharomonospora viridis	257054422	--L---L-----GQ--HFL SAE-Y--M--L--L---Y
	Gordonia westfalica	40445323	--L---LM-G--TT--M-FL STE-Y-----MY
	Kribbella flavida	284030714	-----L-----AK--L-I- -EEVY-----
	Propionibacterium acnes	282854317	--IL--L-----AY--M-FM HQET--F-----M-
	Aeromicrobium marinum	293166465	--L---V-----AQ--TTF- DDELY-----
	Nocardioides sp. JS614	119717364	--L--ML--S--AY--M-V- -DEVY-----