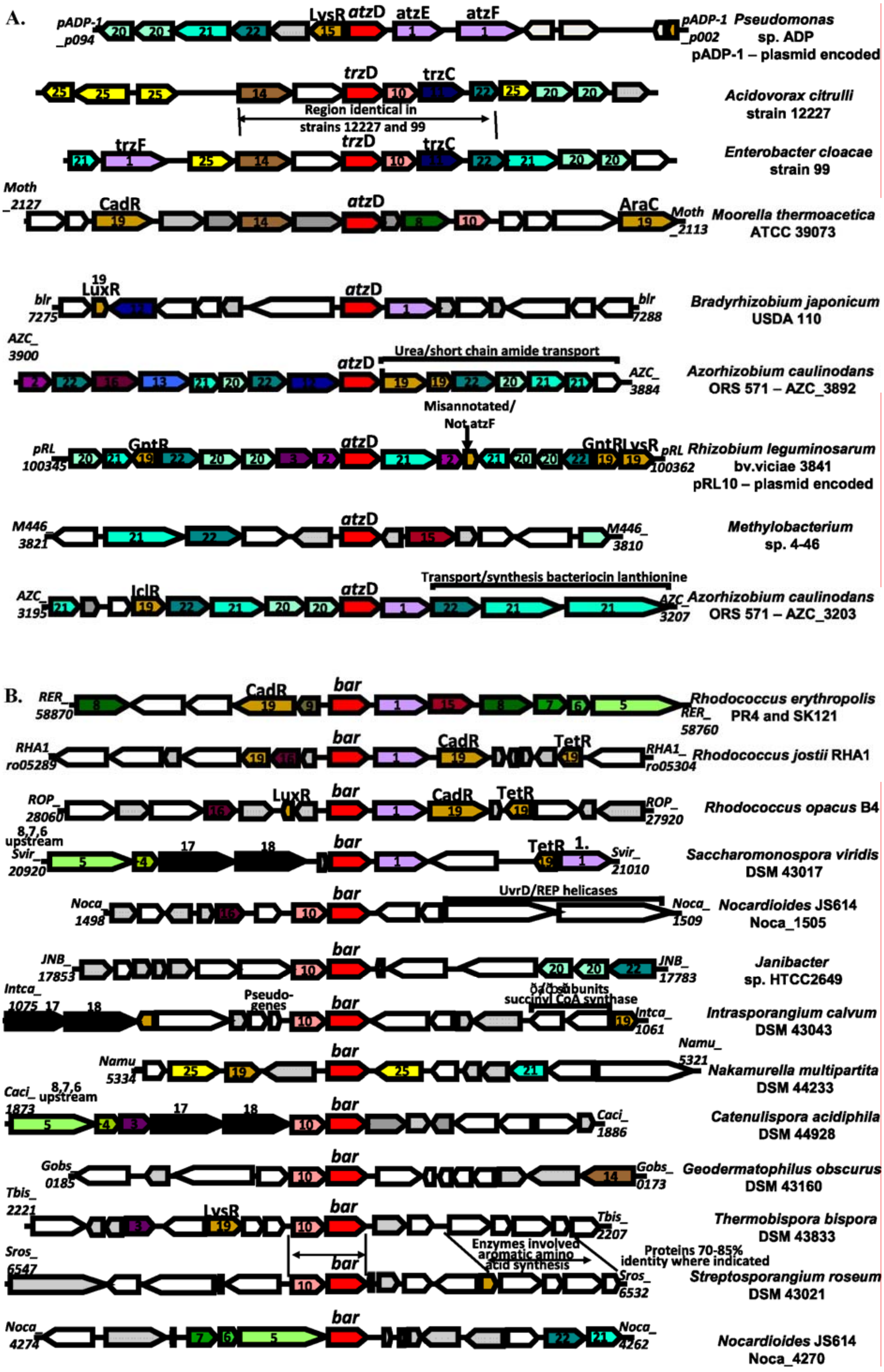
Fig. S1

Gene context diagram for representative organisms in (A) AtzD/TrzD clustered homologs and (B) barbiturase cluster. The numbering represents proteins belonging to these various groups: 1. COG0154, Asp-tRNA Asn/Glu-tRNA Gln amidotransferase belongning to the amidase signature superfamily; 2. COG1335, amidase related to nicotinamidase and isochorismatase that belong to the cysteine hydrolase superfamily; 3. COG0388, Predicted amidohydrolase that belong to the nitrilase superfamily like nitrilase and cyanide hydratase; 4-7. COG3427, COG1529, COG2080, and COG1319, uracil/thiaime dehydrogenase – precursor to barbiturase; 8. COG2233, Xanthine/uracil permeases; 9. COG0035, Uracil phosphoribosyltransferase; 10. COG0549, Carbamate kinase which belongs to the amino acid kinase superfamily, including uridylate, carbamate, asparto-, and acetylglutamate kinases; 11. trzC, ammelide amidohydrolase. 12. OG0402, Cytosine deaminase and related metal-dependent hydrolases; 13. COG0044, Dihydroorotase and related cyclic amidohydrolases, likely L-Hydantoinases or Allantoinase; 14. FdrA like protein; 15. COG0161, Adenosylmethionine-8-amino-7-oxononanoate aminotransferase that belongs to acetyl ornithine aminotransferase family of the pyridoxal phosphate (PLP)-dependent aspartate aminotransferase superfamily, incluiding ornithine aminotransferase, acetylornithine aminotransferase, alanine-glyoxylate aminotransferase, dialkylglycine decarboxylase, 4-aminobutyrate aminotransferase, beta-alanine-pyruvate aminotransferase, adenosylmethionine-8-amino-7-oxononanoate aminotransferase, and glutamate-1-semialdehyde 2,1-aminomutase; 16. COG1028, Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases); 17. COG0145, N-methylhydantoinase A/acetone carboxylase, beta subunit; 18. COG0146, N-methylhydantoinase B/acetone carboxylase, alpha subunit; 19. regulator; 20. ABC transporter permease – representing various COGs; 21. ABC trasporter ATP binding domain – representing various COGs; 22. ABC transporter periplasmic component; 23. hypothetical; 24. DUF – domain of unknown function; and 25. transposase.


A.
pADP-1_p094
20
20
21
22
LysR
15
atzD
atzE
1
atzF
1
pADP-1_p002
Pseudomonas sp. ADP
pADP-1 – plasmid encoded
25
25
25
14
trzD
10
trzC
11
22
25
20
20
Region identical in strains 12227 and 99
Acidovorax citrulli strain 12227
21
trzF
1
25
14
trzD
10
trzC
11
22
21
20
20
Enterobacter cloacae strain 99
Moth_2127
CadR
19
14
atzD
8
10
AraC
19
Moth_2113
Moorella thermoacetica ATCC 39073
blr 7275
19
LuxR
11
atzD
1
blr 7288
Bradyrhizobium japonicum USDA 110
AZC_3900
2
22
16
13
21
20
22
12
atzD
Urea/short chain amide transport
19
19
22
20
21
21
AZC_3884
Azorhizobium caulinodans ORS 571 – AZC_3892
Misannotated/ Not atzF
pRL 100345
20
21
GntR
19
22
20
20
3
2
atzD
21
2
21
20
20
22
GntR
LysR
19
19
pRL 100362
Rhizobium leguminosarum bv.viciae 3841
pRL10 – plasmid encoded
M446_3821
21
22
atzD
15
M446_3810
Methylobacterium sp. 4-46
AZC_3195
21
IclR
19
22
21
20
20
atzD
1
Transport/synthesis bacteriocin lanthionine
22
21
21
AZC_3207
Azorhizobium caulinodans ORS 571 – AZC_3203
B.
RER_58870
8
CadR
19
9
bar
1
15
8
7
6
5
RER_58760
Rhodococcus erythropolis PR4 and SK121
RHA1_ro05289
19
16
bar
1
CadR
19
TetR
19
RHA1_ro05304
Rhodococcus jostii RHA1
ROP_28060
16
LuxR
bar
1
CadR
19
TetR
19
ROP_27920
Rhodococcus opacus B4
8,7,6 upstream
Svir_20920
5
4
17
18
bar
1
TetR
19
1.
1
Svir_21010
Saccharomonospora viridis DSM 43017
Noca_1498
16
10
bar
UvrD/REP helicases
Noca_1509
Nocardioides JS614 Noca_1505
JNB_17853
10
bar
20
20
22
JNB_17783
Janibacter sp. HTCC2649
Intca_1075
17
18
Pseudo-genes
10
bar
α and β subunits succinyl CoA synthase
19
Intca_1061
Intrasporangium calvum DSM 43043
Namu_5334
25
19
bar
25
21
Namu_5321
Nakamurella multipartita DSM 44233
Caci_1873
8,7,6 upstream
5
4
3
17
18
10
bar
Caci_1886
Catenulispora acidiphila DSM 44928
Gobs_0185
10
bar
14
Gobs_0173
Geodermatophilus obscurus DSM 43160
Tbis_2221
3
LysR
19
10
bar
Tbis_2207
Thermobispora bispora DSM 43833
Sros_6547
bar
Enzymes involved aromatic amino acid synthesis
Proteins 70-85% identity where indicated
10
Sros_6532
Streptosporangium roseum DSM 43021
Noca_4274
bar
7
6
5
22
21
Noca_4262
Nocardioides JS614 Noca_4270

Gene context of AtzD/barbiturase family members. Despite the diverse origins of the source organisms and the highly divergent protein sequences, the characterized proteins throughout Clade 1 had only cyanuric acid hydrolase activity, while the only functional member of Clade 2 had barbiturase activity. To determine if additional functional information could be identified by genomic context, the flanking regions of the genes were compared (Fig. S2).

Gene context – Clade 1. Group 1 proteins were comprised of proteins found in the phylum Firmicutes. *Moorella thermoacetica* ATCC 39073 was shown previously to exhibit cyanuric acid hydrolase activity (1). Although the *Paenibacillus* and *Bacillus* proteins are also members of the Firmicutes, the tree location of these proteins was method-specific, clustering in either Group 1 or in an adjacent cluster. For this reason they were not included specifically in Group 1. The Firmicutes are diverse in sequence and organismal origins. *Clostridium asparagiforme* DSM 15981 was isolated from human feces, *Paenibacillus* sp. JDR-2 was isolated from sweet gum stem wood buried in surface soil and was shown to degrade hemicelullose, and *Bacillus cellulosilyticus* DSM 2522 was isolated from soil and is known to degrade cellulose. The AtzD homologs in these organisms have only 65, 58, and 41% sequence identity to the *Moorella* cyanuric acid hydrolase, respectively. *Clostridium* and *Bacillus* strains share a common metal-dependent amidohydrolase gene similar to cytosine deaminase (COG0402) located downstream of the *atz*D homolog genes. Members of this amidohydrolase superfamily are responsible for the atrazine degradation pathway upstream of cyanuric acid. It has yet to be determined whether the amidohydrolase protein encoded adjacent to the *atz*D gene in these organisms is producing cyanuric acid.

Gene context information is not available for all organisms in Group 2 since individual genes were sequenced, and the genome sequence is unknown. An operon was found to be present in *P*. ADP, encoding *atz*D, *atz*E, and *atz*F (2). The AtzE and AtzF proteins are members of the amidase signature superfamily and catalyze consecutive reactions, resulting in complete mineralization of cyanuric acid to ammonia and carbon dioxide (Fig. 1). From the sequence information available, an identical transposon associated region, containing *trz*D and its precursor gene *trz*C, was located in both *Acidovorax citrulli* strain 12227 and *Enterobacter cloacae* strain 99 (Fig. S2A). These two genes encoded enzymes that convert ammelide to cyanuric acid and catalyze the subsequent ring-opening reaction. TrzF, a homologous protein to AtzE and AtzF, located upstream of the TrzD protein in *Enterobacter cloacae* strain 99, was shown to catalyze the allophonate hydrolase reaction in that strain (24). A similar homolog is adjacent to the *atz*D gene in *Bradyrhizobium japonicum* USDA 110. This enzyme shares 27% identity to AtzE and 25% identity to AtzF from *P.* ADP, and could possibly be involved in deamination of the products generated by the cyanuric acid hydrolase.

Group 2 contains the only protein found in a eukaryote, the common photosynthetic marine alga *Micromonas* sp. RCC299. Along with the USDA 110 protein, this is the only other Group 2 protein from an organism not isolated for its *s*-triazine degradation ability. The *atz*D homolog gene is located on chromosome 7 and is surrounded by mainly predicted proteins that have no close function to proteins of known function. Unlike the USDA100 protein, there is no evidence of potential *s*-triazine degradation pathway genes located proximal to the *atzD* homolog. This protein has 45, 41, 42, and 32 % identity to AtzD (*Pseudomonas sp*. ADP), TrzD (*Acidovorax citrulli* strain 12227), the *Moorella* cyanuric acid hydrolase, and barbiturase (*Rhodococcus erythropolis* JCM 3132), respectively. Because of its sequence divergence and unique origin, this protein is a good candidate for future functional studies.

Group 3 enzymes were found solely in α-proteobacteria (Fig. 2). The proteins had 37-55% identity to AtzD, TrzD, the *Moorella* cyanuric acid hydrolase, or barbiturase (Table 1). No functionality was determined among this group prior to this study. It should be noted that although all of these proteins originate in α-proteobacteria, they are not found in a large range of bacteria and are not present in all organisms in that genera. For instance, *Methylobacterium radiotolerans* JCM 2831 and sp. 4-46 contain AtzD homologs, while the genomes for *Methylobacterium extorquens* AM1 and *M. nodulans* ORS 2060 do not. All of the Group 3 organisms, with the exception of *Oceanicola granulosus* HTCC2516, are involved in nitrogen fixation or are plant pathogens. The relevance of this observation has yet to be determined. The two *Rhizobium* enzymes are encoded on large plasmids and have similar gene context, which include nitrilase and cysteine hydrolase superfamily members directly upstream of *atz*D and a sugar ABC transporter and cysteine hydrolase superfamily member directly downstream. Likewise, the two *Bradyrhizobium* strains in this group have similar gene context, including a downstream TonB-dependent receptor protein and putative hydroxylase. Gene context for all other organisms in this group are highly divergent (Fig. 4A). *Oceanicola granulosus* HTCC2516 and *Agrobacterium vitis* S4 have one copy of the *atz*E/F protein pair in close proximity to the *atz*D genes. The product of the AtzE-catalyzed reaction is allophonate that will spontaneously decarboxylate to form urea, which many organisms can metabolize. This may account for only one AtzE/F homolog being present in these organisms. The other strains, though lacking AtzE/F homologs, have other potential amidases belonging to the cysteine hydrolase or nitrilase superfamilies in close proximity, which could be catalyzing AtzE/F type reactions.

Azorhizobium caulinodans ORS 571 has two cyanuric acid hydrolase/barbiturase family members. The protein encoded by locus AZC_3892 was placed in Group 3, while locus AZC_3203 was not placed in either Clade 1 or 2 (Fig. 3). Upstream of the locus AZC_3892 gene is a metal-dependent amidohydrolase superfamily member similar to proteins that are in the upper pathway in atrazine degradation leading to cyanuric acid production. Downstream of the gene is a urea/short chain amide transport system. A cysteine hydrolase superfamily protein upstream may also be involved in hydrolysis of the AtzD homolog products. Interestingly, the AZC_3203 locus has one *atz*E/F homolog directly downstream of the *atz*D homolog gene.

The identical protein in *Acidithiobacillus ferrooxidans* ATCC 53993 and ATCC 23270 was placed adjacent to Group3 in Fig. 3, but was not formally included in Group 3 for reasons stated above. These strains are γ-Proteobacteria isolated from acid mine drainage. The AtzD homolog has 34-46% identity with the experimentally determined cyanuric acid hydrolases or barbiturases. It has the ability to fix nitrogen, which is common to most of the Group 3 source organisms. The genomes revealed a combination of genes around the AtzD homolog similar to multiple other groups. There are uracil permease and dehydrogenase genes similar to Group 4 (described below), AtzE/F homologs similar to group 5 (described below) and some examples in Groups 2 and 3, and multiple metal-dependent amidohydrolases, which are homologous to the upstream pathway genes in *s*-triazine degradation. This last example demonstrates the variability of the regions found flanking the Clade 1 genes. While some genes, such as those encoding amidase proteins, were observed adjacent to *atzD* homologs with some frequency, no consistent patterns in the genes flanking Clade 1 *atzD* homologs could be identified.

Gene context – Clade 2. Clade 2 consists of proteins found solely in certain Actinobacteria species (Fig. 2). The proteins in this clade share greater than 72% identity with the experimentally determined barbiturase from *Rhodococcus erythropolis* JCM 3132. With such similarity, it might be suspected that these proteins belong in the same group and catalyze the same reaction. However, gene context data suggested the division of this clade into two groups.

Group 4 genes encode proteins that have an upstream enzyme belonging to the amino acid kinase superfamily (Fig. S2). This superfamily includes such enzymes as uridylate, carbamate, asparto-, and acetylglutamate kinase. Uridylate kinase is a bidirectional enzyme that converts uridine 5'-diphosphate (UDP) to uridine 5'-monophosphate (UMP). UMP can then be converted to uridine and then uracil by 5' nucleotidase (EC 3.1.3.5), and uridine nucleosidase (EC 3.2.2.3) or uridine phosphorylase (EC 2.4.2.3) to feed into the pathway for uracil conversion to barbituric acid, and then malonate and urea (Fig. 1). Although the Group 4 proteins have high sequence identity to the known *Rhodococcus erythropolis* JCM 3132 barbiturase (74-86%), no enzyme in this group has been purified and the function of the Group 4 proteins has not been validated. The one organism in this group that lacks the kinase gene is *Nakamurella multipartite* DSM 44233. This organism has transposons flanking the barbiturase homolog, suggesting a mobilization event may have occurred to produce the current gene arrangement.

The second grouping in Clade 2, Group 5, contains the experimentally determined barbiturase from *Rhodococcus erythropolis* JCM 3132 and proteins with greater than 92% identity to this enzyme, and are found within other *Rhodococcus* species and *Saccharomonospora viridis* DSM 43017 (Fig. 3). Gene context data (Fig. S2B) revealed that this group has a conserved amidase homologous to AtzE/AtzF directly after the barbiturase. We propose that, analogous to the cyanuric acid degradation pathway in *P*. ADP, this amidase catalyzes the next step in the barbituric acid degradation pathway, the conversion of ureidomalonic acid to malonate and urea (Fig. 1). Although there is no genome for *Rhodococcus erythropolis* JCM 3132, two other *Rhodococcus erythropolis* species, PR4 and SK121, show conserved uracil/xanthine permease and uracil/thiamine dehydrogenase (small, middle, and large subunits) in close proximity (Fig. S2B). These enzymes catalyze the transport and conversion of uracil to barbituric acid, respectively. The other *Rhodococcus* strains in this group have these enzymes as well.

Due to a possible gene duplication or horizontal gene transfer event, *Nocardioides* JS614 has two proteins within Clade 2. The protein encoded by gene locus Noca_4270 has the uracil/thiamine dehydrogenase (small, middle, and large subunits), the enzyme producing barbituric acid, encoded directly upstream of the barbiturase homolog. This protein however shares only 72% identity with the experimentally determined barbiturase. Noca_1505 shares 86% identity with barbiturase but has the gene context described in Group 4. It is possible that both group 4 and 5 proteins catalyze the barbiturase reaction, that Noca_4270 encodes a protein having a different function, or that the coding sequence is a pseudogene. Additional experimentation is required to confirm these functions.

Frankia sp. EuI1c is an actinobacterium nitrogen-fixing symbiont found within root nodules of woody trees and shrubs. The cyanuric acid hydrolase/barbiturase homolog in this strain has the lowest sequence identity to the enzymes of known function, 36-40%. The genome sequence does not reveal any genes related to cyanuric acid or barbaturic acid metabolism. Experimental investigation is required to determine the function of this protein.

Fig. S2

Whole family sequence alignment. 80% identity are highlighted by the various colors with the most stringent condition being displayed: 80% identity with a single residues are colored with a grey background, negative residue conservation are red (DE), Ser/Thr are cyan, aliphatic are grey with a yellow background (ILV), positive are blue (HKR), tiny are green (AGS), aromatic are blue with a yellow background (FHWY), charged are pink (DEHKR), small are forest green (ACDGNPSTV), polar are light blue (CDEHKNQRSTV), big residues are dark blue with a peach background (EFHIKLMQRWY), and hydrophobic are black with a yellow background (ACFGHILMTVWY).

```
Janibacter_sp._HTCC2649          ----MP---E AIEVRKVPLH SVSDASELAK LIDDGVMEAS RVIAIIGKTE GNGGVNDYTR
Intrasporangium_calvum_DSM_430   ----MP---E AIEVRKVPIH SVADASELAR LIDDGVMEAA RVVAIIGKTE GNGGVNDYTR
Noca_1505_1Nocardioides_sp._JS   ----MP---D AIEVRKVPIH SVADASELAK LIDDGVMQAE RVIAIIGKTE GNGGVNDYTR
PR4_Rhodococcus_erythropolis_P   ----MP---E AIEVRKVPLH SVSDASELAK LIDDGVLEAD RVIAVIGKTE GNGGVNDYTR
SK121_Rhodococcus_erythropolis   ----MP---E AIEVRKVPLH SVSDASELAK LIDDGVLEAD RVIAVIGKTE GNGGVNDYTR
Barbiturase_Rhodococcus_erythr   ----MP---E AIEVRKVPLH SVSDASELAK LIDDGVLEAD RVIAVIGKTE GNGGVNDYTR
jostii_RHA1_Rhodococcus_jostii   ----MP---D AIEVRKVPLH SVSDAGELAK LIDDGVLEAD RVVAVIGKTE GNGGVNDYTR
B4_Rhodococcus_opacus_B4         ----MP---E AIEVRKVPLH SVSDAGELAK LIDDGVLEAD RVVAVIGKTE GNGGVNDYTR
Saccharomonospora_viridis_DSM_   ----MP---D AIEVRKVPIH SVSDASELAK LIDDGVMEAD RIIAIIGKTE GNGGVNDYTR
Nakamurella_multipartita_DSM_4   (11)TP---E PIEVRKIPLH SVSDASELTA LFDAGIIDAD RVVAVIGKTE GNGGVNDYTR
Catenulispora_acidiphila_DSM_4   ----MP---E AIEVRKVPLR HVSDASGLAE LITAGVVEAD RVVAVVGKTE GNGGVNDYTR
Streptosporangium_roseum_DSM_4   ----MP---E PIEVRKVAIE SVTDASGLTR LIDDGVIEAH RVLAVIGKTE GNGGVNDYTR
Thermobispora_bispora_DSM_4383   ----MP---D PIEVRKVPIE SVTDASGLAK LIDEGIIEAD RVLAVIGKTE GNGGVNDYTR
Geodermatophilus_obscurus_DSM_   ----MP---A PIEVRKVPLH NVSDASELAK LIDDGVMEAD RVVAVIGKTE GNGGVNDYTR
Noca_4270_2Nocardioides_sp._JS   ----MT---R PIEVRKVPIE HVSDAAGLAD LIDAGVFSAD DVIAVVGKTE GNGGVNDYTR
Aminobacter_aminovorans          ---------- ---------- ---------- ---------- ---------- ----------
Chelatobacter_heintzii           ---------- ---------- ---------- ---------- ---------- ----------
TRZD_Enterobacter_cloacae        ---------M QAQVFRVPMS NPADVSGVAK LIDEGVIRAE EVVCVLGKTE GNGCVNDFTR
TrzD_Acidovorax_citrulli         ---------M QAQVFRVPMS NPADVSGVAK LIDEGVIRAE EVVCVLGKTE GNGCVNDFTR
Moorella_thermoacetica_ATCC_39   --------MQ KVEVFRIPTA SPDDISGLAT LIDSGKINPA EIVAILGKTE GNGCVNDFTR
Clostridium_asparagiforme_DSM_   ---------M KIDVFKIEMS APDDVRELKK LIDAGEVDPG EIIAVLGKTE GNGCVNDFTR
AD25_Arthrobacter_sp._AD25       --------MY HIDVFRIPCH SPGDTSGLED LIETGRVAPA DIVAVMGKTE GNGCVNDYTR
Comamonas_sp._A2                 ---------- ---------- ---------- ---------- ---------- ----------
gi|75369013|sp|Q6PMK9|Q6PMK9_9   --------MY HIDVFRIPCH SPGDTSGLED LIETGRVAPA DIVAVMGKTE GNGCVNDYTR
AtzD_ADP_Pseudomonas_sp._ADP     --------MY HIDVFRIPCH SPGDTSGLED LIETGRVAPA DIVAVMGKTE GNGCVNDYTR
Pseudomonas_nitroreducens        ---------- ---------- --GDTSGLED LIETGRVAPA DIVAVMGKTE GNGCVNDYTR
USDA_110_Bradyrhizobium_japoni   ----MR--TT SVGVFKIVTK GPGDVSGLMA MIGSGAIDPK SILAVLGKTE GNGGVNDFTR
viciae_3841_Rhizobium_legumino   ----MP--SL RAHVFRVPAD GPDDVAGVEA LFASG-LQAN NIVAVLGKTE GNGCVNDFTR
trifolii_WSM1325_Rhizobium_leg   ----MP--SL RAHVFRVPAD GPDDVAGVEA LFASG-LQAN NVVAVLGKTE GNGCVNDFTR
Agrobacterium_vitis_S4           ---MTT--CF RASVYRVAAS APDDISGVEA LIGSG-LNPH SIVAVLGKTE GNGCVNDFTR
ORS278_Bradyrhizobium_sp._ORS2   ----MPTPLR RAHVHRLPMR SPDDVGALEA AIVSGAIDPA GIVAILGKTE GNGCVNDFTR
BTAi1_Bradyrhizobium_sp._BTAi1   ----MPTALR RATVHRLPMR SPDDVGALDA AIAQGAIDPA GIIAILGKTE GNGCVNDFTR
4_46_Methylobacterium_sp._4_46   ----MP---R RAEILRLPMA APDDVSAIAA SLRDGRLDPG DVVAVFAKTE GNGCVNDFTR
AZC_3892_2Azorhizobium_caulino   ----MP---- IAKVHRIATA SPDDVSGLAA AIATGAIAPA GILAIFGKTE GNGCVNDFSR
Paenibacillus_sp._JDR-2          --------MM KCSVIKVPAR SPNDTAALQR YIDESLINPK DVVAVLGKTE GNGCVNDFTR
Oceanicola_granulosus_HTCC2516   ---------M RADVIRFDLP DPGDVSALAG AIDAGRVDPG RIVAIIGKTH GNGLVNDYTR
Acidithiobacillus_ferrooxidans   ---------M RTSVYRFFTT DPADVSGLVR AIAEGVINPQ SIVAIIGKTE GNGGMNDFTR
JCM_2831_Methylobacterium_radi   --------MP ECLVFRVPAQ HPADVSGVMD LVTSGTVGAP EIVAIFGKTE GNGCVNDFTR
Micromonas_sp._RCC299            ----MVLDKP VAVVAVVDTA HPGDTGPLER WMDAN--PLA DILCVLGKTE GNGCVNDFTR
AZC_3203_1Azorhizobium_caulino   ---------M KVGVHKLAMS APGDVSELAA LIETGAVNPR EIVALVGKTE GNGGANDFTR
Bacillus_cellulosilyticus_DSM_   ---------M NYELIRCDMD HPGDVSALAK LLNEEEISAE SIRAIIAQTE GDG----YSR
Frankia_sp._EuI1c                (10)RSARYP KASMHVVPML APNDTAAFRA LFASGAVDPA SVVALIAKSE GSGLHNDHAR
Consensus/80%                    ...MMPh... .hpVb+lshp sssDst.l.. blssG.lps. pllAllGKTE GNGsVNDaTR
```

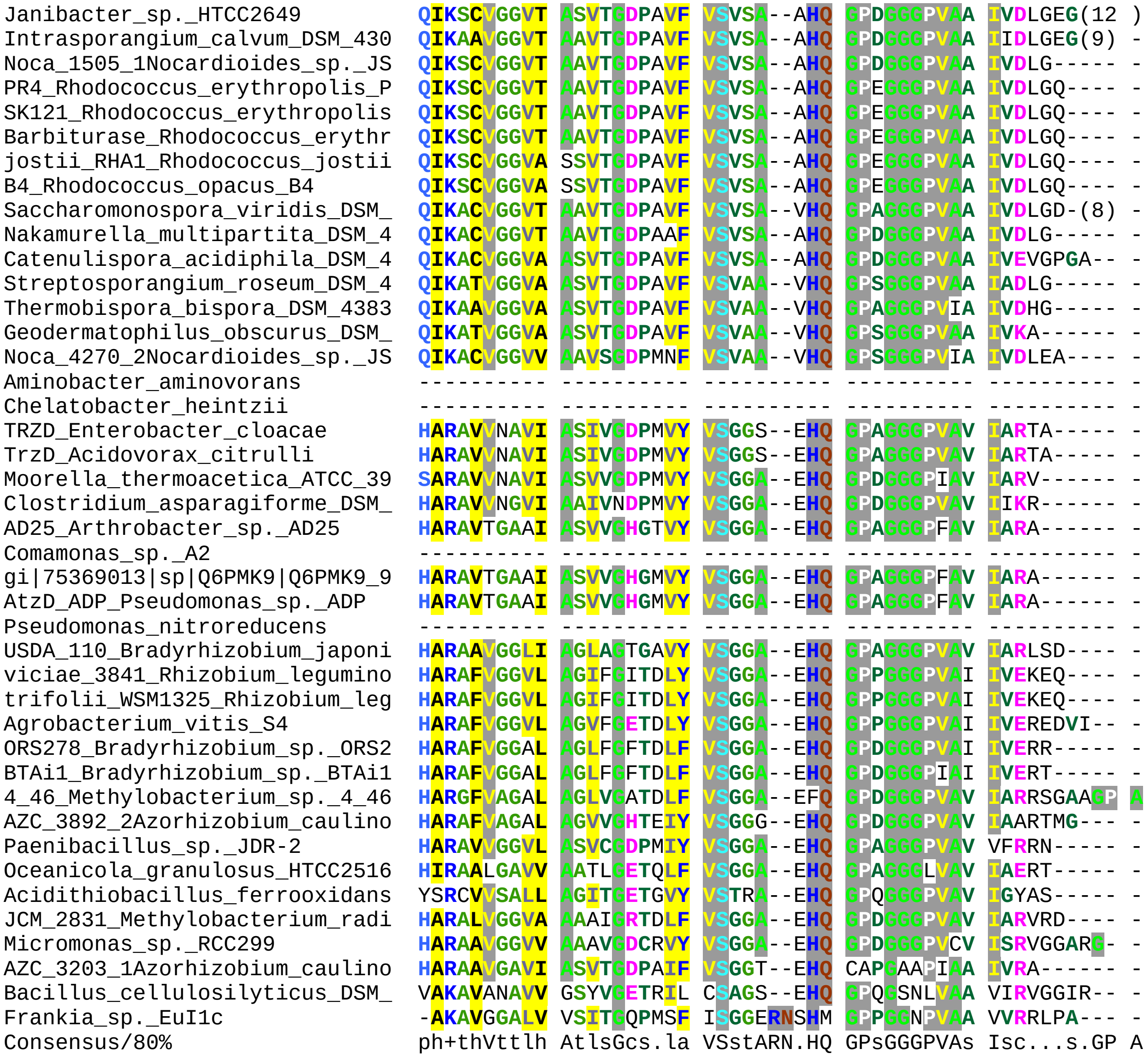
```
Janibacter_sp._HTCC2649          IIADRAFREV LIEKG-APA- ----DQVKQV PIVWSGGTDG IISPHATIFA TVP----ADK
Intrasporangium_calvum_DSM_430   IIADRAFRQV LVEKG-APA- ----DQVSQI PIVWSGGTDG VISPHATIFA TLP----ADR
Noca_1505_1Nocardioides_sp._JS   IIADRAFREV LVEKG-APA- ----EQVKQV PIVWSGGTDG VISPHATIFA TVP----PED
PR4_Rhodococcus_erythropolis_P   IIADRAFREV LSAKGNRSP- ----EEVAEV PIVWSGGTDG VISPHATIFA TVP----ADK
SK121_Rhodococcus_erythropolis   IIADRAFREV LSAKGNRSP- ----EEVAEV PIVWSGGTDG VISPHATIFA TVP----ADK
Barbiturase_Rhodococcus_erythr   IIADRAFREV LSAKGNRSP- ----EEVAEV PIVWSGGTDG VISPHATIFA TVP----ADK
jostii_RHA1_Rhodococcus_jostii   IIADRAFREV LSAKGRRSA- ----DEVAEV PIVWSGGTDG VISPHATIFA TVP----AGK
B4_Rhodococcus_opacus_B4         IIADRAFREV LSAKGSRST- ----DEVAQV PIVWSGGTDG VISPHATIFA TVP----AEK
Saccharomonospora_viridis_DSM_   IIADRAFREV IVEKGSRTM- ----EEVRQI PIVWSGGTDG VISPHATVFA TVP----EDR
Nakamurella_multipartita_DSM_4   IIADRAFREA IVAGGTRTL- ----DQVRQI PIVWSGGTDG VLSPHATVFA TVP----ADS
Catenulispora_acidiphila_DSM_4   IIADRAFRET LMRLGTRSE- ----AEVAEV PIVWSGGTDG VLSPHATIFA TTD----AEP
Streptosporangium_roseum_DSM_4   ILADRAFREV LAAKGHPSP- ----ES---V PLVWSGGTDG VLSPHATIFA TTA----DAE
Thermobispora_bispora_DSM_4383   ILADRAFRDV LVQKGTRTP- ----EEVAQV PLVWSGGTDG VLSPHATIFA TVD----PAK
Geodermatophilus_obscurus_DSM_   IIADRAFREV LLEKGTRSK- ----EEVGEI PIVWSGGTDG VISPHATVFA TLP----EDA
Noca_4270_2Nocardioides_sp._JS   IISTHAYRAV LEEKGTRSK- ----EEVAQV PLVWSGGTDG VISPHATIFA YAP----EGR
Aminobacter_aminovorans          ---------- -----GVSR- ---QEVGERI AFIMSGGTEG VMAPHCTIFT VQKTDNKQKT
Chelatobacter_heintzii           ---------- -----GVSR- ---QEVGERI AFIMSGGTEG VMAPHCTIFT VQKTDNKQKT
TRZD_Enterobacter_cloacae        GYTTLAFKVY FSEKLGVSR- ---QEVGERI AFIMSGGTEG VMAPHCTIFT VQKTDNKQKT
TrzD_Acidovorax_citrulli         GYTTLAFKVY FSEKLGVSR- ---QEVGERI AFIMSGGTEG VMAPHCTIFT VQKTDNKQKT
Moorella_thermoacetica_ATCC_39   GFATQSLAMY LAEKLGISR- ---EEVVKKV AFIMSGGTEG VMTPHITVFV RKDVQ---EP
Clostridium_asparagiforme_DSM_   ALSTVSFRSL IMAETGWSE- ---AEVSKKV AFVMSGGTEG VMSPHATIFC KHEDG---AA
AD25_Arthrobacter_sp._AD25       EYATAMLAAC LGRHLQLPP- ---HEVEKRV AFVMSGGTEG VLSPHHTVFA RRP----AID
Comamonas_sp._A2                 -YATAMLAAC LGRHLQLPP- ---HEVEKRV AFVMSGGTEG VLSPHHTVFA RRP----AID
gi|75369013|sp|Q6PMK9|Q6PMK9_9   EYATAMLAAC LGRHLQLPP- ---HEVEKRV AFVMSGGTEG VLSPHHTVFA RRP----AID
AtzD_ADP_Pseudomonas_sp._ADP     EYATAMLAAC LGRHLQLPP- ---HEVEKRV AFVMSGGTEG VLSPHHTVFA RRP----AID
Pseudomonas_nitroreducens        EYATAMLAAC LGRHLQLPP- ---HEVEKRV AFVMSGGTEG VLSPHHTVFA RRP----AID
USDA_110_Bradyrhizobium_japoni   EYAVAALCTA LAPQLGLSP- ---EEVEQRI AFVMSGGTEG VLSPHITVFT RRE----VER
viciae_3841_Rhizobium_legumino   GYATRSFETL FSRYG----- ------VDGV SIIMSGGTEG ALSPHWTVFA RET----VET
trifolii_WSM1325_Rhizobium_leg   GYATRSFETL FSRYG----- ------VDGV SIIMSGGTEG ALSPHWTVFA RET----VET
Agrobacterium_vitis_S4           GFATATFEQL FARLC----- ------VTGV SIVMSGGTEG ALSPHWTIFA RER----VEA
ORS278_Bradyrhizobium_sp._ORS2   AFAVRSLETL LGRHLQAD-- -----AVRQI AMVMSGGTEG ALSPHMIVFE AR--T--GDD
BTAi1_Bradyrhizobium_sp._BTAi1   AFAVRSLQTL LGRHVDGD-- -----AINQI AMVMSGGTEG ALSPHMIVFE ARDAS--GET
4_46_Methylobacterium_sp._4_46   PLAVQALRGL FGPLIGEA-- -----ALGRI AMVMSGGTEG GLSPHWLVIA AR-----EAA
AZC_3892_2Azorhizobium_caulino   GFAVQSLQML LRGHMG-A-- -----AADEV CLVMSGGTEG GMSPHFLVFE RA-----EGN
Paenibacillus_sp._JDR-2          GFAVLALKNL FASYK-EA-- -----TAEPV SYIMSGGTEG ILSPHFSVIS RSE-S--LDD
Oceanicola_granulosus_HTCC2516   GYLTQSLLKL LAARTGEAE- ---AGLQARI PMVFSGGVEG VLSPHFIVFA VHD----DED
Acidithiobacillus_ferrooxidans   TLAMMAVSQV LATHLDCRP- ---ELVEERV VLSFSGGSEG VAAPHILVFA VSGRA--AEQ
JCM_2831_Methylobacterium_radi   QLAVMALETA LATALGCTP- ---AEAGARI ALVMSGGTEG GLSPHFLILA RRE----GPA
Micromonas_sp._RCC299            GYATSAIQAA LGRVRSRRM( 12)SSFTADP AIIMSGGTEG VLCPHFIVFA NSWDE--ADA
AZC_3203_1Azorhizobium_caulino   GLATLSYQLL LARHLGLSP- ---EEVGQRI AFVWSGGTEG VLSPHATLFT RAP-----DD
Bacillus_cellulosilyticus_DSM_   GYATLAFQLL LSEKLNISQ- ---QEVFDTI PMMMIGKTGG LMTPHYTLFI EKED---VEK
Frankia_sp._EuI1c                VFADVSLRTA LAEARGCP-- --VEDLADSV TVAVSGGSPG VISPHVTVVT QEWVADLPAG
Consensus/80%                    hbAs.tbp.h ls.bh..s.. ..Vp-...pl sblbSGGT-G lbSPHh*lFs ...ssNK...
```

```
Janibacter_sp._HTCC2649         AE----QSDE PRLTVGFAMS EAILPEEIGY VGMVTKVADA VKVAMERAGI TDPADVHYVQ
Intrasporangium_calvum_DSM_430  TE----PSDD LRLTVGFAMS EPIKPEEIGY TAMVEKVAAG VQVAMERAGI TDPADVHYVQ
Noca_1505_1Nocardioides_sp._JS  LTGALAPSDE QRLTVGFAMS ERLAPEDIGR TAMITKVADA VKVAMERAGI SDPADVHYVQ
PR4_Rhodococcus_erythropolis_P  VT----KTDE PRLTVGVAMS EQLLPEDIGR TAMITKVAAA VKDAMADAGI TDPADVHYVQ
SK121_Rhodococcus_erythropolis  VT----KTDE PRLTVGVAMS EQLLPEDIGR TAMITKVAAA VKDAMADAGI TDPADVHYVQ
Barbiturase_Rhodococcus_erythr  VT----KTDE PRLTVGVAMS EQLLPEDIGR TAMITKVAAA VKDAMADAGI TDPADVHYVQ
jostii_RHA1_Rhodococcus_jostii  AT----KTDE PRLTVGVAMS EQLLPEDIGR TPMITKVAAA VKMAMANAGI TDPADVHYVQ
B4_Rhodococcus_opacus_B4        AT----KTEE PRLTVGVAMS EQLLPEDIGR TAMITKVAAA VKMAMDNAGI TDPADVHYVQ
Saccharomonospora_viridis_DSM_  VE----KTDE PRLTVGIAMS EPILPEEIGR RPMIEKVAAG VKDAMARAGI TDPADVHYVQ
Nakamurella_multipartita_DSM_4  VP----ATDE PRLTVGFAMS EQLRPEDIGR TAMITKVAAA VREAMARAGI TDPADVHYVQ
Catenulispora_acidiphila_DSM_4  CD-------E PRLTVGFAMS EPLLPEEIGR SAMITKVADA VHEAMKRAGI DDPADVHYVQ
Streptosporangium_roseum_DSM_4  P------GDE PRVSVGIAMS DVILPEDIGR PAMVEKVAAG VREAMKIAGI DDPADVHYVQ
Thermobispora_bispora_DSM_4383  AP----KTDE PRLSVGVAMS EPILPEDIGR PAMIEKVANG VREAMKIAGI EDPADVHYVQ
Geodermatophilus_obscurus_DSM_  VE----KTDE PRLTVGYAMS DVLLPEDIGR IAMVEKVADG VRRAMAEAGI TDPADVHYVQ
Noca_4270_2Nocardioides_sp._JS  YL----PTDE PRVTVGYAMS EVLLPEDIGR PAMVEKVAAG VRVAMERAGI TDPADVHYVQ
Aminobacter_aminovorans         AA------EG KRLAVQQIFT REFLPEEIGR MPQVTETADA VRRAMREAGI ADASDVHFVQ
Chelatobacter_heintzii          AA------EG KRLAVQQIFT REFLPEEIGR MPQVTETADA VRRAMREAGI ADASDVHFVQ
TRZD_Enterobacter_cloacae       AA------EG KRLAVQQIFT REFLPEEIGR MPQVTETADA VRRAMREAGI ADASDVHFVQ
TrzD_Acidovorax_citrulli        AA------EG KRLAVQQIFT REFLPEEIGR MPQVTETADA VRRAMREAGI ADASDVHFVQ
Moorella_thermoacetica_ATCC_39  AK------PG KRLAVGVAFT RDFLPEELGR MEQVNEVARA VKEAMKDAQI DDPRDVHFVQ
Clostridium_asparagiforme_DSM_  GT------GE KRLAVSSGFT RDFLPEEQGT MAVVEECRRV TLELMAQAGI DDPADVHFVQ
AD25_Arthrobacter_sp._AD25      AHR----PAG KRLTLGIAFT RDFLPEEIGR HAQITETAGA VKRAMRDAGI ASIDDLHFVQ
Comamonas_sp._A2                AHR----PAG KRLTLGIAFT RDFLPEEIGR HAQITETAGA VKRAMRDAGI ASIDDLHFVQ
gi|75369013|sp|Q6PMK9|Q6PMK9_9  AHR----PAG KRLTLGIAFT RDFLPEEIGR HAQITETAGA VKRAMRDAGI ASIDDLHFVQ
AtzD_ADP_Pseudomonas_sp._ADP    AHR----PAG KRLTLGIAFT RDFLPEEIGR HAQITETAGA VKRAMRDAGI ASIDDLHFVQ
Pseudomonas_nitroreducens       AHR----PAG KRLTLGIAFT RDFLPEEIGR HAQITETAGA VKRAMRDAGI ASIDDLHFVQ
USDA_110_Bradyrhizobium_japoni  RPAG---LSG KRLSIGMAHT RDFLPEELGR AAQIAETAAA VKAAMADAGI ADPADVHFVQ
viciae_3841_Rhizobium_legumino  PG-------E RALAIGVSRT PALSPEHLGR REQILLVAEG VKSAMRDAGI DDPADAHFVQ
trifolii_WSM1325_Rhizobium_leg  PG-------E RALAIGVSRT PALPPEHLGR REQILLVAEG VKSAMRDAGI DDPADVHFVQ
Agrobacterium_vitis_S4          AP-------N RSLAIGVART ASLLPEHLGR KEQVDLVAQG VRAAMADAGI DDPADVHFVQ
ORS278_Bradyrhizobium_sp._ORS2  GPRA----FG TSLALGRART PTLPGEHLGR MQQVAQVAAG VRAAMADAGI EDPADVHYVQ
BTAi1_Bradyrhizobium_sp._BTAi1  QPRA----FD KSLALGRART PLLPSEHLGR MQQVAQVAAG VRAAMADAGI ADAADVHYVQ
4_46_Methylobacterium_sp._4_46  GP-------G PALAVGQART PPLAAEDLGR RAQVEMVAAG VRAAMREAG- LDAGQVHYVQ
AZC_3892_2Azorhizobium_caulino  AP-E----AA PALAIGRAHT PDLPFEALGR MGQVRMVAQA VRRAMAAAGI TDPEDVHFVQ
Paenibacillus_sp._JDR-2         AIKR----EQ SSLAIGVART PDFMPEELGR LAQVDQVAEA VRRAMDGAGI HDAADVHFVQ
Oceanicola_granulosus_HTCC2516  AP------TE PGLAVATAMT PPLGPEDLGR QRQIDLTAAA VEEAMARADI HSPDDVHFVQ
Acidithiobacillus_ferrooxidans  AY------PI KRLALASAHT RPFSAQEVGR MAMVRETARA VTRLMHTLQV E-PADVHLVQ
JCM_2831_Methylobacterium_radi  AP------ER KALAIGTAFT HAFRPEEIGR MSQVEATAAA VGRAMAEAGL TDPADVHLVQ
Micromonas_sp._RCC299           DEN(4)DDDD AAFELHVTRT RDFKPEEVGT PVMARATRRA IEEYLGYGGN APNWDDMFVQ
AZC_3203_1Azorhizobium_caulino  GP----MPAE PRLALGIGIT RDIAPEEVGT TAMVEAVAGA VHTALAEAGI TEPADVHYVQ
Bacillus_cellulosilyticus_DSM_  EN------KE KRFSFGVAST RVLLPEEIGT VAQVDLVADA VNKAMLEAGI ESIDDVKCVE
Frankia_sp._EuI1c               LP-------G VGLVVGRGHT EPILPEDIGR TAQVDKVADA VAAAMLDAGV TDPDDVHLVM
Consensus/80%                   s.p.LA.s.. .pLslG.tb* ..b.PE-lGR .sblpbsA.t V+.AM.pAGI s-ssDVHaVQ
```

```
Janibacter_sp._HTCC2649         TKTPLLTIHT IRDAKSRGQK VWTEHT---H ESMDLSNGTT GLGIAVALG- ---EIEMP--
Intrasporangium_calvum_DSM_430  TKTPLLTIHT IRDAKSRGMK VWTEHT---H ESMDLSNGTT GLGVAVALG- ---EIEMP--
Noca_1505_1Nocardioides_sp._JS  TKTPLLTIHT IRDAKSRGKT VWTEHT---H ESMDLSNGCT ALGIAVALG- ---EIEMP--
PR4_Rhodococcus_erythropolis_P  TKTPLLTIHT IRDAKSRGKT VWTEQT---H ESMDLSNGGT ALGIAVALG- ---EIDMP--
SK121_Rhodococcus_erythropolis  TKTPLLTIHT IRDAKSRGKT VWTEQT---H ESMDLSNGGT ALGIAVALG- ---EIDMP--
Barbiturase_Rhodococcus_erythr  TKTPLLTIHT IRDAKSRGKT VWTEQT---H ESMDLSNGGT ALGIAVALG- ---EIDMP--
jostii_RHA1_Rhodococcus_jostii  TKTPLLTIHT IRDAKSRGET VWTEHT---H ESMDLSNGVT ALGIAVALG- ---EIDMP--
B4_Rhodococcus_opacus_B4        TKTPLLTIHT IRDAKSRGET VWTEHT---H ESMDLSNGVT ALGIAVALG- ---EIDMP--
Saccharomonospora_viridis_DSM_  TKTPLLTIDT IRDAKSRGKS VWTEDT---L RSMDLSNGTT GLGIAVALG- ---EIDMP--
Nakamurella_multipartita_DSM_4  TKTPLLTIHT IRDAKQRGQT VWTEHT---H ESMDLSNGCT ALGVAVALG- ---EIEMP--
Catenulispora_acidiphila_DSM_4  TKTPLLTIQT IREAKARGET VWTEHT---H ESMDLSNGCT ALGIAVALG- ---EVEMP--
Streptosporangium_roseum_DSM_4  TKTPLLTLAT INDAKSRGKD VVIEDT---G PSMDISNSTT ALGVAVALG- ---EIEMP--
Thermobispora_bispora_DSM_4383  TKTPLLTLET INDAKRRGHT VVTEDT---L KSMDISNSTT ALGIAVALG- ---EIEMP--
Geodermatophilus_obscurus_DSM_  TKTPLLTIET IRDAKARGEE TYYDEP---H GSMDLSNGTT ALGIAVALG- ---EIDMP--
Noca_4270_2Nocardioides_sp._JS  TKTPLLVQDT INDAERRGET VYTHNT---L ESMDVSNATT ALGIAVALG- ---EIEMP--
Aminobacter_aminovorans         VKCPLLTAGR MHDAVERGHT VATEDT---Y ESMGYSRGAS ALGIALALG- ---EVEKANL
Chelatobacter_heintzii          VKCPLLTAGR MHDAVERGHT VATEDT---Y ESMGYSRGAS ALGIALALG- ---EVEKANL
TRZD_Enterobacter_cloacae       VKCPLLTAGR MHDAVERGHT VATEDT---Y ESMGYSRGAS ALGIALALG- ---EVEKANL
TrzD_Acidovorax_citrulli        VKCPLLTAGR MHDAVERGHT VATEDT---Y ESMGYSRGAS ALGIALALG- ---EVEKANL
Moorella_thermoacetica_ATCC_39  IKCPLLTAER IEDAKRRGKD VVVNDT---Y KSMAYSRGAS ALGVALALG- ---EISADKI
Clostridium_asparagiforme_DSM_  IKCPLLTSER ILDAKSRGKD VVVHDT---Y ESMGYNRGAS ALGVAIALG- ---EIPAEAV
AD25_Arthrobacter_sp._AD25      VKCPLLTPAK IASARSRGCA PVTTDT---Y ESMGYSRGAS ALGIALATE- ---EVPSSML
Comamonas_sp._A2                VKCPLLTPAK IASARSRGCA PVTTDT---Y ESMGYSRGAS ALGIALATE- ---EVPSSML
gi|75369013|sp|Q6PMK9|Q6PMK9_9  VKCPLLTPAK IASARSRGCA PVTTDT---Y ESMGYSRGAS ALGIALATE- ---EVPSSML
AtzD_ADP_Pseudomonas_sp._ADP    VKCPLLTPAK IASARSRGCA PVTTDT---Y ESMGYSRGAS ALGIALATE- ---EVPSSML
Pseudomonas_nitroreducens       VKCPLLTPAK IASARSRGCA PVTTDT---Y ES-------- ---------- ----------
USDA_110_Bradyrhizobium_japoni  IKCPLLTSDR VEAASARGNK TATTSA---Y GSMAYSRGAS ALGVAVALG- ---ETG-SDI
viciae_3841_Rhizobium_legumino  IKCPLLTSRR IAEAEAAGRT VATHDT---L KSMGLSRGAS ALGVAVALG- ---EIDATSI
trifolii_WSM1325_Rhizobium_leg  IKCPLLTSRR IAEAEAAGRT VATHDT---L KSMGLSRGAS ALGVAVALG- ---EIDATSI
Agrobacterium_vitis_S4          IKCPLLTSAR VAEAQARGKR TVIAET---L KSMGYSRGAS ALGVAVALG- ---ELDMDDI
ORS278_Bradyrhizobium_sp._ORS2  VKCPLLTMER IEAAEARGAR TAVRDT---L KSMGFSRGAS ALGVAVALG- ---ELAFDEL
BTAi1_Bradyrhizobium_sp._BTAi1  VKCPLLTMER IEAAEARGTT TAVRDT---L KSMSFSRGAS ALGAAVALG- ---ELTLSDL
4_46_Methylobacterium_sp._4_46  VKCPLLTSER IGAALARGAA PATRDT---L KSMGLSRAAA ALGAALALG- ---EVPAAAI
AZC_3892_2Azorhizobium_caulino  VKCPLLTAMR VKEAEARGAT TATSDT---L KSMGLSRGAS ALGIALALG- ---EVAEDAL
Paenibacillus_sp._JDR-2         IKCPLLTAEQ VQEANLRGAS TVTEDT---Y KSMGYSRGAS ALGAAVALG- ---EIQRERL
Oceanicola_granulosus_HTCC2516  VKGPAFAQA- --DLAGAADG LASDDP---G KLMGFCRAAS AIGVARALG- ---EVPAAEA
Acidithiobacillus_ferrooxidans  IKG-AIPAP- --DSTTTGN- VRDEHL---R CDMAWSRGAS ALGVALALG- ---EVTESEL
JCM_2831_Methylobacterium_radi  VKCPLLTSAR IADAAARGHG VATHDT---Y ASMGLSRGAS ALGIALALG- ---EVDREQL
Micromonas_sp._RCC299           IKCPLLTPER IAEAAARGET CATTDC---Y ESMALSRGAS SLGAMVAVE( 10)DARVDDA
AZC_3203_1Azorhizobium_caulino  VKGPLLTPAT IADADRRGAR LVTRDP---N GSKPYARGAM ALGVALGLG- ---EVAAERI
Bacillus_cellulosilyticus_DSM_  VKCPWGVGG- ---------- ---------- ---SLAKATS ALGSAVALK- ---EVDREVI
Frankia_sp._EuI1c               VKGPALSSRA VADALSRGKT VVTGDY(4)M GSMCWSNDAS ALGVAVALG- ---EVKRDLV
Consensus/80%                   hKsPLLT..p l.sA..RGps sh*pcT...b cSMsbSpGs* ALGlAlALG. ...El..s.l
```

```
Janibacter_sp._HTCC2649         TDADVMHNRS LFSSVASCSS GVELDQAQIV VVG---NARG VGGRYRIGHS VMRDALDADG
Intrasporangium_calvum_DSM_430  SDEDIMHNRS LYSSVASCSS GVELDQAQVV VVG---NTRG IGGRYRIGHS VMRDALDADG
Noca_1505_1Nocardioides_sp._JS  SDEDVMHDRS LYSSVASCSS GVELDQAQVV VVG---NAPG VGGRYRIGHS VMKDALDQDG
PR4_Rhodococcus_erythropolis_P  TDEDVMHSRE LFSSVASCSS GVELDRAQIV VVG---NARG VGGRYRIGHS VMKDPLDQDG
SK121_Rhodococcus_erythropolis  TDEDVMHSRE LFSSVASCSS GVELDRAQIV VVG---NARG VGGRYRIGHS VMKDPLDQDG
Barbiturase_Rhodococcus_erythr  TDEDVMHSRE LFSSVASCSS GVELDRAQIV VVG---NARG VGGRYRIGHS VMKDPLDQDG
jostii_RHA1_Rhodococcus_jostii  SDADVMHRRD LFSSVASCSS GVELDRAQVV VVG---NARG IGGRYRIGHS VMKDPLDQDG
B4_Rhodococcus_opacus_B4        DDADVMHRRD LYSSVASCSS GVELDRAQIV VVG---NARG IGGRYRIGHS VMTDPLDQDG
Saccharomonospora_viridis_DSM_  SDEDVLKNRD LFSSVASCSS GVELDRAQIV VVG---NARG VGGRYRIGHS VMQDALDQDG
Nakamurella_multipartita_DSM_4  GDADVMHDRS LFSSVASCSS GVELDQAQVV VVG---NARG VGGRYRIGHS VMQDALDADG
Catenulispora_acidiphila_DSM_4  SDEAVLHDRA LYSSVASCSS GVELDRAQVV VVG---NASG VGGRYRIGHA VMRDALDAEG
Streptosporangium_roseum_DSM_4  TADQIHRDLS LYSSVASCSS GVELDRAQIV VVG---NVRG IGGRYRIGHS VMKDALDADG
Thermobispora_bispora_DSM_4383  REDQIHRDLS LYSSVASCSS GVELDRAQIV VMG---NVRG IGGRYRIGHS VMKDALDTDG
Geodermatophilus_obscurus_DSM_  EQGQVMRDLS LFSAVASCSS GVELDRAQIV VVG---NARG HGGAYRIGHS VMRDALDQDG
Noca_4270_2Nocardioides_sp._JS  TAEQIFHDLS LYSSVASCSS GVELDQAQIV VVG---NARG VGGRFRVGHS IMKDALDMDG
Aminobacter_aminovorans         SDEVITADYS LYSSVASTSA GIELMNNEII VMG---NSRA WGGDL----- ----------
Chelatobacter_heintzii          SDEVITADYS LYSSVASTSA GIELMNNEII VMG---NSRA WGGDL----- ----------
TRZD_Enterobacter_cloacae       SDEVITADYS LYSSVASTSA GIELMNNEII VMG---NSRA WGGDLVIGHA EMKDAIDGAA
TrzD_Acidovorax_citrulli        SDEVITADYS LYSSVASTSA GIELMNNEII VMG---NSRA WGGDLVIGHA EMKDAIDGAA
Moorella_thermoacetica_ATCC_39  SNEAICHDWN LYSSVASTSA GVELLNDEII VVG---NSTN SASDLVIGHS VMKDAIDADA
Clostridium_asparagiforme_DSM_  SQETICHDYS LYSKVASTSA GVELLNCEII LMG---NSVN STSDYRIGHS VMEDAIDFKA
AD25_Arthrobacter_sp._AD25      VDESVLNDWS LSSSLASASA GIELEHNVVI AIG---MSEQ ATSELVIAHG VMSDAIDAAS
Comamonas_sp._A2                VDESVLNDWS LS-------- ---------- ---------- ---------- ----------
gi|75369013|sp|Q6PMK9|Q6PMK9_9  VDESVLNDWS LSSSLASASA GIELEHNVVI AIG---MSEQ ATSELVIAHG VMSDAIDAAS
AtzD_ADP_Pseudomonas_sp._ADP    VDESVLNDWS LSSSLASASA GIELEHNVVI AIG---MSEQ ATSELVIAHG VMSDAIDAAS
Pseudomonas_nitroreducens       ---------- ---------- ---------- ---------- ---------- ----------
USDA_110_Bradyrhizobium_japoni  SDGDVLRRYD LFSKVASTSA GIELMHNVVI VLG---NSAA SASEFEIGHA VMNDAIDAAA
viciae_3841_Rhizobium_legumino  NDADICTRFD LFSRCASTSS GVELTDHEII VLG---MSAK WSGPLSIDHA VMRDAIDAHS
trifolii_WSM1325_Rhizobium_leg  GDADICTRFD LFSRCASTSS GGELTDHEII VLG---MSAK WSGPLSIDHA VMLDAIDAHS
Agrobacterium_vitis_S4          KDTDICRTLD LYSARASTSA GVELTDHEII VLG---MSDH WSGPFAMTHT VMFDAMDAPS
ORS278_Bradyrhizobium_sp._ORS2  SDADICADYS RYSERAATSG GVELLDHEIM VAG---MSAR WTGPLAIDHG VMRDAIDIEP
BTAi1_Bradyrhizobium_sp._BTAi1  SDAMICADYA RYSERAATSG GVELLDHEIM VAG---MSAD WTGPLAIDHA VMADAIDIEP
4_46_Methylobacterium_sp._4_46  GETVAETDPG RHARRCGASA GVELLDHEVV VMG---MSPD WTGPLVIDHA VMADAIDLRP
AZC_3892_2Azorhizobium_caulino  SDAVICADYG LWSARASCSS GIELLGHEIV VLG---MSEG WSGPLAIAHG VMADAIDVTP
Paenibacillus_sp._JDR-2         QDTDICHSWE LYSKVASTSA GSELAYCEVI VFG---NSAY AEGPYFIDHD VMKDAIDAGA
Oceanicola_granulosus_HTCC2516  NEAAVLRDFS VRSNVASCSA GLEVSCNEVI VIG---MAPG WSGPLAIAHA PMADALDVAA
Acidithiobacillus_ferrooxidans  SDDVINQDWS LYSSRASVSA KPLLQRSEMV LFA---NSCW WEGDLVIAHG VMQDIIDVPA
JCM_2831_Methylobacterium_radi  TDPAIGTRRD LFSGRASASA GIELMRNEII VLG---NAPG WTGPLAIAHR VMDDGIDLPA
Micromonas_sp._RCC299           VYAKTCEDYS VFSDHVSSSA GVELMHCEIL MML(6)SSKT PRSKFRMARG RMDDAIDLAG
AZC_3203_1Azorhizobium_caulino  TPDMIACDME VFSAVASTSA GGELTKCEVL LFG---NAPG ATSAFRIGHG VLKDAIDVAG
Bacillus_cellulosilyticus_DSM_  DEDSLNEDHS LFSVKTSVSA GQEQVAARVI VMG---NSVK STSNLYIGSG VMKDALDLKG
Frankia_sp._EuI1c               ADDRIRSDWD LFSAVAATSS GGEKRGGEVL LLA---NSAQ SASELRIGHG ITRDMADTEG
Consensus/80%                   s-.slhpcbs LaSshASsSt GlEL.pspll VhG....t.. hst.b.ItHt VMpDtlD..t
```

```
Janibacter_sp._HTCC2649         IWEAIKDAGL ELP--DRPHT SDIQGRLVNV FLKCEVSQDG QVRGRRNAML DDSDVHWH-R
Intrasporangium_calvum_DSM_430  IWEAIKDAGL DLP--ERPHW TDIDGRLVNV FLKCEVSQDG MVHGRRNAML DDSDVHWH-R
Noca_1505_1Nocardioides_sp._JS  IWEAIKDAGL DLP--ERPRT SDLDGRLVNV FLKCEASQDG LVRGRRNAML DDSDVHWH-R
PR4_Rhodococcus_erythropolis_P  IWAAIRDAGL ELP--ERPHS SDLDGQLVNV FLKCEASQDG TVRGRRNAML DDSDVHWH-R
SK121_Rhodococcus_erythropolis  IWAAIRDAGL ELP--ERPHS SDLDGQLVNV FLKCEASQDG TVRGRRNAML DDSDVHWH-R
Barbiturase_Rhodococcus_erythr  IWAAIRDAGL ELP--ERPHS NDLDGQLVNV FLKCEASQDG TVRGRRNAML DDSDVHWH-R
jostii_RHA1_Rhodococcus_jostii  IWAAIRDAGL ELP--ERPHT SDLDGQLVNV FLKCEASQDG TVRGRRNAML DDSDVHWH-R
B4_Rhodococcus_opacus_B4        IWSAIRDAGL ELP--ERPHT SDLDDRLVNV FLKCEASQDG TVRGRRNAML DDSDVHWH-R
Saccharomonospora_viridis_DSM_  IWQAIRDAGL KLP--ENPHP KDIDGRLVNV FLKCEASPDG TVRGRRNAML DDSDVHWH-R
Nakamurella_multipartita_DSM_4  IWAAIRDAGL DLP--DRPHP RDLDGRLVNV FLKCEASQDG QVRGRRNAML DDSDVHWH-R
Catenulispora_acidiphila_DSM_4  IWDAIRAAGL ELP--ERPRS SDLDGRLVNV FLKCEASTDG MVRGRRNAML DDSDVHWH-R
Streptosporangium_roseum_DSM_4  IWEAIRSSGI DLP--DRPHP SDLGGRLVNV FMKCEADPSG SVRGRRNIML DDSDVHWH-R
Thermobispora_bispora_DSM_4383  IWEAIRNAGL DLP--ERPHP SDLKGRLVNV FLKCEADPTG RVRGRRNIML DDSDVHWH-R
Geodermatophilus_obscurus_DSM_  IWNAIRDAGL DLP--ERPHP SDLGDRLVNV FLKCEADPTG YVRGRRNAML DDSDVFWH-R
Noca_4270_2Nocardioides_sp._JS  VWAAIRDAGL DDMPVDCIHP RHIKGRLVNL FLKCEADPTG RVRGRRNIML DDSDVAWH-R
Aminobacter_aminovorans         ---------- ---------- ---------- ---------- ---------- ----------
Chelatobacter_heintzii          ---------- ---------- ---------- ---------- ---------- ----------
TRZD_Enterobacter_cloacae       VRQALRDVGC CEN---DLPT VDELGRVVNV FAKAEASPDG EVRNRRHTML DDSDINST-R
TrzD_Acidovorax_citrulli        VRQALRDVGC CEN---DLPT VDELGRVVNV FAKAEASPDG EVRNRRHTML DDSDINST-R
Moorella_thermoacetica_ATCC_39  VRAALKDAGL KFD---CCPP AEELAKIVNV LAKAEAASSG TVRGRRNTML DDSDINHT-R
Clostridium_asparagiforme_DSM_  VIEAIQSAGL EVD---KYPT DEQAKEIVNV FAKAEADSTG YVRGRRNTMH TDSDINHT-R
AD25_Arthrobacter_sp._AD25      VRRTIESLG- IRSD----DE MDR---IVNV FAKAEASPDG VVRGMRHTML SDSDINST-R
Comamonas_sp._A2                ---------- ---------- ---------- ---------- ---------- ----------
gi|75369013|sp|Q6PMK9|Q6PMK9_9  VRRTIESLG- IRSD----DE MDR---IVNV FAKAEASPDG VVRGMRHTML SDSDINST-R
AtzD_ADP_Pseudomonas_sp._ADP    VRRTIESLG- IRSD----DE MDR---IVNV FAKAEASPDG VVRGMRHTML SDSDINST-R
Pseudomonas_nitroreducens       ---------- ---------- ---------- ---------- ---------- ----------
USDA_110_Bradyrhizobium_japoni  VTSALKCVG- LGVA----PQ AEAGRELVNI FAKAEASPDG SVRGFRHTML EDTDISST-R
viciae_3841_Rhizobium_legumino  VRKARERLP- ---------- -ENS-RLAAV LAKAEPDPSG EIDGRRHTML DDSDIAGT-R
trifolii_WSM1325_Rhizobium_leg  VRKARERLP- ---------- -ENS-RLAAV LAKAEPDPSG RIDGRRHTML DDSDIAGT-R
Agrobacterium_vitis_S4          VRAAWSDMP- ---------- -LG--KLSAV LAKAEPDPRG TVRGKRHTML EDSDISGT-R
ORS278_Bradyrhizobium_sp._ORS2  ARAALSRLGL DGP--GQLPA ASRG-RIAAV LAKAEAAQSG KVRGLRHTML DDSDVSST-R
BTAi1_Bradyrhizobium_sp._BTAi1  ARAALARLGI DSA--GQVPA AARG-RIAAV LAKAEAAQSG RLRGQRHTML DDSDIAST-R
4_46_Methylobacterium_sp._4_46  VAACLGRLGL LGPD-GFVDE AGRA-RLAAL LAKAEASADG AIRGRRHTML TDSDIAPT-R
AZC_3892_2Azorhizobium_caulino  VKAALSALG- ---------- AEAG-EATIV LAKAEPSRSG RIRGKRHTML DDSDISPT-R
Paenibacillus_sp._JDR-2         LKRILEKYP- ---------- --DA-EMVQV LAKAEADPTG YIRGQRHTML DDSDINHT-R
Oceanicola_granulosus_HTCC2516  VHHALAGVGL AAAP--QLAA GDAA-RLRAV FVKGEAARSG LIRGRGHTML NDGDIDQQ-R
Acidithiobacillus_ferrooxidans  IHAVLNQLGL QPVN-GQLAD VDTK-KVLAV FAKSDADPRG HIRHRRHTMW TDADISDM-R
JCM_2831_Methylobacterium_radi  ILGVLSDLGF PEAG--PLSR QESE-RVIAL LVKAEPSHDG LIRGRRHIMN DDSDINAT-R
Micromonas_sp._RCC299           VTRCVDAAR- ---------- GSRG-TNTIV TAIVKADPAA SVRGRRTTML SDSDIHAT-R
AZC_3203_1Azorhizobium_caulino  VKEALRSAGL SFHG---TPS EEQAHEIVAV FAKAEAPVNG RLRGRRTTML SDADIHYE-R
Bacillus_cellulosilyticus_DSM_  MQSAFKDAGI NHS---DTLS EEEQAKIVQV FVNAGADAIN EVRGRRHTMH TDALAMHSGI
Frankia_sp._EuI1c               IKTAIRTAGV DFD---CCLS PAQQAQVVQV FGKFVLPGSD VLRGQHITAL DDHEAHHV--
Consensus/80%                   lb.slpshGl ..ssV.p... .-..sclssV bhKsEss.sG .lRGpRpsML sDSDlp.pGR
```



```
Janibacter_sp._HTCC2649         QIKSCVGGVT ASVTGDPAVF VSVSA--AHQ GPDGGGPVAA IVDLGEG(12 )
Intrasporangium_calvum_DSM_430  QIKAAVGGVT AAVTGDPAVF VSVSA--AHQ GPDGGGPVAA IIDLGEG(9) -
Noca_1505_1Nocardioides_sp._JS  QIKSCVGGVT AAVTGDPAVF VSVSA--AHQ GPDGGGPVAA IVDLG----- -
PR4_Rhodococcus_erythropolis_P  QIKSCVGGVT AAVTGDPAVF VSVSA--AHQ GPEGGGPVAA IVDLGQ---- -
SK121_Rhodococcus_erythropolis  QIKSCVGGVT AAVTGDPAVF VSVSA--AHQ GPEGGGPVAA IVDLGQ---- -
Barbiturase_Rhodococcus_erythr  QIKSCVGGVT AAVTGDPAVF VSVSA--AHQ GPEGGGPVAA IVDLGQ---- -
jostii_RHA1_Rhodococcus_jostii  QIKSCVGGVA SSVTGDPAVF VSVSA--AHQ GPEGGGPVAA IVDLGQ---- -
B4_Rhodococcus_opacus_B4        QIKSCVGGVA SSVTGDPAVF VSVSA--AHQ GPEGGGPVAA IVDLGQ---- -
Saccharomonospora_viridis_DSM_  QIKACVGGVT AAVTGDPAVF VSVSA--VHQ GPAGGGPVAA IVDLGD-(8) -
Nakamurella_multipartita_DSM_4  QIKACVGGVT AAVTGDPAAF VSVSA--AHQ GPDGGGPVAA IVDLG----- -
Catenulispora_acidiphila_DSM_4  QIKACVGGVA ASVTGDPAVF VSVSA--AHQ GPDGGGPVAA IVEVGPGA-- -
Streptosporangium_roseum_DSM_4  QIKATVGGVA ASVTGDPAVF VSVAA--VHQ GPSGGGPVAA IADLG----- -
Thermobispora_bispora_DSM_4383  QIKAAVGGVA ASVTGDPAVF VSVAA--VHQ GPAGGGPVIA IVDHG----- -
Geodermatophilus_obscurus_DSM_  QIKATVGGVA ASVTGDPAVF VSVAA--VHQ GPSGGGPVAA IVKA------ -
Noca_4270_2Nocardioides_sp._JS  QIKACVGGVV AAVSGDPMNF VSVAA--VHQ GPSGGGPVIA IVDLEA---- -
Aminobacter_aminovorans         ---------- ---------- ---------- ---------- ---------- -
Chelatobacter_heintzii          ---------- ---------- ---------- ---------- ---------- -
TRZD_Enterobacter_cloacae       HARAVVNAVI ASIVGDPMVY VSGGS--EHQ GPAGGGPVAV IARTA----- -
TrzD_Acidovorax_citrulli        HARAVVNAVI ASIVGDPMVY VSGGS--EHQ GPAGGGPVAV IARTA----- -
Moorella_thermoacetica_ATCC_39  SARAVVNAVI ASVVGDPMVY VSGGA--EHQ GPDGGGPIAV IARV------ -
Clostridium_asparagiforme_DSM_  HARAVVNGVI AAIVNDPMVY VSGGA--EHQ GPDGGGPVAV IIKR------ -
AD25_Arthrobacter_sp._AD25      HARAVTGAAI ASVVGHGTVY VSGGA--EHQ GPAGGGPFAV IARA------ -
Comamonas_sp._A2                ---------- ---------- ---------- ---------- ---------- -
gi|75369013|sp|Q6PMK9|Q6PMK9_9  HARAVTGAAI ASVVGHGMVY VSGGA--EHQ GPAGGGPFAV IARA------ -
AtzD_ADP_Pseudomonas_sp._ADP    HARAVTGAAI ASVVGHGMVY VSGGA--EHQ GPAGGGPFAV IARA------ -
Pseudomonas_nitroreducens       ---------- ---------- ---------- ---------- ---------- -
USDA_110_Bradyrhizobium_japoni  HARAAVGGLI AGLAGTGAVY VSGGA--EHQ GPAGGGPVAV IARLSD---- -
viciae_3841_Rhizobium_legumino  HARAFVGGVL AGIFGITDLY VSGGA--EHQ GPPGGGPVAI IVEKEQ---- -
trifolii_WSM1325_Rhizobium_leg  HARAFVGGVL AGIFGITDLY VSGGA--EHQ GPPGGGPVAI IVEKEQ---- -
Agrobacterium_vitis_S4          HARAFVGGVL AGVFGETDLY VSGGA--EHQ GPPGGGPVAI IVEREDVI-- -
ORS278_Bradyrhizobium_sp._ORS2  HARAFVGGAL AGLFGFTDLF VSGGA--EHQ GPDGGGPVAI IVERR----- -
BTAi1_Bradyrhizobium_sp._BTAi1  HARAFVGGAL AGLFGFTDLF VSGGA--EHQ GPDGGGPIAI IVERT----- -
4_46_Methylobacterium_sp._4_46  HARGFVAGAL AGLVGATDLF VSGGA--EFQ GPDGGGPVAV IARRSGAAGP A
AZC_3892_2Azorhizobium_caulino  HARAFVAGAL AGVVGHTEIY VSGGG--EHQ GPDGGGPVAV IAARTMG--- -
Paenibacillus_sp._JDR-2         HARAVVGGVL ASVCGDPMIY VSGGA--EHQ GPAGGGPVAV VFRRN----- -
Oceanicola_granulosus_HTCC2516  HIRAALGAVV AATLGETQLF VSGGA--EHQ GPAGGGLVAV IAERT----- -
Acidithiobacillus_ferrooxidans  YSRCVVSALL AGITGETGVY VSTRA--EHQ GPQGGGPVAV IGYAS----- -
JCM_2831_Methylobacterium_radi  HARALVGGVA AAAIGRTDLF VSGGA--EHQ GPDGGGPVAV IARVRD---- -
Micromonas_sp._RCC299           HARAAVGGVV AAAVGDCRVY VSGGA--EHQ GPDGGGPVCV ISRVGGARG- -
AZC_3203_1Azorhizobium_caulino  HARAAVGAVI ASVTGDPAIF VSGGT--EHQ CAPGAAPIAA IVRA------ -
Bacillus_cellulosilyticus_DSM_  VAKAVANAVV GSYVGETRIL CSAGS--EHQ GPQGSNLVAA VIRVGGIR-- -
Frankia_sp._EuI1c               -AKAVGGALV VSITGQPMSF ISGGERNSHM GPPGGNPVAA VVRRLPA--- -
Consensus/80%                   ph+thVttlh AtlsGcs.la VSstARN.HQ GPsGGGPVAs Isc...s.GP A
```

Fig. S3
Mass spectra of cyanuric acid hydrolase reaction products using (A) unlabeled and (B) labeled cyanuric acid substrate. Peak distribution from electrospray ionization in negative mode over the approximate reaction time course for m/z of 244.0193 ± 0.058 Da and 247.0196 ± 0.038 Da for unlabeled and labeled, respectively. These values are the averaged ion mass corresponding to the carboxybiuret anion complexed with potassium and acetate ions.