

A comprehensive curation shows the dynamic evolutionary patterns of prokaryotic CRISPRs

Guoqin Mai^{1,*}, Ruiquan Ge^{1,2,*}, Guoquan Sun¹, Qinghan Meng^{1,2}, Fengfeng Zhou^{3,1,#}

¹ Shenzhen Institutes of Advanced Technology, and Key Lab for Health Informatics, Chinese Academy of Sciences,

² Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong, P.R. China, 518055.

³ College of Computer Science and Technology, Jilin University, Changchun, Jilin, P.R. China, 130012.

* These authors contribute equally to this work.

[#]To whom correspondence should be addressed. Corresponding author: Fengfeng Zhou, e-mail: FengfengZhou@gmail.com, or ffzhou@jlu.edu.cn. Web site: <http://www.healthinformatics-lab.org/ffzhou/>.

Figure S1. Structure of a CRISPR.

Structure of a CRISPR. DR is the direct repeat flanking the spacer sequence. The grey box represents the DR sequence, and DRs of the same CRISPR are homologous to each other. The patterned boxes represent the spacer sequence.



Figure S2. 6 cases with one new DR

(a) CRISPRdb only detected 9 spacers for a CRISPR (NC_010125_2181482_2182111) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 10 spacers for this CRISPR (NC_010125_2181482_2182177). (b) CRISPRdb only detected 28 spacers for a CRISPR (NC_010125_62935_64832) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 29 spacers for this CRISPR (NC_010125_62935_64899). (c) CRISPRdb only detected 21 spacers for a CRISPR (NC_010125_2253747_2255055) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 22 spacers for this CRISPR (NC_010125_2253747_2255112). (d) CRISPRdb only detected 28 spacers for a CRISPR (NC_011365_460172_461907) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 29 spacers for this CRISPR (NC_011365_460172_461964_29). (e) CRISPRdb only detected 3 spacers for a CRISPR (NC_011365_388303_388536) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 4 spacers for this CRISPR (NC_011365_388303_388602). (f) CRISPRdb only detected 96 spacers for a CRISPR (NC_008553_670839_677982) in the *Gluconacetobacter diazotrophicus* PAI 5, but we detected 97 spacers for this CRISPR (NC_008553_670839_677982). The new spacers are marked by yellow background, and the new DRs are highlighted in red.

> *Gluconacetobacter diazotrophicus* PAI5_NC_010125_2181482_2182177_10_our results

AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	CTGGCCGGTAAATTGCGTGACGGCGGGTTC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	ATCGCATGACCTTTGGTTTCGACCGGTAT
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	TCGAAGACCGCGCTGGACGACATGGGGAAG
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	CGATTCCGATACCTTGC GCGTGCGCACTGG
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GCAAGATAACGGCCTTCGGCCACACGAAGA
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GCAGATTCGATACCGGCAACGACGGTCTT
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GAGCGGAGGCAGCTTGTGGCCAATATGGC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GATCGCGCATGATGGCCGGATCAACGCTCA
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	ATCGAGCGCCGGCCGGACGTCGTTGTCGCC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	CATGGTGTGAGCTTGC TCGGCGGTTTCTC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	

> *Gluconacetobacter diazotrophicus* PAI5_NC_010125_2181482_2182111_9_gold standard

AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	CTGGCCGGTAAATTGCGTGACGGCGGGTTC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	ATCGCATGACCTTTGGTTTCGACCGGTAT
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	TCGAAGACCGCGCTGGACGACATGGGGAAG
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	CGATTCCGATACCTTGC GCGTGCGCACTGG
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GCAAGATAACGGCCTTCGGCCACACGAAGA
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GCAGATTCGATACCGGCAACGACGGTCTT
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GAGCGGAGGCAGCTTGTGGCCAATATGGC
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	GATCGCGCATGATGGCCGGATCAACGCTCA
AGCCTACCATCGGCAAAATCGGTAGGGAAACCACGGC	ATCGAGCGCCGGCCGGACGTCGTTGTCGCC

> *Gluconacetobacter diazotrophicus* PA15_NC_010125_62935_64899_29_our results

GTTTTAATCCCCGCTCCCGCTGGGGAGCGAC	GACAACTTGC GACTCCTGTCCGACGCGGTCCGGC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GGGTGCGATCCTCCGCATGGTGCAGGAGGACATC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CGCCTCGTCGATGATGACGGGCGAATTCATCACGCC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GATCACGCGCACGCCAGCCACTCGTCGTAGG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CAACTGTCCGATATCTCGGCCAATCTCTCCAACG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CGTTTCTCGATACGCGGTCCAATCCCCTCCTTCTT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CAACGTCTCCGACCCTTGACGTGGTCTCACGCCT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GAGCCGGGACGAGCGTGAATCCTGCTGCGCCG
.....	
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CTTCTGTTCCGCATCGCCGGCGGTCTGGTCCG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	AAGATCCAATGTCCTTCCAAGCGCATTCCCTAAAT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GGGAAACGAACCGTGATGTCGTGGCGAAAGAAAG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	




> *Gluconacetobacter diazotrophicus* PA15_NC_010125_62935_64832_28_gold standard

GTTTTAATCCCCGCTCCCGCTGGGGAGCGAC	GACAACTTGC GACTCCTGTCCGACGCGGTCCGGC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GGGTGCGATCCTCCGCATGGTGCAGGAGGACATC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CGCCTCGTCGATGATGACGGGCGAATTCATCACGCC
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GATCACGCGCACGCCAGCCACTCGTCGTAGG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CAACTGTCCGATATCTCGGCCAATCTCTCCAACG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CGTTTCTCGATACGCGGTCCAATCCCCTCCTTCTT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CAACGTCTCCGACCCTTGACGTGGTCTCACGCCT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	GAGCCGGGACGAGCGTGAATCCTGCTGCGCCG
.....	
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	CTTCTGTTCCGCATCGCCGGCGGTCTGGTCCG
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	AAGATCCAATGTCCTTCCAAGCGCATTCCCTAAAT
GTTTCAATCCACGCTCCCGCACAGGGAGCGAC	

> *Gluconacetobacter diazotrophicus* PA15_NC_010125_2253747_2255112_22_our results

AGATTCATCCCTGCATATGCGAGGAACAC	AAGGCGGCCTCGTCCAGGATCACCAGACCCTG
CGGTTTATCCCCGCACGTGCGGGGAACACG	TCCCCCTGCTCGGCAACCCCAAATCAAGAT
CGGTTTATCCCCGCACGTGCGGGGAACAC	GCCGCGACGGTGAAGGGGAGGTTGTTGACGC
CGGTTTATCCCCGCACGTGCGGGGAACAC	GAGTCAACGGATTCCCACTCATCCTAATATC
CGGTTTATCCCCGCACGTGCGGGGAACAC	CTTTTCTCAGCCGACACTCTCCATGCAAGAAC
CGGTTTATCCCCGCACGTGCGGGGAACAC	ACGCGGGTCTATGAGATTCCGATCATGCAGGA
CGGTTTATCCCCGCACGTGCGGGGAACAC	CATTGACCCGGCACCTCCCATGTGACGTTT
CGGTTTATCCCCGCACGTGCGGGGAACAC	TGGCTCGGACCATCTTAGCGGACGACATGGC
.....	
CGGTTTATCCCCGCACGTGCGGGGAACAC	TTGCGTCAAACGAAGCAATATCAGGGCAG
CGGTTTATCCCCGCACGTGCGGGGAACAC	TTGCCTCTAACGAGGCGCAGTATCAGGCACAG
CGGTTTATCCCCGCACGTGCGGGGAACAC	CCGTCGTTTCGATCGTACAGGGGATCATCGGC
CGGTTTATCCCCGCACGTGCGGGGA	



> *Gluconacetobacter diazotrophicus* PA15_NC_010125_2253747_2255055_21_gold standard

AGATTCATCCCTGCATATGCGAGGAACAC	AAGGCGGCCTCGTCCAGGATCACCAGACCCTG
CGGTTTATCCCCGCACGTGCGGGGAACACG	TCCCCCTGCTCGGCAACCCCAAATCAAGAT
CGGTTTATCCCCGCACGTGCGGGGAACAC	GCCGCGACGGTGAAGGGGAGGTTGTTGACGC
CGGTTTATCCCCGCACGTGCGGGGAACAC	GAGTCAACGGATTCCCACTCATCCTAATATC
CGGTTTATCCCCGCACGTGCGGGGAACAC	CTTTTCTCAGCCGACACTCTCCATGCAAGAAC
CGGTTTATCCCCGCACGTGCGGGGAACAC	ACGCGGGTCTATGAGATTCCGATCATGCAGGA
CGGTTTATCCCCGCACGTGCGGGGAACAC	CATTGACCCGGCACCTCCCATGTGACGTTT
CGGTTTATCCCCGCACGTGCGGGGAACAC	TGGCTCGGACCATCTTAGCGGACGACATGGC
.....	
CGGTTTATCCCCGCACGTGCGGGGAACAC	TTGCGTCAAACGAAGCAATATCAGGGCAG
CGGTTTATCCCCGCACGTGCGGGGAACAC	TTGCCTCTAACGAGGCGCAGTATCAGGCACAG
CGGTTTATCCCCGCACGTGCGGGGAACAC	

> *Gluconacetobacter diazotrophicus* PA15_NC_011365_460172_461964_29_our results

AGATTCATCCCTGCATATGCGAGGA	ACACAAGGCGGCCTCGTCCAGGATCACCAGACCCTG
CGGTTTCATCCCCGCACGTGCGGGAA	CACGTCCCCTGCTCGGCAACCCAACTCAAGAT
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCCGCGACGGTGAAGGGGAGGTGTTCGACGC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCCTGATCCTGCGCCGACGCCGAACCAGTCCG
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCCGCGACGGTGAAGGGGAGGTGTTCGACGC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCCTGATCCTGCGCCGACGCCGAACCAGTCCG
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGAGTACCCGATTCCCATCTCATCCTAAATATC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCTTTTCTCAGCCGACACTCTCCATGCAAGAAC
.....	
CGGTTTCATCCCCGCACGTGCGGGAA	ACACAACGTGGAATGGCACCAGATCGAAACGATGCT
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCGAAGACCTGCGCCTGGAGCGCCTGCTTGTG
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCGTGGGAATATCCTGACACACATGGGCGT
CGGTTTCATCCCCGCACGTGCGGGAA	

> *Gluconacetobacter diazotrophicus* PA15_NC_011365_460172_461907_28_gold standard

AGATTCATCCCTGCATATGCGAGGA	ACACAAGGCGGCCTCGTCCAGGATCACCAGACCCTG
CGGTTTCATCCCCGCACGTGCGGGAA	CACGTCCCCTGCTCGGCAACCCAACTCAAGAT
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCCGCGACGGTGAAGGGGAGGTGTTCGACGC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCCTGATCCTGCGCCGACGCCGAACCAGTCCG
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCCGCGACGGTGAAGGGGAGGTGTTCGACGC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCCTGATCCTGCGCCGACGCCGAACCAGTCCG
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGAGTACCCGATTCCCATCTCATCCTAAATATC
CGGTTTCATCCCCGCACGTGCGGGAA	ACACCTTTTCTCAGCCGACACTCTCCATGCAAGAAC
.....	
CGGTTTCATCCCCGCACGTGCGGGAA	ACACAACGTGGAATGGCACCAGATCGAAACGATGCT
CGGTTTCATCCCCGCACGTGCGGGAA	ACACGCGAAGACCTGCGCCTGGAGCGCCTGCTTGTG
CGGTTTCATCCCCGCACGTGCGGGAA	

> *Gluconacetobacter diazotrophicus* PA15_NC_011365_388303_388602_4_our results

AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	TCAGCCGGCTGGGCATTGCCCTACGAGAGG
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	AGGGGAGCGGTACCTTCGACAAGGGCCT
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	CCTCGTTGCTCGCGCGCGGGTGCAGGGCTG
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	TCCGCACGFTCACGCGTACCAGCATCGCGG
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	

> *Gluconacetobacter diazotrophicus* PA15_NC_011365_388303_388536_3_gold standard

AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	TCAGCCGGCTGGGCATTGCCCTACGAGAGG
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	AGGGGAGCGGTACCTTCGACAAGGGCCT
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	CCTCGTTGCTCGCGCGCGGGTGCAGGGCTG
AGCCTACCATCGGCAAAATCGGTAGGGAAACACGGC	

```

GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC TTTCCGGATTAGAATAATAAGGAACTCCGGTACGGG
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC GGTATACTAAGCGTACCGTTCGTCTCAACATCGACTA
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC ACACCATATGCTCCGGGCAGAGGAGCATGTTCAACT
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC CCAGTCTCCTCCTCAGGAGGACTGGCTTTTCAA
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC CACAATAAATCCCCCTTGA AAAACGAGTTCAAGC
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC TATGTAGCTGCATATCTACACCTCCACGGATATAAGAG
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC TCTTCTGGCTCTGGCACGGCGAGA ACTATCTCGCCGT
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC TCCAAATTACCAAGACAAACATGCTTGT CATGGGTG
.....
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC GCCCATATCTCTTCGTGATGCTCCGGGTCTGCTG
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC AGCAGCATCTTTTTCACCTCATGATATCATTATGAC
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC TTTCTATACTGGGGATGTGTACTTCTTGCCTTCCTT
GTCGAAGAGCGAGTTCCAGGAAAACAAGGATTGAAAC

```

> *Methanosaeta thermophila* PT_ NC_008553_670839_677982_97_our results

> *Methanosaeta thermophila* PT_ NC_008553_670839_677982_96_gold standard

Figure S3. 4 cases with two new DRs

(a) CRISPRdb only detected 26 spacers for a CRISPR (NC_014152_2078344_2080300) in the *Thermincola sp.* JR, but we propose two more DRs for this CRISPR, , although one of the two new DRs is identical to the other DRs in half of its region. These two new spacers comes from their BLAST matches to two known spacers in the other genomes in the SpacerDB with 91.3% and 94.4% in matching identity percentages, respectively. The CRISPRs, *i.e.* NC_015865_1907425_1908328 in *Thermococcus sp.* 4557 (b), NC_015738_2085666_2087297 in *Eggerthella sp.* YY7918 (c), and NC_014209_791663_793738 in *Thermoanaerobacter mathranii subsp. mathranii str. A3* (d), are also expanded with two more DRs. The new spacers are marked by yellow background, and the new DRs are highlighted in red.

> *Thermincola sp.* JR_NC_014152_2078201_2080300_28_our results

```

TTTCGTCCTCCCTACGGGGATAAGGAAGATTA AAC TCTTTGGCAAAGGCAAAACATCAGTTGACAGTTGACA ← 146bp
TTAACAGTTAACAGTTGAGATAAGGAAGATCA AAC TCTTTGGCAAAGGCAAAATATATCATTTCGGCAACCCAAA
TTTCGTCCTCCCTACGGGGATAAGGAAGATTA AAC GCAAAACCCCGGAACAGGCCCGGGGATGGGACC
GTTTCGTCCTCCCTACGGGGATAAGGAAGATTA AAC TAAATCCGAATGGATACTATTCTTGTCCAGATGCTG
.....
GTTTCGTCCTCCCTACGGGGATAAGGAAGATTA AAC ATGAATTAAGGATGGAATAGTTAATTCGACACAGT
GTTTCGTCCTCCCTACGGGGATAAGGAAGATTA AAC

```

aag gcaaacatcagttgacagtt
|||||
aag gcaaacatcaattg-cagtt 91.3%_NC_003106_6_11_spacer

gcaaaaggcaaaatatca
|||||
gcaaaaggcaaa-atatca 94.4%_NC_015555_4_102_spacer

Thermincola sp. JR_NC_014152_2078344_2080300_26_gold standard

> *Thermococcus sp.* 4557_NC_015865_1907425_1908328_13_our results

```

GTTGCAATAAGACTCTGGGAGAATTGAAAG TCGGGAACGCCAAGTTTATTACGGCGCCGATTACCT
GTTGCAATAAGACTCTGGGAGAATTGAAAG CCGTTGGCTCATCGGTAGAAACGGGCTCATCTATAATC
.....
GTTGCAATAAGACTCTGGGAGAATTGAAAG CAGGCGCTCAGGGCCGGGACATACGGGTGCGCCG
GTTGCAACAAGACTCTCTAGGAGAATTGAAA TGGTGACGATGATGGCCCTGCTCCCTGCTCCACTTGG ← 136bp
GTTGCAATAAGACCCCAAGGAGATCGAAAG

```

Thermococcus sp. 4557_NC_015865_1907425_1908192_11_gold standard

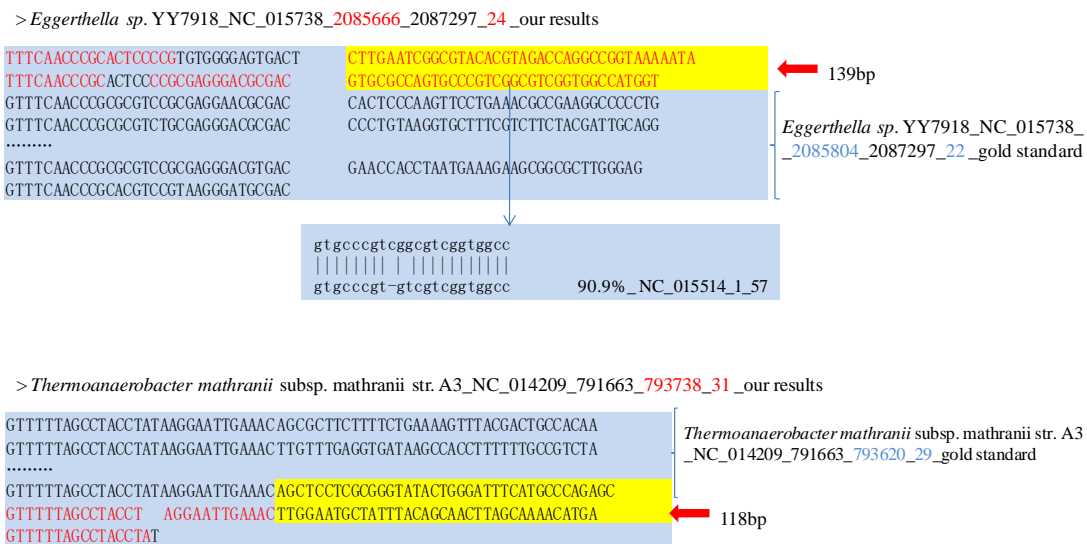
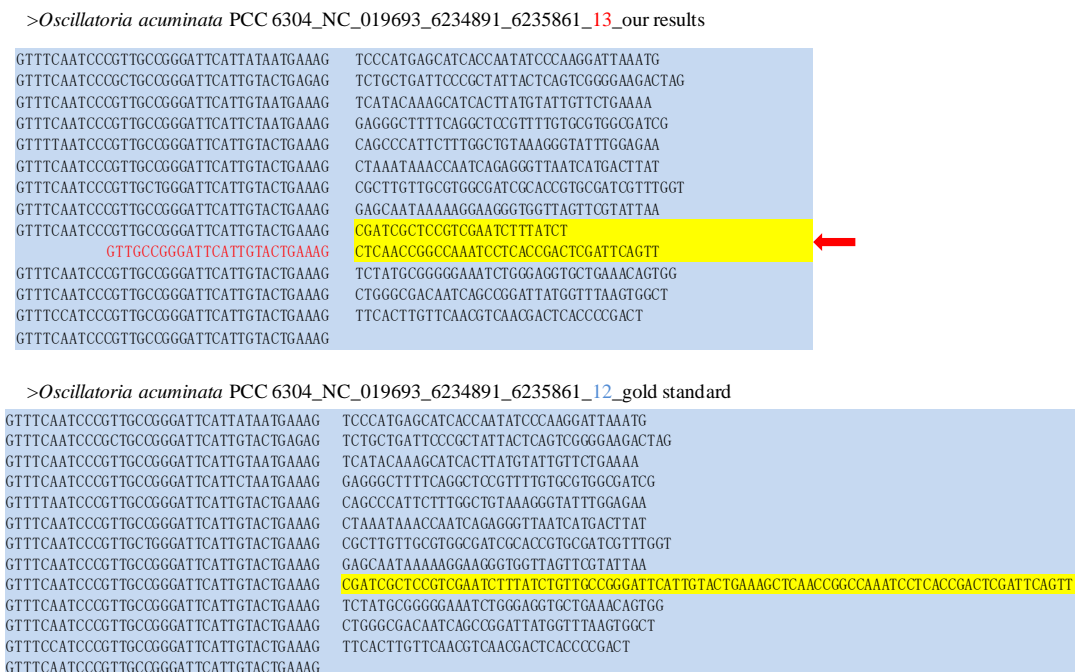


Figure S4. 3 cases of long spacers with a partial DR

- (a) A CRISPR in *Oscillatoria acuminata* PCC 6304 has a long spacer with a truncated DR inside.
- (b) A CRISPR in *Chlorobium phaeobacteroides* DSM 266 has a long spacer with a truncated DR inside.
- (c) A CRISPR in *Frankia* sp. Cc13 has a long spacer with a truncated DR inside. The long spacer is in yellow and the truncated DR is in red. DRs and spacers are represented in the left and right columns, respectively.



>Chlorobium phaeobacteroides DSM 266_NC_008639_1625359_1633049_115_our results

GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGGCTATCCGTATCTGGCTTTACGCTGTGAAGCT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CTGTCAATGATGGCCGATGATTCCCGTAAATAG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AAGGCCAAATOGACCACATCGACGGCGCGGTCT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	GTGATGTGCGTTCCGGCAGCAGCGCTCAAAAA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CTGCATCCACGGCAGTTCAGGGGAGGATTCCGTTAT
GTTTCAATCCACGGCTCCCGCAGGGGGCGGAC	GGTGTCACTTACGGTGTGTTTTTGTATCGGAATCGA
TTCATCCACGGCCCGCAGGGGGCGGTAC	CATTGAATACTACAAGATGGAACCTCAACTACTC
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CATGCTTGCATTACCTGATCGAAGCCACCCAA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AAAAAGTGATTGTGATCACACACCTGCCAAGCAT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	TCGGATTGACTCTTGCCTAAGTGGGTGATCGGCA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGTTGTCCAGGCCCTGCACGACCTCAATCAACGG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGGCTATCCGTATCTGGCTTTACGCTGTGAAGCT
.....	
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	GCATGGTCGAGAAAGAGTTGCCTGAATCGGCACCTGT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CAATGCTTCTGTTAAGTTGAGCAAGCGCAATCACTG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CACAGCCCGCCTTATCAGCAGGCGTTGTCGAGGC
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	

>Chlorobium phaeobacteroides DSM 266_NC_008639_1625359_1633049_114_gold standard

GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGGCTATCCGTATCTGGCTTTACGCTGTGAAGCT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CTGTCAATGATGGCCGATGATTCCCGTAAATAG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AAGGCCAAATOGACCACATCGACGGCGCGGTCT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	GTGATGTGCGTTCCGGCAGCAGCGCTCAAAAA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CTGCATCCACGGCAGTTCAGGGGAGGATTCCGTTAT
GTTTCAATCCACGGCTCCCGCAGGGGGCGGAC	GGTGTCACTTACGGTGTGTTTTTGTATCGGAATCGA
GTTTCAATCCACGGCCCGCAGGGGGCGGTAC	CATTGAATACTACAAGATGGAACCTCAACTACTC
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CATGCTTGCATTACCTGATCGAAGCCACCCAA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AAAAAGTGATTGTGATCACACACCTGCCAAGCAT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	TCGGATTGACTCTTGCCTAAGTGGGTGATCGGCA
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGTTGTCCAGGCCCTGCACGACCTCAATCAACGG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CTCGTATCAAGGATGACAGGTGACGCTAGACATC
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	AGGCTATCCGTATCTGGCTTTACGCTGTGAAGCT
.....	
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	GCATGGTCGAGAAAGAGTTGCCTGAATCGGCACCTGT
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CAATGCTTCTGTTAAGTTGAGCAAGCGCAATCACTG
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	CACAGCCCGCCTTATCAGCAGGCGTTGTCGAGGC
GTTTCAATCCACGGCCCGCAGGGGGCGGAC	

>Frankia sp. CcI3_NC_007777_3904715_3905896_16_our results

CTTGGGAACCTCGCCGGCGGGCGCCCGGGCTGCT	CTTCTGGGTCGGATATGGTGCCCTTATCGTGACTC
GTTGTGATCTCGCCGCTGCCCTGGGGCAGATCAAGC	GCCCAAGTTCGGCATCTTCGC
GTTGCGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TGGTGGGGAGGTGGGCTGCGCCAGCGTCAGTT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GACGACCTGGTGCTGATCTGCACCAACGGCACCGCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CGTCCGCGGACAGAAAGACCTACACGCCACCACC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GTGTCCGCTCCCGTAGCTGCTCGTCCGCTGGTT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CTGGGACACGGCCGGATCCTGCCAGGGCACCGCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GATGGTGACGACCTGGTGCTGATCTGCCAACG
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TCGATGGTGGCTGGGCTGGTCAAGCCTTGGAT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	AAGTGCTCGAACATGTGTGACTGCATCCGGGT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CCCCGGCCGGGCTCACCGCTGCCCGCTGGTCTAT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TCGGAAACACCAAGCTGTGGAGCCACGGGGTA
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TGCGCCCGGCTATCTGGTCCGCTTTCTCGGGA
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GACCGCTGCTGAGCCGACGGTCAACAATGACCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CCGCTCCGGCCTGGATGCCAGATGCCGCCCTTGG
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CTGTAGGGGACGTTGTAATGCTTCCGGCCGGAG

>Frankia sp. CcI3_NC_007777_3904716_3905896_15_gold standard

CTTGGGAACCTCGCCGGCGGGCGCCCGGGCTGCT	TCTTGGGTCGGATATGGTGCCCTTATCGTGACTC
GTTGCGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TGGTGGGGAGGTGGGCTGCGCCAGCGTCAGTT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GACGACCTGGTGCTGATCTGCACCAACGGCACCGCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CGTCCGCGGACAGCAAGACCTACACGCCACCACC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GTGTCCGCTCCCGTAGCTGCTCGTCCGCTGGTT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CTGGGACACGGCCGGATCCTGCCAGGGCAACGGCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GATGGTGACGACCTGGTGCTGATCTGCCAACG
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TCGATGGTGGCTGGGCTGGTCAAGCCTTGGAT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	AAGTGCTCGAACATGTGTGACTGCATCCGGGT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CCCCGGCCGGGCTCACCGCTGCCCGCTGGTCTAT
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	TCGGAAACACCAAGCTGGTGGAGCCACGGGGTA
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GACCGCCCGGCTATCTGGTCCGCTTTCTCGGGA
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	GCGCGCTGCTGAGCCGACGGTCAACAATGACCC
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CCGCTCCGGCCTGGATGCCAGATGCCGCCCTTGG
GTTGTGATCCTCGCCGAGGGCGGTCCCTGGGGCTGC	CTGTAGGGGACGTTGTAATGCTTCCGGCCGGAG

Figure S5. 59 cases with DRs broken by a transposon

(a) CRISPRs with DRs inserted by transposons in the *Microcystis aeruginosa* NIES-843. DRs and spacers are represented in the left and right columns, respectively. The newly annotated CRISPR regions are in red, and the new spacers are marked by yellow background. The last number in the CRISPR name is the DR copy number. (b)~(k) The other CRISPRs with DRs inserted by transposons.

GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATT	TTCAGAATTCGACACAATTATTGCTATAAAAAATGATAGC TTGTAATTGAGCTTCGTGTACGCCAATTTAATC MITE(180bp)	<i>Microcystis aeruginosa</i> NIES-843_NC_010296_2814769_2822874_112_our results
CTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGA	TCCTGCTTGGTAAATTAAGTACGCCATAAGC CAGCCCGAATGGTACGAGACAGCCTGGCAGACCAA transposase (513bp)-transposase (675bp)-MITE(167bp)	<i>Microcystis aeruginosa</i> NIES-843_NC_010296_2823052_2824590_21_our results <i>Microcystis aeruginosa</i> NIES-843_NC_010296_2823103_2824595_20_gold standard
TTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAAAC GTTCCAATTAATCTTAAACCTATTAGGGATTGAA	AGCAATATTTTGGCATCGGATTCGGTCAAACT CTTAAGCAGTGCCTTTGAGCGATAAGGCGATTATC transposase (513bp)-transposase (675bp)-MITE(167bp)	<i>Microcystis aeruginosa</i> NIES-843_NC_010296_2826180_2829164_41_our results <i>Microcystis aeruginosa</i> NIES-843_NC_010296_2826228_2829166_40_gold standard

GGGACTTAGGCATTTTCGGGCAGCAAAAAAGGCTCAAACCATTACGGGACAAGGGTTAACCTCGATTATGATTTTCCTTGATCTATCTGGGTTTCAGCGATTTTG
GGCTCTCGAAGTAAATCCCAAGCAGGGATTGG GGTGAGGCTTTTCTCGATTTGTTTTGTCGTAAGTCCCA (180bp, appears 19 times)

TTGTTACTCGAAAAAGCTCGAAGTATAAGAGTTTGTGCTGCCTATTTTCGAAAAATAGGTTGTAATGGGAGACTCTTTGTTTCCTAGACAAAAATAATTACT
TTTCAAAAAACACTTTTCTATTTTTTTGTTGGGTA TTTTATTCTCTGGAAGTCCCT(167bp, appears 8 times)

CACTATTTTCAGGATAGGTAGGCTAAAA GTTTCAATTCCTCATTAGGTAGGCTAAAAAC GTTTCAATTCCTCATTAGGTAGGCTAAAAAC GTTTCAATTCCTCATTAG	Transposase (1080bp)- Transposase (426bp) AAACGAGGTAGGATCAAGCGGGTGTATCTGCGCTGATAA CCCAATCCGAAATAGAGGTAGGAGAAATGCAGGAGT CATCATTGAGCGGATTTTGGATCTCGGTTTATTTA Transposase (1221bp)	<i>Thermoanaerobacter italicus</i> Ab9_NC_013921_2330986_2331339_5_our results <i>Thermoanaerobacter italicus</i> Ab9_NC_013921_2330986_2331286_4_gold standard
ATAGGTAGGCTAAAAAC GTTTCAATTCCTCATTAGGTAGGCTAAAAAC GTTTCAATTCCTCATTAGGTAGGCTAAAAAC GTTTCAATTCCTCATTAGGTAGGCTAAAAAC	CCCGCTTGACATGATATATTTTCGGCGAGGCGCGGT GAAGATGCAATAGAGTAGCAGATAAAATGCTGAA TCTTTCCTATTCTGTTTAACTACTATAGAATAAT	<i>Thermoanaerobacter italicus</i> Ab9_NC_013921_2332916_2334204_19_our results <i>Thermoanaerobacter italicus</i> Ab9_NC_013921_2332916_2334204_19_gold standard

GTCAGAAGGACTTCCTGATGAAGAAGGATTAAAGC GTCAGAAGGACTTCCTGATGAAGAAGGATTAAAGC GTCAGAAGGACTTCCTGATGAAGAAGGATTAAAGC GATGAAGAAGGATTAAAGC	GCCGCGCGCTCTTGTGCGCGCATGTACTG GGCTGGCCCGCTCCATGGCATCCATGTCACGGC IS66 (927bp)-phage DNA methylase (641bp) CTTCATCAGAAATGGCGCTCGTGCATCGGTGGG	<i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_1466441_1467521_15_our results <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_1466441_1467533_15_gold standard <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_1469911_1471052_16_our results <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_1469966_1471052_15_gold standard
TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC CGCGGCAGCGAAC	CAGGGATAGGTGATGGTTCTATTACGGTTATG ATCATCGAGAAGCCGCAAGTGGCAGCCA ISPsy4 (222bp)- istB(171bp) CCCTCAGATATCGCGCGGCGTGTGATCTGGAT	<i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2520192_2523887_61_our results <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2520192_2523887_61_gold standard <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2928156_2928605_7_our results <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2928217_2928605_6_gold standard
TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC TTTCTGAGCTGCCTGCGCGGCAGCGAAC	CTGGCCCAAGCCGACGACCCCGGCTGCTACGA IS110(1383bp) GGCGGCAGTCCGGTGGTCCGATTTTGGATGA GCTGGCGGGAACCTCATGAGCATGGCTGGCGA GGGGCGATCAGTCCGGGGTTCGAGGGATCGA	<i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2931057_2932046_16_our results <i>Thioalkalivibrio nitratireducens</i> DSM 14787_NC_019902_2931057_2932046_16_gold standard

GTTTCAATTCCTCATAGGTACGCTGAGAAC	GCATTGCCTTTCATCTCCCTCGGCATGCCCGC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2439391_2440133_11_our results
GTTTCAATTCCTCATAGGTACGCTGATAAC GTTTCAATTCCTCATAG	GCTCTGGATGCTTGGAGAAAATGCAAGGAACGGAATA IS110 (2800bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2439391_2440146_11_gold standard
AGGTACGCTGAGAAC GTTTCAATTCCTCATAGGTACGCTGAGAAC	ACATAATATTCCTTGTATGCATAAACCACGCCGTCAGGA TCTGCACATTGGGGCATAAAGCGACGCCCCAAAAACC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2443573_2445992_36_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATA	TTCTCGGCTGGCAGGCTGATTTAGAGGAGGTGTTCTAA IS1634(1701bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2443641_2446006_35_gold standard
TCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	CGGTAGATAGTCCCGACGAGGCTCTT CGATAATGTTCAAACTTGCATTCAATTCGTTCT	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2447892_2448160_4_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAG	GAAATATACCATGCGGGGTGTTTCAGTCCGGTTCA IS110 (1215bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2447944_2448173_3_gold standard
AGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	TAAAACCTGATTCCTCCAGTCTTTTCTTTGGGCA ACGGAACCACTCCTGCCAATATGCTCTTTTCGTTTC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2449719_2451060_20_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	CCCGGACAATACCCATCATGATACCGTCTCGGTTCTTC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2449704_2451060_20_gold standard
GTTTCAATTCCTCATAGGTACGCTGAAAAAC	TTTGACCACCCTCATCTTAGCAAATGCTAACGGA	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2453564_2455390_27_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAG	ACATTTTCAATTTTATGTAGATTGAAATATCGTTCG MITE (300bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2453564_2455337_26_gold standard
AGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	CGGTGTGTGATGTATGAAAAGATGACGATGACCAGGGC ATGTGCCGAGGTGTGCCGAATGCGCAGATATCC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2455691_2458143_37_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATA	GCATGGACTGTGAGCGGATGTACGTCTCTGGGAGTA IS1634(1701bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2455676_2458157_37_gold standard
TCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	GCCGAGAAAATGTTCTCGGTGCGCTGGGATTACGCT GATGCCCTCCTTCTTGCGATTGAACGGCAAGAAGTG	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2460043_2463456_51_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAG	CGGACCCTTCAACCCAGGAGGTTTGTGACCTG MITE (300bp)	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2460106_2463402_49_gold standard
AGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	GACTGGCAGAGATTCGGGATGTCGTGAGGCAGTATC	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2463757_2464503_11_our results
GTTTCAATTCCTCATAGGTACGCTGAAAAAC GTTTCAATTCCTCATAGGTACGCTGAAAAAC	GAGGGAGAATGCCGATGGTCGGTCGAGTCCGCGG	<i>Thermacetogenium phaeum</i> DSM 12270_NC_018870_2463742_2464503_11_gold standard

GTCGTAATCCCTTCAA AAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	MITE(49bp) TGAAGATTAAAGACCGAAGCAATCCAGATTTTGAAGA CATAGGCTGATTAAGGGCATGTAAGCATGGAGCTTGT GGAAGGGCAACAACCCTTCCGGAGGTTTTAAAATGA	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1428433_1430685_30_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1428495_1430685_29_gold standard
GTCGTAATCCCTTCAA AAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	IS4 (1368bp) MITE(49bp) AGTATATCTTAACTTAAACAATGGATTAGAAGACTTTT TCGGGTGCTGCGTATTATGCAACTGCTCCACGGAAAGG	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1709402_1709566_2_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1709462_1709573_1_gold standard
GTTGTAATCCCTTCAA AAATCAGGTCATGGATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	MITE(47bp) GGTATAGATAATAAAAAATGTCATGATTTGATATA AATAAATACTCAAGAACCTTTTGAATGAGATAGAA AAATTCGGACTGAGCAATTAGGATGGCAAAAAA	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1709819_1710226_5_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1709819_1710226_5_gold standard
GTTGTAATCCCTTCAA AAATCAGGTCATGGATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	MITE(47bp) GTTGAAGCACTCGGCGCAGTTCTATCTTTGAGCAGA AGAGAGATTTCTCCATTTCCATAATTAATACAGTTGAT TACACAATAAAACAGATACTTGTATTGCTTTTCTTA	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1710472_1710885_5_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1710472_1710885_5_gold standard
GTTGTAATCCCTTCAA AAATCAGGTCATGGATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	MITE(47bp) CTGAAGCGAATCAATTATATAAAAAATCTTTCTCTTCT AGTGATCGGGGTGGTACGTCGTTGATATCGAAA GCCGGAAGGAGCACAACTTCGGAGGTTACAAAA	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1711131_1712652_20_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1711131_1712652_20_gold standard
GTTGTAATCCCTTCAA AAATCAGGTCATGAATCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT GCAATCCCTTCAAATCAGGTCATGAATCCAAT GTCGTAATCCCTTCAAATCAGGTCATTAATTCCAAT	MITE(47bp) TGGAAATATATAGATACCTGTCTTCCGGGTTGA TTAAAAAGAAATCCAAAGAAAATAGAGAACTCCCAAG ATCTTCGTCCTCGTGGGATATCCCGAGGACTGAAAAA	<i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1734298_1735756_19_ourresults <i>Flexistipes sinusarabici</i> DSM 4947_NC_015672_1734298_1735756_19_gold standard

ACATTGCGCAATATCTCATCGCAAATGTTTCTGTTCAGAGCTTGCTGG (49bp, appears 3 times)

ACCCAGCCATCATATAATCGGCTGGGTTCTGTGTCGAGCTTGCTGG (47bp, appears 4 times)

GTTTCAATTACCGTAGGTACTACAGAAC TTTCAAAGCCTTAAAGGTACTATCAGAAC	TCATATAGGTGCATTCCTTCATTCGCATAATCGCAAA TTGGACAAGCTAAAGCTAGGGCTTTACGGCGCATCCTA	<i>Thermobacillus compostii</i> KWC4_NC_019897_3270701_3271073_5_our results
GTTTCAATTCCTTATAGGTACTACAGAAC TTCAATTCCTTATAGGTACTATCAGAAC GTTTCAATTCCTCATAGGTACGATCAAAAC	CGATACGGGCGATGTCATAGTATACTTCGCCAGATGAGA IS110(1105bp) TTCACACTTCGGCATGTCGGCCCTCGGACTTACGAC	<i>Thermobacillus compostii</i> KWC4_NC_019897_3270768_3271074_4_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTACGATCAAAAC GTTTCAATTCCTCATAGGTACGATCAAAAC	TATAGTGCTCGGGTCAAGGACCGGAAATATACCGG IS110(1227bp) GGAGCCTAAACCATGCGTCAATATCGGGACGAAT GAGCAGCCGGGCGCCAAATCCGATGCACAGACGAG	<i>Thermobacillus compostii</i> KWC4_NC_019897_3272406_3273774_20_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3272406_3273720_19_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTACGATCAAAAC GTTTCAATTCCTCATAGGTACGATCAAAAC	ACTCTGCGAGCGCTCGAGGCGCCAGCCGACTTT IS110(1227bp) CATCAGCGCGCCGACACCGGCCACGTCGACACCG GCAAACGGAAACCGCTGCGCGCCGCCCGTGTTTGT	<i>Thermobacillus compostii</i> KWC4_NC_019897_3275324_3277594_34_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3275311_3277541_33_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTACGATCAAAAC GTTTCAATTCCTCATAGGTACGATCAAAAC	CGGGACGATCTATGTCATGTTCTCCGGCGTGAAGT IS110(1227bp) AAAGGAATGGGACAGACATCGTCTTCCTGAAATTA CATGTCGAGCTCCACCATGAGAGCCGATCCGCCGCGC	<i>Thermobacillus compostii</i> KWC4_NC_019897_3279144_3280208_16_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3279131_3280155_15_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTACGATCAAAAC GTTTCAATTCCTCATAGGTACGATCAAAAC	CTGTATCGTATATTCAGTTCCTGATCCTCGGAAGCAIT IS110(1228bp) GTTACGGGTGTGCTCCACTAGCCGCTCCACGCTTC GTGATACTGTCCGAATACTCGTACCGTCATATCCGT	<i>Thermobacillus compostii</i> KWC4_NC_019897_3281758_3283032_19_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3281745_3282977_18_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC	GTGAGCCATCAGAAATATCTCGTTCCTATCCGT IS(1293bp)-IS(1278bp)-IS110(1224bp) GATCGAAAAGCGAGAAACGATTGAGTTGAACGTTGACGA GAAGGAGTTTACCCTCGCCGTCATACCGCTGGAGT	<i>Thermobacillus compostii</i> KWC4_NC_019897_3284583_3285001_6_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3284570_3285001_6_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC	TGAGGTGGTGTGAATTATAGAAGCTGAAAACAGG IS110(1227bp) AAACCTACATCCGATGGACTACACTGGTACAGATCGCTAC GATGCCCTGAAGTTCCAGACGGCTGATACGCCA	<i>Thermobacillus compostii</i> KWC4_NC_019897_3298570_3299110_8_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3298557_3299057_7_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTAAGATCAA GTTTCAATTCCTCATAGGTACGATCAAAAC	TCAGTTTGGCCATGTATGTGATCTTCGGCATTTCGT IS110(1225bp) ACCGATGCCGAGAACAATGGAGTAAGAATCCGTATG TTTGATGTACATGGCAGCGCGGAACAGAGATCG GGCCACGATCTCACCCACGCTGTCGTACCGGTGT	<i>Thermobacillus compostii</i> KWC4_NC_019897_3300646_3300996_5_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3300647_3300943_4_gold standard
GTTTCAATTCCTCATAGGTACGATCAAAAC GTTTCAATTCCTCATAG ATAGGTACGATCAAAAC GTTTCAATTCCTCATAGGTACGATCAAAAC		<i>Thermobacillus compostii</i> KWC4_NC_019897_3302548_3302761_3_our results <i>Thermobacillus compostii</i> KWC4_NC_019897_3302535_3302761_3_gold standard

GTTTTAGCCTACCTATAAGGAATTGAAAC	AGCTGAAAAGACATTTTAAAGTAGCAGGGCAAGGTACAG IS110(1221bp)	<i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_121660_124897_48_our results <i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_121660_124897_48_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	AGTATTAATAATTAAGGAGTAAATAATGAAAAA CTAATTTGCTTTAAGCTTGTTCACAAAGATAGACCTTG CGCCGAGGTCCTCTCTGTCGATATAATGCGACAGG	<i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_126473_128016_23_our results <i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_126529_128029_22_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	AGAGGAACAATGGAATGATGAAATAGGACAAAAGA AGGTCAAAACCTGGGGCAAGACTTACGACGGGAAA TGCCATCACTATCATATCTCCATCCACTCTTA	<i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_131265_133542_34_our results <i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_131317_133555_33_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	TGGATAGCAAAGGCAAGTGCAGAACTGGAATAAAAAA IS110(1221bp) CCTATCAGCTCCGGTGGCAGCCAAAGCAATAGC GATAGCCTTCAGTATTAACCTTGCCGCTTCCAAATG	<i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_135132_136869_26_our results <i>Thermoanaerobacter brockii</i> subsp. finnijii Ako-1_NC_014964_135184_136882_25_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT	CCATGTGTGTTTTACGCTCAATCTCGTTACGTAGAAA IS110(1221bp)	

GTTTTAGCCTACCTATAAGGAATTGAAAC	AGCTGAAAGACATTTTAAAGTAGCAGGGCAAGGTACAG	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_121655_124879_48_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	AGTATTAATAAATAAGGAGGTTTAAAAATGGAAAA IS110 (1221bp) CTAATTGGCTTTAAGCTTGTGTGCTACAAGATAGACCTTG	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_121655_124892_48_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC	CGCCGAGGTCTTCTCTGTCGGATATAATGCGACAGG	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_126469_128012_23_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	AGAGGAACAATTGGAATGTATGAAATAGGACAAAGA IS110 (1221bp) - IS110 (1221bp)	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_126525_128025_22_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC	AGGTCAAAACCTGGGGCAAAGACTTACGACGGGAAA TGCCATCAACCTATCATACTATCTCCATCCACTCTTA	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_131261_134422_47_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTATAAGGAATTGAAAC CTATAAGGAATTGAAAC	CACTTACGGCAGTCCGGTAGCTCCTGATCATGTTCGG MITE (309bp) GTTTTTGTACATTTCACATCAAGGATTTTTCTAT	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_131313_134493_47_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC	TGCCGTCTGTGCTTGCAGTATTTCTAGTCGCGGA	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_134731_135597_13_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	CCGCCACATACATCCAAACAGGAAGTACCAGCATGG IS110 (1221bp) AGAAGAATATTGACAGTTTAACTGCTCAAAGATAGC	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_134784_135610_12_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC	ACTGTCCCTCAATTTCCCTGTTTTCTCCCTTAAAGCTT	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_137187_138123_14_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT CTATAAGGAATTGAAAC	TGGATAGCAAAGGCAAGTGCAGAACTGGATAAAAGA IS110 (1221bp) CCTATCAGCTCCGGTGGCACCCAAACGCATAGC	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_137241_138136_13_gold standard
GTTTTAGCCTACCTATAAGGAATTGAAAC	GATAGCCTTCAGCTATTAACTTTCGCGCTTCAATTTG	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_139713_141450_26_our results
GTTTTAGCCTACCTATAAGGAATTGAAAC GTTTTAGCCTACCTAT	CCATGTTGGTTTTACGCTCAATCTCGGTTACGTAGAAA IS110 (1221bp)	<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223_ _NC_010321_139765_141463_25_gold standard

GTCTTATCTGAACTATGAGGGATGTAAC	ATAGAGGTACCAGGACAAGCTTGGCTCGAAGGAGAAG	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _344075_347921_58_our results
GTCTTATCTGAACTATGAGGGATGTAAC GTCTTATCTGAACTATGAGGGATGTAAC GTCTTATCTGAACTATGAGGGATGTAAC	CTTTTCAGTTCAAATAATATCTCGTTCGCAAGGAAA ATPase (792bp)-IS (1230bp)-IS (1071bp)	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _344075_347921_58_gold standard
GTCTTATCTGAACTATGAGGGATGTAAC	CITTTTCTTTTTCAAAGACCTTGCACAAAAAATC	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _351757_353150_21_our results
GTCTTATCTGAACTATGAGGGATGTAAC GTCTTATCTGAACTATGAG	CATAGAAAGCGGACTTTTTATTACTGGAAGGATGAAG IS (909bp)-IS (1230bp)-IS (438bp)	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _351757_353160_21_gold standard
GTCTTATCTGAACTATGAGGGATGTAAC	ACAAAAAATGCAAGGTCTTTGGAACAAATATTGATGA	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _356437_357973_23_our results
GTCTTATCTGAACTATGAGGGATGTAAC	TCAACTTTTCGGTTCTCTTTCAGCTTGTTTGACTGC	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _356504_357983_22_gold standard
GTCTTATCTGAACTATGAGGGATGTAAC GTCTTATCTGAACTATGAG	ATTTCCAATGTTGCACATATAGGCAAAATATCCGCTGCTA IS (1248bp)-IS (486bp)	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _539664_541402_26_our results
GTCTTATCTGAACTATGAGGGATGTAAC	IS (987bp)-IS (1230bp)-IS (321bp) ACATGTTTGGGCTCCCGATTTAGAGCCGATTGAGCCG	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _539731_541402_25_gold standard
GTCTTATCTGAACTATGAGGGATGTAAC	TTGCATAACTAGTGGATCAATTAATCGGCTCTATAATA	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _2513322_2520034_102_our results
GTCTTATCTGAACTATGAGGGATGTAAC GTTTTATCTAACTATGAGGGATGTAAC	AAGGGGGTGTGGAAAACTACTATTGCACCTGCAC IS (1131bp) - IS (1221bp)	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _2513322_2519967_101_gold standard
GTTTACATCCCTCATAGTTCAGATAAGAC	TAAAGTTATTAATCCTCTTTCAGCCACTTGTGCTA AGCTTTTTGGAACGACTAATCCAGATAGCCCTTATCA IS110 (1287bp)	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _2522005_2524601_39_our results
GTTTACATCCCTCATAGTTCAGATAAGAC	TATCAACATTGAACCAAAGCTGTTATGTTGGTCATCA TAACGCAGACACGGCAACAAAAGAAAAGCGTTGATAA	<i>Caldicellulosinuptor saccharolyticus</i> DSM 8903_ _NC_009437_ _2521995_2524601_39_gold standard
GTTTACATCCCTCATAGTTCAGATAAGAC GTTTACATCCTTATAGTTCAGATAAGAC	TGATGATTATTACCAACAACATACAGATATAAAAA	

ATAGGTAGGCTAAAAAC TTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC ATAGGTAGGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC	GAGCATTACCGCTTGTAAAAGTATTGTATCTAATTTAA AAATTTTTGCAGGGGAGAGACATAAAACGAAATGAA CCACTCCGCCCCAAATTCAAAACCTCATTTTTCCTT IS110 (1220bp) TAGAGAGGGTAAAAGCGGCATGCCAAGTTTGTAAA TTAAAGGAGGAAAGATAAAAATGATAACTTCTGTAAA GAAAACCTAACCCGACCGCATGCGGGGAGATAAAAAA IS110 (1221bp) AATGAAATTCAAATCTCAACAGTAAAGATTTGGC CACATACAGGACATCCACACCCGCTCCGATAACT AACCAAACCTTGAAAAAGCTGCATTTTCAAGTTTA IS110 (1221bp) CTTTTCTGACCACCAAAAACCTATGTGGCTGTTCTG GATTCTACTGCATTATACATCTGTGCTACATCA TTTAAAATCGTATAAAGAGCTTATTGAGCTGATTT	<i>Thermoanaerobacter</i> sp. X513_NC_014538_2149198_2149615_6_ourresults <i>Thermoanaerobacter</i> sp. X513_NC_014538_2149254_2149615_5_gold standard <i>Thermoanaerobacter</i> sp. X513_NC_014538_2340512_2341253_11_ourresults <i>Thermoanaerobacter</i> sp. X513_NC_014538_2340499_2341198_10_gold standard <i>Thermoanaerobacter</i> sp. X513_NC_014538_2342843_2347020_63_ourresults <i>Thermoanaerobacter</i> sp. X513_NC_014538_2342830_2346967_62_gold standard <i>Thermoanaerobacter</i> sp. X513_NC_014538_2348597_2362246_205_ourresults <i>Thermoanaerobacter</i> sp. X513_NC_014538_2348597_2362246_205_gold standard
GTTTTAGCTTACCTATAAGGGATTGAAAC GTTTTAGCTTACCTATAAGGGATTGAAAC GTTTTAGCCTACCTAT ATAGGTAGGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC GTTTCAATCCCTTATAGGTAAGCTAAAAAC	AAGGAAAAATGAGTTTTGAAATTTGGCGGGAGTGG TTAAAATTAGATACAAAATACCTTTTACAAGCGGTAATGCTG IS110 (1220bp) TAGAGAGGGTAAAAGCGGCATGGCAAGTTTGTAAA TTAAAGGAGGAAAGATAAAAATGATAACTTCTGTAAA GAAAACCTAACCCGACCGCATGCGGGGAGATAAAAAA IS110 (1221bp) AATGAAATTCAAATCTCAACAGTAAAGATTTGGC CACATACAGGACATCCACACCCGCTCCGATAACT AACCAAACCTTGAAAAAGCTGCATTTTCAAGTTTA IS110 (1221bp) CTTTTCTGACCACCAAAAACCTATGTGGCTGTTCTG GATTCTACTGCATTATTACATCTGTGCTACATCA TTTAAAATCGTATAAAGAGCTTATTGAGCTGATTT	<i>Thermoanaerobacter</i> sp. X514_NC_010320_764967_765384_6_ourresults <i>Thermoanaerobacter</i> sp. X514_NC_010320_764967_765328_5_gold standard <i>Thermoanaerobacter</i> sp. X514_NC_010320_2340417_2341158_11_ourresults <i>Thermoanaerobacter</i> sp. X514_NC_010320_2340404_2341103_10_gold standard <i>Thermoanaerobacter</i> sp. X514_NC_010320_2342748_2346925_63_ourresults <i>Thermoanaerobacter</i> sp. X514_NC_010320_2342735_2346872_62_gold standard <i>Thermoanaerobacter</i> sp. X514_NC_010320_2348515_2363082_219_ourresults <i>Thermoanaerobacter</i> sp. X514_NC_010320_2348502_2363082_219_gold standard

	CRISPR	DR	CRISPR
		GTTCCGTCGCCCTCTCGGGGTTTTGGGTCTGACGAC	
Mycobacterium africanum GM041182	CRISPR_30_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_29		CRISPR_29-IS6110(1355bp)- CRISPR_29
Mycobacterium bovis AF2122/97	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_24		CRISPR_16-IS6110(1355bp)- CRISPR_24
Mycobacterium bovis BCG str. Korea 1168P	CRISPR_19_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_29		CRISPR_18-IS6110(1355bp)- CRISPR_29
Mycobacterium bovis BCG str. Mexico	CRISPR_19_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_29		CRISPR_18-IS6110(1355bp)- CRISPR_29
Mycobacterium bovis BCG str. Pasteur 1173P2	CRISPR_19_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_29		CRISPR_18-IS6110(1355bp)- CRISPR_29
Mycobacterium bovis BCG str. Tokyo 172	CRISPR_19_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_29		CRISPR_18-IS6110(1355bp)- CRISPR_29
Mycobacterium tuberculosis CAS/NITR204	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis CDC1551	CRISPR_19_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_15		CRISPR_18-IS6110(1355bp)- CRISPR_15
Mycobacterium tuberculosis EAI5	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis EAI5/NITR206	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis str. Haarlem/NITR202	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis H37Ra	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis H37Rv	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_23
Mycobacterium tuberculosis str. Beijing/NITR203	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTTTTGGG-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC_CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_22
Mycobacterium canettii CIPT 140060008	ISL3(2846bp)-GGTITTTGGGTCTGACGAC_CRISPR_26		ISL3(2846bp)- CRISPR_25
Mycobacterium tuberculosis F11	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)- CRISPR_18		CRISPR_17-IS6110(1355bp)- CRISPR_18
Mycobacterium tuberculosis KZN 1435	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)- CRISPR_21		CRISPR_16-IS6110(1355bp)- CRISPR_21
Mycobacterium tuberculosis KZN 4207	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)- CRISPR_21		CRISPR_16-IS6110(1355bp)- CRISPR_21
Mycobacterium tuberculosis KZN 605	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)- CRISPR_21		CRISPR_16-IS6110(1355bp)- CRISPR_21
Mycobacterium tuberculosis CDC5079	CRISPR_15_GTTTCCGTCGCCCTCTC-IS6110(1355bp)		CRISPR_14-IS6110(1355bp)
Mycobacterium tuberculosis CDC5180	CRISPR_15_GTTTCCGTCGCCCTCTC-IS6110(1355bp)		CRISPR_14-IS6110(1355bp)
Mycobacterium tuberculosis CTRI-2	CRISPR_18_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-CRISPR_6_GTTTCCGTCGCCCTCTC-IS6110(1355bp)-CTCGGGTTTTGGGTCTGACGAC-CRISPR_15		CRISPR_17-IS6110(1355bp)- CRISPR_5-
Mycobacterium tuberculosis str. Haarlem	CRISPR_11_GTTTCCGTCGCCCTCTCGGGGTTTTGGGTCTGAC-IS6110(1355bp)-CRISPR_6_GTTTCCGTCGCCCTCTCGGGGTT-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC-CRISPR_24		CRISPR_11-IS6110(1355bp)- CRISPR_5-
Mycobacterium tuberculosis str. Beijing/NITR203	CRISPR_17_GTTTCCGTCGCCCTCTCGGGGTTTTGGG-IS6110(1355bp)-GGGTTTTGGGTCTGACGAC-CRISPR_23		CRISPR_17-IS6110(1355bp)- CRISPR_22

Figure S6. 15 CRISPRs broken by undetected DRs inside

(a)~(g) CRISPRs with undetected DRs inside. The two neighboring CRISPRs have almost identical DRs and one undetected DR in between. The undetected DR may be a full copy. The regions of new DRs matched to the existing DRs are in red. The regions of new spacers are marked by yellow background. The last number in the CRISPR name is the DR copy number.



> *Caldicellulosiruptor obsidiansis* OB47_NC_014392_145343_158051_193

GTTTTATCTGAACTATGAGGGATGTAAC	TCTTTTCTCTTTAACAAATCTTGATATGCCAACTCT	} <i>Caldicellulosiruptor obsidiansis</i> OB47_NC_014392_145343_157028_177_gold standard
GTTTTATCTGAACTATGAGGGATGTAAC	AACGACAAGCTAACGTACAGGAAAAATCCCATGACG	
GTTTTATCTGAACTATGAGGGATGTAAC	GTTCAATTCTGATTATATCAGACTTAAAGTCCGTTG	
GTTTTATCTGAACTATGAGGGATGTAAC	TCAGTAGAATTTTAAAAGATTATTTGACTTTAGGTGG	
.....		
GTTTTATCTGAACTATGAGGGATGTAAC	TCAATGGCTAAAG	} 80bp
GTTTTATCTGAACTATGAGGGATGTAAC	TCAAAATTTACAACAGACATTTTCTCTCTTCTCTC	
GTTTTATCTGAACTATGAGGGATGTAAC	GCACAAAACTGTGGACTACACTTTTAAATCATTAAAG	} <i>Caldicellulosiruptor obsidiansis</i> OB47_NC_014392_157108_158061_14_gold standard
GTTTTATCTGAACTATAAGGGATGTAAC	GATACTATCGAATGCACAAATAAAAACTATTTTA	
GTTTTATCTGAACTATGAGGGATGTAAC	TTGAGAATAAGGATAATCTAACGAAGGGTAACACCA	
GTTTTATCTGAACTATGAGGGATGTAAC	TCTCTACAGGAAATAATTTCATTAATTAATTCATAAA	
.....		
GTTTTATCTGAACTATGAGGGATGTAAC	TTTATAGATGATTTTTTCATAGTACAGATATGCAACA	
GTTTTATCTGAACTATGAG		

> *Thermosiphon africanus* TCF52B_NC_011653_309784_311878_30

GTTTAGAATCTACCTATGAGGAATGAAAAC	CATATTTTTAGATAAAAATACACGCATTAAGTCCCCCA	} <i>Thermosiphon africanus</i> TCF52B_NC_011653_309784_310512_10_gold standard
GTTTAGAATCTACCTATGAGGAATGAAAAC	TTGAATTCCTCGACACCCCTCCTTAAAGATTTTCATTT	
GTTTAGAATCTACCTATGAGGAATGAAAAC	TATACTTCTCTATGTGTCATTAATAGCAATATCTCGTG	
GTTTAGAATCTACCTATGAGGAATGAAAAC	ATTCAAATCTTCTTTAACATAATAACACCTCCATACATTT	
.....		
GTTTAGAATCTACCTATGAGGAATGAAAAC	TAAATCAGT	} 81bp
GTTTAGAATCTACCTATGAGGAATGAAAAC	ATCCCGATGTCGCGGCTTCTCTACTATCTGTATGGATTTT	
GTTTAGAATCTACCTATGAGGAATGAAAAC	TTCATTTTGTATTATACATTTCAATTTCCCTTTTAA	} <i>Thermosiphon africanus</i> TCF52B_NC_011653_310593_311880_18_gold standard
GTTTAGAATCTACCTATGAGGAATGAAAAC	ATCACAGAATAATCTGTGTGTCGATCGCCACTATTTAAATC	
GTTTAGAATCTACCTATGAGGAATGAAAAC	GTGAATAAGTGTTTCTTCAAATTTTCAAGACTAATGGTTTC	
GTTTAGAATCTACCTATGAGGAATGAAAAC	CAAAAATGGCACTTGCAACACTTCTCCTTTAGTTTCTCA	
.....		
GTTTAGAATCTACCTATGAGGAATGAAAAC	GTACGAACTGTCTCAATTTGCAACGGGATAATGT	
GTTTAGAATCTACCTATGAGGAATGAAA		

> *Herpetosiphon aurantiacus* ATCC 23779_NC_009973_102032_109407_101

ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TACGGATGAGGCTGCACCTGCGCTCAGGAGCCGCC	} <i>Herpetosiphon aurantiacus</i> ATCC 23779_NC_009973_102032_108752_92_gold standard
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	GTTGAAGGGATTCATGTTTATATTCACCCGCGCT	
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	ACGCATCGATCCGATCATGGATCGCCACAACTGGT	
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	AAACCTTTTATAGCTCGGCGCACTATGGCATGAT	
.....		
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TCGCGCACTTGGCCAACTTGGGCATTTGGACAAACGT	} 112bp
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TATCCAAATAACATTTGCCTATATTTGCATCGCACATC	
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TTAGAGGATACTGAAACATTCATCCATGGTTGCCAGAACCTGGGGCATAG	} <i>Herpetosiphon aurantiacus</i> ATCC 23779_NC_009973_108864_109408_7_gold standard
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TCGATCACGTTGTACGGGCAACGCTTGGCATTAG	
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	AGCAGCTTCATGGCGCTGCGGGTGTGCAATTCG	
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	GCCATGCCTCGGCTGTCTCAACAAATGGCTCCAT	
.....		
ATTTCAATFACTCGATCCGATTAGAGGATACTGAAAC	TGCATGGCGTTGAGTTCAGTGGCATCAAGCCAAGGCT	
ATTTCAATFACTCGATCCGCTGGACGAATATTGA		

> *Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:- str. 08-1736_NC_021820_4332392_4333865_24

<pre> GTTTATCCCCTGGCGCGGGGAACA TTTTTCAGCCCTGTGCGACTGCGGAACCCCT CGGTTTATCCCCTGGCGCGGGGAACAC GCGAAATAGTGGGAAAAACCCCTGGTTAAC CGGTTTATCCCCTGGCGCGGGGAACAC TAGGCCTTGATACCATCGCTCGCACCTCGTCA CGGTTTATCCCCTGGCGCGGGGAACAC GTTTATTAAGTCTAGTTAAATTAATGGGTTG CGGTTTATCCCCTGGCGCGGGGAACAC GATCGAGTAACGTGGCTGGAACGGG TCGGCGCGGGGAACAC AAAATTAAGCCGAGGGTGGCACCGCCCTATT CGGTTTATCCCCTGGCGCGGGGAACAC GCACCTCGAACCCTTTTAAACACTACCGTTT CGGTTTATCCCCTGGCGCGGGGAACAC TGGACCGATGGGGCAACATCGCGAAGCTGG CGGTTTATCCCCTGGCGCGGGGAACAC GTTACGTTTCGTAATGAAAGCGCGAATAT CGGTTTATCCCCTGGCGCGGGGAACAC CCAGAAAGTGCCTAGTGCCTGATGACGAC CGGTTTATCCCCTGGCGCGGGGAACAC CAGCAGAAAAATTATTACTGCTGTGCTCA CGGTTTATCCCCTGGCGCGGGGAACAC </pre>	<p><i>Salmonella enterica</i> subsp. <i>enterica</i> serovar 4,[5],12:i:- str. 08-1736_NC_021820_4332390_4332968_9_gold standard</p> <p>→ 75bp</p> <p><i>Salmonella enterica</i> subsp. <i>enterica</i> serovar 4,[5],12:i:- str. 08-1736_NC_021820_4333042_4333865_13_gold standard</p>
<pre> gatcgagtaacgtg-gctggaacg gatcgagtaacgtg-gctggaacg </pre>	<p><i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. 798</p> <p>0.962_NC_017046_2_10_spacer</p>

- > *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. 41578_NC_021810_6276_7936_27
- > *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. B182_NC_017623_3740803_3742463_27
- > *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. SL476_NC_011083_3051219_3052879_27
- > *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. 14028S_NC_016856_3096850_3098323_24
- > *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. D23580_NC_016854_3069600_3071012_23
- > *Salmonella enterica* subsp. *enterica* serovar Heidelberg str. CFSAN002069_NC_021812_3668815_3670469_26
- > *Salmonella enterica* subsp. *enterica* serovar Typhimurium var. 5_str. CFSAN001921-NC_021814_490784_492562_28
- > *Salmonella typhimurium* LT2_NC_003197_3076613_3078147_25

> *Mycobacterium tuberculosis* RGTB423_NC_017528_3115972_3117672_23

<pre> TCAGGGTTTGGGCTGACGAC TCGCGGAGCAGCTCTCACCCAGCAGCGGTGAGGTTGG GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC ACGGACGAGCTGACCGCATCAGCATGCTGAGCTGAGG GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC TTGTCTCAAGTGTGCGGTGACACCGCTCCAA GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC CTCGACGGCCCAAGCCAGTTAGCGCAATCCA GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC ATGCGAATCCGCTGTCAGCATGGGATCCGAGT GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC TAGGGCGCCCGCGAGGCTGGGGCGGTTTCACGC GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC TGACGAC GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC CAGCGCAGACGCGAGCCCGAGTACTCGCTCTCCTCAG GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC AGGCTGAAATTAAGCCGGAATGACGACGCAATGGT GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC CTAAGCCCGTAATCCGCAAGTGTGTCAGAAA GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC TGATGATTGTCGCGTATGACGTGCTACTGAGGTGTT GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC GTGCAAGAATCCGGTTCAGTGCACACCGTTTAA GTTTCGGTCCCCTCGGGGTTTGGGCTGACGAC </pre>	<p><i>Mycobacterium tuberculosis</i> RGTB423_NC_017528_3115959_3116942_13_gold standard</p> <p>→ 109bp</p> <p><i>Mycobacterium tuberculosis</i> RGTB423_NC_017528_3117050_3117672_8_gold standard</p>
<pre> atcggaatccgctgctcagcacatggga-tccgagt atcggaatccgctgctcagcacatgggat tccgagt </pre>	<p><i>Mycobacterium tuberculosis</i> H37Rv</p> <p>97.1%_NC_000962_6_14_spacer</p> <p><i>Mycobacterium tuberculosis</i> H37Rv</p> <p>94.9%_NC_000962_6-15_spacer</p>

> *Thermococcus onnurineus* NA1_NC_011529_994458_997976_52

<pre> TTTCGATTCTCTTAGAATCTTATTGCAAC TAAGAAATGGGCACTAGCTTCGGCGGTTGGAATTTG GTTTCAATTCCTTAGAGTCTTATTGCAAC CTGTGGTGGTTCACCAGAGCTTTGCAACGATTAC TTTCAATTCCTTAGAGTCTTATTGCAAC CCTTGAATCCGATAAGCGGGAGCTACGAGTTTC GTTTCAATTCCTTAGAGTCTTATTGCAAC TTCTTGCCACGGTCGCTTACATCAAGCTACGATGA GTTTCAATTCCTTAGAGTCTTATTGCAAC TCGCTTCGACGGTGGCTTTGTCTCCGGATTCATCATC GTTTCAATTCCTTAGAGTCTTATTGCAAC GAAGTTGGCACTCCAACATGCGCCC GTTTCAATTCCTTAGAGTCTTATTGCAAC CAGATCGCTGCTGGTCTGGTCACTGCGCGTGTCTGC TTTCAATTCCTTAGAGTCTTATTGCAAC CCGGCATCAGGGCTGACACCTACTCAACGACCTCTATG GTTTCAATTCCTTAGAGTCTTATTGCAAC AGGAGGAAGATGCAAGTATCATGTTTATGATAAGAA GTTTCAATTCCTTAGAGTCTTATTGCAAC ACGAATCTATCAACCGAGTCAATGGCCCGAGTTCTA GTTTCAATTCCTTAGAGTCTTATTGCAAC TCCCGTTCACGAGCTGTGGCCGCCGCAATCTG GTTTCAATTCCTTAGAGTCTTATTGCAAC </pre>	<p><i>Thermococcus onnurineus</i> NA1_NC_011529_994457_994969_7_gold standard</p> <p>→ 88bp</p> <p><i>Thermococcus onnurineus</i> NA1_NC_011529_995057_997976_43_gold standard</p>
---	--

Figure S7. 8 CRISPRs broken at the beginning of circular chromosomes

(a) A CRISPR is broken into two at the beginning of a circular chromosome of *Cyanobacterium aponinum* PCC 10605. One more spacer is proposed to combine the two CRISPRs into one longer CRISPR. The added region is highlighted in red. The last number in the CRISPR name is the DR copy number.

(b)~(h) The other CRISPRs are broken into two at the beginning of a circular chromosome.

```
> Cyanobacterium aponinum PCC 10605_NC_019776_4113500_4114099+1_810_19
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC AAAAGATAAAAAGGAGGTGAGTTAATCCACTFAAAA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC GAAAAGTTGCAAGGTTTAAAAGTTGCTTACG
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC GGGATAGACTTAATAAAGAATTTGAGGGCTTTA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC TGGTTAAAGCGGATAAGTATTATTGTTTTATCCCA
.....
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC AACTTTTTTCATTCAAATCCACATTTGCCATCGTA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC AAAAAATTGACCCCCCCCCCGTGAGACTAACGAGACT
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC AAAACACC AAAATGGGAGTACTT AAGTACAAGCGA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC GTTCTAGCCGATATTTCTCGTCAAACCATTCGT
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC TGACGATGTGATCGGTGAATTGGAAGCTCACATTCAA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC TTTGGATTATTGGGATTAGTATTGGGATTGGGAAT
.....
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC ATCAAAAAGCCAATCAATATCAGTATTAATCTGCA
GTTTCAATCCCTAAGAGGTAATAAAGAGTTTAAAC
```

Cyanobacterium aponinum PCC 10605_NC_019776_4113500_4114042_7_gold standard

110bp

Cyanobacterium aponinum PCC 10605_NC_019776_54_810_10_gold standard

```
> Thermococcus sp. CL1_NC_018015_1949366_1950313+1_682_24
GTTGCAATAAGACTCTGGGAAATTGAAAC GCGGGCACTTCATCACCAGCCGACCCAATCACCAT
GTTGCAATAAGACTCTGGGAAATTGAAAC CACAGGGTATAAAGTCCGTGCTGATGGCGTGAACGTG
GTTGCAATAAGACTCTGGGAAATTGAAAC GGGCGGGTGTGTTATCTCTATCATCATACGTAGGT
GTTGCAATAAGACTCTGGGAAATTGAAAC GAAACTGAAAAGCTTTAACTGCCCTGCTTTCTGC
.....
GTTGCAATAAGACTCTGGGAAATTGAAAC TAGGGTGGGAAATTGTAATCGACCCGAGTAGGGGGAGAA
GTTGCAATAAGACTCTGGGAAATTGAAAC GTGATATGGCGAAGCCTCTTACT
GTTGCAATAAGACTCTGGGAAATTGAAAC TTGGGGTTTTGAGGTTGGGGAAGTTGCCCAT
GTTGCAATAAGACTCTGGGAAATTGAAAC ATTAGTATCACGGCAGGCTTATCGCCCTGCTACTA
GTTGCAATAAGACTCTGGGAAATTGAAAC TAGAGTATCACTTGCTCCACGCCATTGGCTTAGCCCTCA
GTTGCAATAAGACTCTGGGAAATTGAAAC TAAGCATTTTTGCCTGCTGCTGATGTAGTCTTAA
.....
GTTGCAATAAGACTCTGGGAAATTGAAAG CTTTCTCTGCAAAAGTCGAGCAAGCTTTGAGATACT
GTTGCAATAAGACTCTGGGAAAGAGAAAAG
```

Thermococcus sp. CL1_NC_018015_1949366_1950270_13_gold standard

90bp

Thermococcus sp. CL1_NC_018015_48_682_9_gold standard

```
> Candidatus Puniceispirillum marinum IMCC1322_NC_014010_2753369_2753527+1_1792_29
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAACGCTTTCGGTCTCGTGAACCTTTTCGTAGGA
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAACGCTATAAGTCAACAACAATAATCGATGTTG
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC TGTGAGGTGTTTTGTGAGATAGTGAGCTGC
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC CGTGGTGGTCTCAAGCAGATAAACCCCTTCC
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC TGACCCTTGATATCGCATCACCGCACGGC
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC GTAACGCCATTGTGCCGCAAGTCGCGGGCC
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC ATGAGCAGAATTCATTATCACCTTTGACTT
.....
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC CGATCCCTGATGATGGTTGACCGCCGAA
AGTATAGCACTCTGGTATTGAGAGCC TAGAGCAAC CCCCCCA
```

Candidatus Puniceispirillum marinum IMCC1322_NC_014010_2753369_2753527_2_gold standard

39bp

Candidatus Puniceispirillum marinum IMCC1322_NC_014010_40_1792_26_gold standard

```
> Thermocrinis albus DSM 14484_NC_013894_1496455_1500577+1_90_62
CTTTCAACTCCACACGGTACATTAGAAAC ACTATCTGAACTTCTCACAGTATGCAAAAACAGCG
CTTTCAACTCCACACGGTACATTAGAAAC TGCTGGACAAGGCACTGGCAAGGATGAGACAAAGAT
CTTTCAACTCCACACGGTACATTAGAAAC TACTTATCAAGCGTTTCCCTGTAGACTTTTTGACGTAG
CTTTCAACTCCACACGGTACATTAGAAAC TATCAATGACTTGAGGGCTGACACCACTTCTTTTTT
.....
CTTTCAACTCCACACGGTACATTAGAAAC TGTAGTCTGTACTTCTTAATAACCTGTTCTATTGATGC
CTTTCAACTCCACACGGTACATTAGAAAC TGTCTGTCAGGTACGAAAGAAACAGTGTATATAGATGATA
CTTTCAACTCCACACGGTACATTAGAAAC TCCAGCTCGGGCTTTCCTCAGTCCCTTGTCTTGC
CTTTCAACTCCACACGGTACATTAGAAAC
```

Thermocrinis albus DSM 14484_NC_013894_1496455_1500535_60_gold standard

42bp











Thermocrinis albus DSM 14484_NC_013894_1_90_1_gold standard

> *Fervidicoccus fontis* Kam940_NC_017461_1319183_1319206 +1_794_13

GAATCTCTCAGATAGAATTGAAAG	TCTGAAAAATCATCACAACATCAGCTTTTCCCT	 60bp  60bp  60bp  60bp  60bp  60bp  60bp  60bp  60bp  60bp
GAATCTCTCAGATAGAATTGAAAG	TTGGCATAATTTACTCCCCTTGCTCAGAAATCAAGCATT	
GAATCTCTCAGATAGAATTGAAAG	TTGGCATAATTTACTCCCCTTGCTCAGAAATCAAGCATT	
GAATCTCTCAGATAGAATTGAAAG	TTCAACAAATTTAGTCACAAATGCGCTAGGAAACAGAA	
GAATCTCTCAGATAGAATTGAAAG	TCAAACCCATGCTTCTTAGTATTCTGCTGCTTCT	
GAATCTCTCAGATAGAATTGAAAG	CTCTTGCTAATGCTCTAGCAGTCGATCATAGTAGGC	
.....		
GAATCTCTCAGATAGAATTGAAAG	AACTCTTCCATGTATTAACCAAACCTTCCAAATCCT	
GAATCTCTCAGATAGAATTGAAAG		

Fervidicoccus fontis Kam940_NC_017461_37_794_12_gold standard













> *Picrophilus torridus* DSM 9790_NC_005877_1540318_1545895+1_309_87

CTTTCAATCCTATTTAGGTTATTATTAAAC	TCCGCCGAATTTCAAACACTTATAGTAGGACTAGA	 36bp  36bp  36bp  36bp  36bp  36bp  36bp  36bp  36bp  36bp
TTTTCAATCCTATTTAGGTTATTATTAAAC	TATCTCTCTTTTCAATCCTTATCGATTTATCGATCC	
CTTTCAATCCTATTTAGGTTATTATTAAAC	CACGGTGGTATGCAACGATAAGATCAACATTATCA	
CTTTCAATCCTATTTAGGTTATTATTAAAC	TCCATCTAAAGACATCTCTCTAAATATATCAAAGCC	
.....		
CTTTCAATCCTATTTAGGTTATTATTAAAC	TATAATCTGCTTCCCTACATATCTAACAGAACCTCT	
CTTTCAATCCTATTTAGGTTATTATTAAAC	CATTGATTTTTCATGAAATCATGAATTCGGATAA	
CTTTCAATCCTATTTAGGTTATTATTAAAC	CTGGATTTCTTGGCCATATTTGCGAGATACTATA	
CTTTCAATCCTATTTAGGTTATTATTAAAC	ATATATCGGATTTCAAACCTTATAGATAGACTAGA	
CTTTCAATCCTATTTAGGTTATTATTAAAC	GCCAGAGGCATTTCTCCGTGGTGGTCTTAAATGA	

Picrophilus torridus DSM 9790_NC_005877_1540318_1545873_82_gold standard

Picrophilus torridus DSM 9790_NC_005877_15_309_4_gold standard













> *Thermococcus litoralis* DSM 5473_NC_022084_2214800_2215172+1_2779_46

CTTTCAATCTTTTTAAAGCTTATTGGAAC	AAAGCGTTAATAAGATGATGTCCTACTCTCCAT	 38bp  38bp  38bp  38bp  38bp  38bp  38bp  38bp  38bp  38bp  38bp  38bp
TTTTCAATCTTTTTAAAGCTTATTGGAAC	CTGTTTCATCTACGGCGACAAGAAATGGATGATC	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	ATAGGAGTGCTTCCAGTCAATCTTCTCTGGTATGT	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	ATATATATAAATTAATTCGTTACTGTACAATAAATA	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	CACCTCAGAATATGTTGAAGTTTAGGCTGTGATGCT	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	ACCGAAAAGAAACCAAAAAGGAGGTTGAAATAGA	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	AGCTTGATGGCTTTGAATGAAAAACTGCCCTTTGA	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	TAAACTGTGAGATTGCGGCGATTTACTATA	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	AAAGACAACCTGTAACAATGGTGTGCTTGTGCTGT	
CTTTCAATCTTTTTAAAGCTTATTGGAAC	TTGAAAAAGAACTCAGGAGCAATCAACCTATACAAGCA	

Thermococcus litoralis DSM 5473_NC_022084_2214800_2215162_5_gold standard

Thermococcus litoralis DSM 5473NC_022084_29_2779_40_gold standard

> *Acidilobus saccharovorans* 345-15_NC_014374_1495994-1496453+1-1050_23

GTTTCAATAAGTTTCACGGTTTACTTCTGGGAGTGAGGACCTGGATAGCCGCCACAGC	ACTTCTGGGAGTGAGGACCTGGATAGCCGCCACAGC	 42bp  42bp  42bp  42bp  42bp  42bp  42bp  42bp  42bp  42bp  42bp  42bp
TTTTCAACACCATTCTTGGTTTC	TTTCAGGATAGGAGAAATGCCAAATATGACAGAGGTGTTACGCTGTCTATGAA	
GTTTCAACACCATTCTTGGTTTC	CTCCACGTCTACGGCAATGAACTGAACAGCAAGCAACGC	
GTTTCAACACCATTCTTGGTTTC	AGAAATCAGTAGTCCAGGTATAGACGTCCCAAACAGATA	
.....		
GTTTCAACACCATTCTTGGTTTC	ACTGCCTTCAGTCCACGGTTGGCGGATAAACACTG	
GTTTCAACACCATTCTTGGTTTC	CTGGAGAGGCCCCACAGGCCTCCCTGAGGCCAGACATCCA	
GTTTCAACACCATTCTTGGTTTC	CCATTGTAAATGAGATGTGTTACGGATCCAGAGCATAA	
GTTTCAACACCATTCTTGGTTTC	GCCGTGGAAGTAGTCTCGTGTGACCAACCTCTCCCA	
GTTTCAACACCATTCTTGGTTTC	TGATGTAACCTGGGCGCTGACAGGTACATGTAGGTG	

Acidilobus saccharovorans 345-15_NC_014374_1495994_1496411_6_gold standard

Acidilobus saccharovorans 345-15_NC_014374_1_1050_16_gold standard