

## Coding sequences of genes cytidine deaminase

>A2\_HOMO

ATGGCCCAGAAGGAAGAGGCTGCTGTGGCCACTGAGG**CTGCC**TCCCAG**GAATGG**GGAGGATCTGGAGAACCTGGA  
CGACCCTGAGAAGCTGAAAGAGCTGATTGAG**CTGCC**GCCCTTTGAGATTGTCACAGGAGAACGG**CTGCC****CTGCCA**  
ACTTCTTTAAATTCCAGTTCCG**GAATG**TGGAGTACAGTTCCGGGAGGAA**CAAGA**CCCTTCCTCTGCTATGTGGTT  
GAAGCACAGGGCAAGGGGGGCCAAGTGCAGGCATCTCGGGGATACCTAGAG**GATGA**GCATGCGG**CTGCC**CATGC  
AGAGGAAGCTTT**TTCAA**CACCATC**CTGCCA**GCCTTCGACCCAGCCCTGC**GGTACA**ATGTCACCT**TGGTA**TGTGT  
CCTCCAGCCCCTGTGCAGCGTGTGCTGACCGCATTATCAAACCCTTAG**CAAGA****CAAGA**ACCTGCGTCTGCTC  
**ATTC**TGGTGGGTGCGACTCTTCATGTGGGAGGAGCCGGAGATCCAGGCTGCTCTGAAGAAGCTGAAGGAGGCTGG  
CTGTAA**ACTGCG**GCATC**ATGAA**GCCCCAGGACTTCGAATATGTCTGGCAGAATTTTGTGGAG**CAAGA**AGAGGG**TG**  
**AAT**CCAAGGCCCTTCAGCCCTGGGAGGA**CATTCA**GGAGA**ACTT**CCTATACTACGAGGAGAAGTTGGCAGACATC  
CTGAAG

24 668 3.6

>A1\_HOMO

ATGACTCCAGAGGAGGAAGTCCAGAGACAGAGCACCATGACTTCTGAGAAAGGTCC**TTCAA**CCGGTGACCCAC  
TCTGAGGAGAAGAATCGAACCCCTGGGAGTTTGACGTCTTCTATGACCCAGAGAACTTCGTAAAGAGGCCTGTC  
TGCTCTACGAAA**TCAAG**TGGGGCATGAGCCGG**AAGAT**CTGGCGAAGCTCAGGCAAAAACACCACCAATCACGTG  
GAAGTTAATTTTATAAAAAAATTTACGTCAGAAAGAGATTTTACCCATCCATGAGCTGCTCCATCACCTGGTT  
CTTGTCCCTGGAGTCCCTGCTGG**GAATG**CTCCAGGCTATTAGAGAGTTTCTGAGTCGGCACCCCTGGTGTGACTC  
TAGTGATCTACGTAGCTCGGCTTTTTTGGCACATGGATCAACAAAATCGGCAAGGTCTCAGGGACCTTGTTAAC  
AGTGGAGTA**ACTATTCA**GATTATGAGAGCATCAGAGTATTATC**ACTGC**TGGAGGAATTTTGTCAACTACCCACC  
TGGG**GATGAA**GCTCACTGGCCACAATACCCACCTCTGTG**GATGAT**TGTTGTACGCACTGGAGCTGC**ACTGC**ATAA  
TTCTAAGTCTTCCACCCTGTTTA**AAGATTTCAAGAAGATG**GCAAATCATCTTACATTTTTTCAGACTTTCATCT**TT**  
**CAA****ACTGCCATT**ACCAAACGATTCGGCCACACATCCTTTTAGCTACAGGGCTGATACATCCTTCTGTGGCTTG  
GAGA

21 740 2.8

>AID\_HOMO

ATGGACAGCCTCTT**GATGAA**CCGGAGGAAGTTTCTTTACCA**ATTCAA**AAATGTCCGCTGGGCTAAGGGTCCGGC  
TGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTC  
GCAATAAGAACGG**CTGCCA**CGTGGAATTGCTCTTCCCTCCGTACATCTCGGACTGGGACCTAGACCCTGGCCGC  
TGCTACCGCGTCACCTGGTTACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCG  
AGGGAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACC**GCAAGGCTGAGCCG**  
AGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAG**GCCAT**CATGACC**TTCAAAGAT**TATTTTT**TACTGC**TGG  
AATACTTTTGTAGAAAACC**ATGAA**AGAACT**TTCAA**AGCCTGGGAAGGGCTGC**ATGAA****AATTCAG**TTTCGTCTCTC  
CAGACAGCTTCGGCGCATCCTTTTGCCCCTGTATGAGGTT**GATGACTT**ACGAGACGCATTT**CTACTT**TGGGAC  
TT

17 590 2.9

>A3A\_HOMO

ATGGAAGCCAGCCAGCATCCGGGCCAGACACTTGATGGATCCACACAT**ATTCACTT**CCA**ACTTTAAC****AATGG**  
CATTGGAAGGCATAAGACCTACCTGTGCTACGAAGTGGAGCGCCTGGAC**AATGG**CACCTCGG**TCAAGATG**GACC  
AGCACAGGGGCTTTCTACACAACCAGGCTAAGAATCTTCTCTGTGGCTTTTACGGCC**GCCAT**GCGGAGCTGCGC  
TTCTTGACCTGGTTCCCTTCTTTGCAGTTGGACCCGGCCAGATCTACAGGGTCACTTGGTT**CATCTCCTGGAG**  
CCCCTGCTTCTCCTGGGGCTGTGCCGGGGAAGTGCCTGCGTTCCCTCAGGAGAACACACACGTGAG**ACTGCG**CA  
TCTTCG**CTGCC**CGCATCTATGATTACGACCCCTATATAAGGAGGCGCTGCAAATGCTGCGGGATGCTGGGGCC  
CAAGTCTCCATCATGACCTAC**GATGAAT**TTAAGC**ACTGCT**TGGGACACCTTTGTGGACCACCAGGGATGTCCCTT  
CCAGCCCTGGGATGGACT**AGATGAG**CACAGCCAAGCCCTGAGTGGGAGGCTGCGG**GCCATTC**TCCAGAATCAGG  
GAAAC

19 593 3.2

>A3B\_N\_HOMO

**ATGAAT**CCACAGATCAGAAATCCGATGGAGCGGATGTATCGAGACA**CATTC**TACGACAACCTTTGAAAACGAACC  
CATCCTCT**ATGGT**CGGAGCTACACTTGGCTGTGCT**ATGAA**GTGAAAATAAAGAGGGGCCGCTCAAATCTCCTTT  
GGGACACAGGGGTCTTTCGAGGCCAGGTGTAT**TTCAAG**CCTCAGTACCACGCAGAAATGTGCTTCTCTCTTGG  
TTCTGTGGCAACCAG**CTGCC**TGCTTACAAGTGTTCAGATCACCTGGTTTGTATCCTGGACCCC**CTGCC**CGGA  
CTGTGTGGCGAAGCTGGCCGAATTCCTGTCTGAGCACCCCAATGTCACCCTGACCATCT**CTGCC**CGCCGCTCT  
ACTAC**ACTG**GGAAAGAGATTACCGAAGGGCGCTCTGCAGGCTGAGTCAGGCAGGAGCCCGCGTGACGATCATG  
GACT**ATGAA**GAATTTGCA**TACTGC**TGGGAAAACCTTTGTGTACA**ATGAA**GGTCAGCA**ATTCATGCCTTGGTACAA**  
ATTC**GATGAA**AATTATG**CATTC**CTGCACCGCACGCTAAAGGAGATTCTCAGA

21 566 3.7

>A3B\_C\_HOMO

TACCTGATGGATCCAGACA**CATTC**ACT**TTCAA**CTTTAATAATGACCCTTTGGTCCTTCGACGGCGCCAGACCTA  
CTTGTGCTATGAGGTGGAGCGCTGGAC**AATGG**CACCTGGGTCCTGATGGACCAGCACATGGGCTTTCTATGCA  
ACGAGGCTAAGAATCTTCTCTGTGGCTTTTACGGCC**GCCAT**GCGGAGCTGCGTTCTTGGACCTGGTTCTTCT  
TTGCAGTTGGACCCGGCCAGATCTACAGGGTCACTTGGTTCATCTCCTGGAGCCCCCTGCTTCTCCTGGGGCTG  
TGCCGGGAAGTGCCTGCGTTCTTTCAGGAGAACACACACGTGAG**ACTGCG**CATCTTCG**CTGCC**CGCATCTATG  
ATTACGACCCCTATATAAGGAGGCGCTGCAATGCTGCGGGATGCTGGGGCCCAAGTCTCCATCATGACCTAC  
**GATGA**GTTTGA**GTACTGC**TGGGACACCTTTGTGTACCGCCAGGGATGTCCCTTCCAGCCCTGGGATGGACTAGA  
GGAGCACAGCCAAGCCCTGAGTGGGAGGCTGCGGG**GCCATTC**TCCAGAATCAGGGAAAC

14 572 2.5

>A3C\_HOMO

**ATGAAT**CCACAGATCAGAAACCC**GATGAA**GGCAATGTATCCAGGCA**CATTC**TACTTCCAATTTAAAAACCTATG  
GGAAGCCAACGATCGGGACGAAACTTGGCTGTGCTTCACCGTGGAAGGTATAAAGCGCCGCTCAGTTGTCTCCT  
GGAAGACGGGCGTCTTCCGAAACCAGGTGGATTCTGAGAC**CCATT**GTCATGCAGAAAGGTGCTTCTCTCTTGG  
TTCTGCGACGACA**TACTG**TCTCCTAACACAAAGTACCAGGTCAC**TGGTAC**ACATCTTGGAGCCCTTGCCAGA  
CTGTGCAGGGGAGGTGGCCGAGTTCCTGGCCAGGCACAGCAACG**TGAAT**CTCACCATCTTACCAGCCCGCTCT  
ACTACTTCCAGTATCCATGTTACCAGGAGGGGCTCCGCAGCCTGAGTCAGGAAGGGGTGCTGTGGAGATCATG  
GACT**ATGAAGAT**TTTAAATATTGTTGGGAAAACCTTTGTGTACAATGATAATGA**GCCATTC**AAGCCTTGAAGGG  
ATTTAAAAACCAACTTTTCGACTTCTGAAAAGAAGGCTACGGGAGAGTCTCCAG

18 566 3.2

>A3D\_N\_HOMO

**ATGAAT**CCACAGATCAGAAATCCGATGGAGCGGATGTATCGAGACA**CATTC**TACGACAACCTTTGAAAACGAACC  
CATCCTCT**ATGGT**CGGAGCTACACTTGGCTGTGCT**ATGAA**GTGAAAATAAAGAGGGGCCGCTCAAATCTCCTTT  
GGGACACAGGGGTCTTTCGAGGCC**GGTACT**ACCCAAACGTCAGTCGAATCACAGGCAGGAGGTGATTTCCGG  
TTTGAGAACCACGCAGAAATGTGCTTCTTATCTTGGTTCTGTGGCAACCG**ACTGCC**TGCTAACAGGCGCTTCCA  
GATCACCTGGTTTGTATCATGGAACCC**CTGCC**TGCCCTGTGTGGTGAAGGTGACCAAATTCCTTGGCTGAGCACC  
CCAATGTCACCCTGACCATCT**CTGCC**CGCCGCTCTACTACTACCGGATAGAGATTGGCGGTGGGTGCTCCTC  
AGGCTGCATAAGGCAGGGGCCCGTGTG**AAGAT**CATGGACT**ATGAA**GACTTTGCAT**TACTGC**TGGGAAAACCTTTGT  
GTGCA**ATGAA**GGTCAG**GCCATTC**ATGCCT**TGGTAC**AAATTC**GATGA**CAATTATGCATCCCTGCACCGCACGCTAA  
AGGAGATTCTCAGA

23 602 3.8

>A3D\_C\_HOMO

AACCCGATGGAGGCAATGTACCCACACATATTC**TACTT**CCACTTTAAAAAC**TACTG**AAAGCCTGTGGTGGAA  
CGAAAGCTGGCTGTGCTTACCATGGAAGTTACAAAGCACCCTCAGCTGTCTTCCGGAAGAGGGGCGTCTTCC  
GAAACCAGGTGGATCCTGAGAC**CCATT**GTCATGCAGAAAGGTGCTTCTCTCTTGGTTCTGTGACGACAT**TACTG**  
TCTCCTAACACAAACTACGAGGTCAC**TGGTAC**ACATCTTGGAGCCCTTGGCCAGAGTGTGCAGGGGAGGTGGC  
CGAGTTCTGGCCAGGCACAGCAACG**TGAAT**CTCACCATCTTACCAGCCCGCTCTGCTACTTCTGGGATACAG  
ATTACCAGGAGGGGCTCTGCAGCCTGAGTCAGGAAGGGGCTCCGTG**AAGAT**CATGGGCTACA**AAGAT**TTTGT

TCTTGTGGAAAACTTTGTGTACAGT**GATGATGAGCCATTCAAG**CCTTGGAAAGGGACTACAAACCAACTTTTCG  
ACTTCTGAAAAGAAGGCTACGGGAGATTCTCCAG

16 548 **2.9**

>A3F\_N\_HOMO

**ATGAAG**CCTCACTTCAGAAACACAGTGGAGC**GAATG**TATCGAGACA**CATTC**TCCTACAACCTTTTATAATAGACC  
CATCCTTTCTCGTCGGAATACCGTCTGGCTGTGCTACGAAGTAAAACAAAGGGTCC**CAAG**GCCCCGTTTGG  
ACGCA**AAGAT**CTTTCGAGGCCAGGTGTATTCCCAGCCTGAGCACCACGCAGAAATGTGCTTCCTCTCTTGGTTC  
TGTGGCAACCAG**CTGCC**TGCTTACAAGTGTTCAGATCACCTGGTTTGTATCCTGGACCC**CTGCC**CGGACTG  
TGTGGCGAAGCTGGCCGAATTCTGTCTGAGCACCCCAATGTCACCCCTGACCATCTCCGCCGCCGCCTCTACT  
AC**TACTG**GGAAAGAGATTACCGAAGGGCGCTCTGCAGGCTGAGTCAGGCAGGGGCCCGTGT**AAGAT**TATGGAC  
**GATGA**GAATTTGC**TACTGC**TGGAAAACCTTTGTGTACAGTGAAGGTCA**GCCATTCA**TGCCT**TGGTAC**AAATT  
**CGATGA**CAATTATG**CATTC**CTGCACCCGCACGCTAAAGGAGATTCTCAGA

20 563 **3.6**

>A3F\_C\_HOMO

AACCCGATGGAGGCAATGTATCCACACATATTCTACTTCCACTTTAAAAACCTACGCAAAGCCT**ATGGT**CGGAA  
CGAAAGCTGGCTGTGCTTCACCATGGAAGTTGTAAAGCACCCTCACCTGTCTCCTGGAAGAGGGGCGTCTTCC  
GAAACCAGGTGGATCCTGAGAC**CCATT**GTGCATGCAGAAAGGTGCTTCCTCTCTTGGTTCTGTGACGACA**TACTG**  
TCTCCTAACACAAACTACGAGGTCAC**TGGTAC**ACATCTTGGAGCCCTTGCCAGAGTGTGCAGGGGAGGTGGC  
CGAGTTCCTGGCCAGGCACAGCAACG**TGAAT**CTCACCATCTTCACCCGCCCTCTACTACTTCTGGGATACAG  
ATTACCAGGAGGGGCTCCGCAGCCTGAGTCAGGAAGGGCCCTCCGTGGAGATCATGGGCTACA**AAGAT**TTTAAA  
TATTGTTGGAAAACCTTTGTGTACAAT**GATGA**TG**GCCATTCAAG**CCTTGGAAAGGGACTAAAATACAACCTTCT  
ATTCCTGGACAGCAAGCTGCAGGAGATTCTCGAG

14 548 **2.6**

>A3G\_N\_HOMO

**ATGAAG**CCTCACTTCAGAAACACAGTGGAGC**GAATG**TATCGAGACA**CATTC**TCCTACAACCTTTTATAATAGACC  
CATCCTTTCTCGTCGGAATACCGTCTGGCTGTGCTACGAAGTAAAACAAAGGGTCC**CAAG**GCCCCCTTTGG  
ACGCA**AAGAT**CTTTCGAGGCCAGGTGTATTCCGAACCTAAGTACCACCAG**AGATGA**GATTCTTCCACTGGTTC  
AGCAAGTGGAGGAAGCTGCATCGTGACCAGGAGTATGAGGTCAC**TGGTAC**ATATCCTGGAGCCCCGTCACAAA  
GTGTACAAGGGATATGGCCACGTTCTGGCCGAGGACCCGAAGGTTACCCTGACCATCTTCGTTGCCCGCCTCT  
ACTACTTCTGGGACCCAGATTACCAGGAGGCGTTTCGCAGCCTGTGTACAGAAAGAGACGGTCCGCG**TGCCACC**  
**ATGAAGATCATGAAT**TATGACGAATTTTCAGCACTGTTGGAGCAAGTTCGTGTACAGCCAAAGAGAGCTATTTGA  
GCCTTGGAAATAAT**CTGCC**TAAATATTATATAT**TACTGC**ACATCATGCTGGGGGAGATTCTCAGA

17 578 **2.9**

>A3G\_C\_HOMO

CACTCGATGGATCCACCCA**CATTC**ACT**TTCAA**CTTTAACA**ATGAA**CCTTGGGTCAGAGGACGGCATGAGACTTA  
CCTGTGTTATGAGGTGGAGCGCATGCACAATGACACCTGGGTCCTGCTGAACCAGCGCAGGGGCTTCTATGCA  
ACCAGGCTCCACATAAACACGGTTTCTTGAAGGCC**GCCATG**CAGAGCTGTGCTTCTGGACGTGATTCCCTTT  
TGGAAAGCTGGACCTGGACCAGGACTACAGGGTTACCTGCTTCACCTCCTGGAGCCCCGCTTTCAGCTGTGCCCA  
GGA**AATGG**CTAA**ATTCATTTCAA**AAAACAAACACGTGAGCCTGTGCATCTTC**ACTGCC**CGCATCTAT**GATGATC**  
**AAGGAAGATG**TCAGGAGGGGCTGCGCACCTGGCCGAGGCTGGGGCCAAAAT**TTCAA**TAATGACATACAG**TGAA**  
**TTTAAGCACTGC**TGGGACACCTTTGTGGACCACCAGGGATGTCCCTTCCAGCCCTGGGATGGACT**AGATGA**GCA  
CAGC**CAAGA**CCTGAGTGGGAGGCTGCGG**GCCATTC**TCCAGAATCAGGAAAAC

23 566 **4.1**

>A3H\_HOMO

ATGGCTCTGTTAACAGCCGAAA**CATTC**CGCTTACAGTTTAAACAACAAGCGCCGCCTCAGAAGGCCTTACTACCC  
GAGGAAGGCCCTCTTGTGTTACCAGCTGACGCCGCA**GAATGG**CTCCACGCCACGAGAGGCTACTTTGAAAACA  
**AGAAAAGTGCAT**GCAGAAATTTGCTTTATTAACGAGAT**CAAG**TCCATGGGACTGGACGAAACGCAGTGCTAC

CAAGTCACCTGTTACCTCACGTGGAGCCCCTGCTCCTCCTGTGCCTGGGAGCTGGTTGACTTCA**TCAAG**GCTCA  
CGACCATCTGAACCTGGGCATCTTCGCCCTCCCGCCT**GTACT**ACCACTGGTGCAAGCCCCAGCAGAAGGGGCTGC  
GGCTTCTGTGTGGATCCCAGGTCCCGGTGGAGGTCATGGGCTTCCCAGAGTTTGCTG**ACTGC**TGGGAAAACTTT  
GTGGACCACGAGAAACCGCTTTC**TCA**ACCCCTAT**AAGATG**TTAGAGGAGCTAGATAAAAACAGTCGA**GCCAT**  
AAAGCGACGGCTTGAGAGGATAAAGCAGTCC

14 545 2.6

>A3-PSEUDO\_HOMO

**ATGAA**GCCTCAGATCAGAAACCTGATGGAGTGGATGTATCCAGGCA**CATTC**TACTACAACCTTTGAAAACAGACC  
CATCCTCTCAGGTTGGAACACCACCTGGCTGTGCTACAAA**ATGAA**AACAAAGAAGGACCCCTCAAAGCCCCCTT  
TGGACGCAAGGATCTTTGGAGGCCAGGTCTATTCCAAGCCTGAACACCACCCAG**AGATG**AGATTCTGTAGATTGG  
TTCTGCAACTCGAGGCTGCATCGTGACCAGGACTACCTGGTCATC**TGGTAC**ATCTCCTGGAGTCCCTGCTCAGA  
GTATGCAGGGAACGTGGCAGAGTTCCTGGCCAAGGATGGCAAGGTCACCCTGACCATCTTCGTTGCCACCTCT  
ACTACTTCTGGGAAGCAGATTAC**CAAG**AGGAGCTTCAC**AGATGGT**GTCAGAAAAAACAGCCACAT**TGCCAGCATG**  
**AAGATCATGAA**CTAT**BATGAATTTCAA**CACTGTTGGGACAAATTTGTGTACAACCAA**GAATG**CTATTTGATCC  
TTGGAAGCAAC**TGAAT**ACAAATTATGCAT**TA**CTGCACAGCATGCTGGGGGAGATTCTCAGACACTTCACTTACA  
ATTTTACCAATGATCCTTCAGTCTCTGGGCAGCACCAGACCTACCTGTGCTGCAAGGTGGAG**TGCCA**AAACAAT  
GATGCCTGGGTCCCTCTGGACCAGCAACAAGGGCATCTACCCAGCCAGGCTT**TGAAT**CAGTTGCTATTAAGTT  
CCATGCAGAGCAGTGCTTCCTGTACCTGATTTCTTTTGGAAAGCTGGACCGGGCTCAGTGCTACAGGGTCACCT  
GGTTCATCTCCTGGAGTCCCTGCTTACAGCTGTGCCCAGCAAGTGGCTACATTTT**CAT**TGGGAGAAC**AGATG**TGTG  
AGCCTGCACATCTTCGCTGCATGCATCTATAATTATCTCCCAGGATATGAGGGGCTGTGCATGCTGCAGAGGGC  
TGGGACCCAAATCTCCATCATGACCCGGAAGTTTAGGC**ACTGC**TGGGTACCTTTGTGGACCACCAGGGATGTC  
CCTTCCAGCCCTGGGATGGACT**AGATG**AGCACAG**TCAAG**CCCTAAGTGGGAGGCTGCAG**GCCATTC**TCCAGAAT  
CAGGAAAAC

32 1115 2.9

>A2\_MUS

ATGGCTCAGAAGGAAGAGGCCGCTGAGGCTGCGGCGCCAGCCTCTC**GAATGGAGATGA**TTTGGAGAACCTGGA  
AGACCCTGAGAAGCTGAAAGAGCTGATCGATCTTCCGCCCTTCGAGATTGCTACTGGGGTGCGGCTACCAGTCA  
ACTTCT**TTCAAG**TTTTCAGTTCGGGAACGTGGAATACAGTTCGGGGCGGAATAAGACCTTTCTCTGCTATGTGGTC  
GAAGTACAGAGTAAGGGCGGCCAAGCGCAGGCGACGCAGGGCTACCTGGAG**GATGAA**CACGCAGGTGCCACGC  
CGAGGAGGCTTTCTTTAACACCATC**CTGCCA**GCTTTTCGACCCGGCC**TCAAG**TACAATGTCACCT**TGGTA**TGTGT  
CCTCCAGCCCCTGTGCAGCCTGCGCTGACCGGATTCTCAAACCCTCAG**CAAGA**CTAAGAACCTTCGTCTGCTC  
**ATTC**TGGTGAGCCGGCTCTTCATGTGGGAGGAGCCAGAGGTCCAGGCCGCTCTGAAAAAGCTGAAGGAGGCCGG  
CTGCAAACCTTCGCATC**ATGAA**ACCCAGGACTTCGAGTACATCTGGCAGAATTTTGTGGAG**CAAGA**AGAGGGTG  
**AAT**CCAAGGCCTTTGAGCCCTGGGAGGA**CATTC**AGGAGAACTTCTTATACTATGAGGAGAAGTTGGCAGACATC  
CTGAAG

19 668 2.8

>A1\_MUS

ATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCCTCTGAGGAGAAGAATTGAGCCCCACGAGTTTGAAGT  
CTTCTTTGACCCCCGGGAGCTTCGAAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGTGGAAAGGCACAGTG  
TCTGGCGACACACGAGCCAAAACACCAGCAACCACGTTGAAGTCAACTTCTTAGAAAAATTTACTACAGA**AAGA**  
**TACT**TTTCGTCCGAACACC**AGATG**CT**CCATT**ACCTGGTTCCCTGTCTCCTGGAGTCCCTGCGGGGAGTGCTCCAGGGC  
**CATT**ACAGAGTTTCTGAGCCGACACCCCTATGTAACCTCTGTTTATTTACATAGCACGGCTTTATCACCACACGG  
ATCAGCGAAACCGCCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACTATCCAGATCATGACAGAG**CAAG**AG  
TATTGT**TA**CTGC**TGG**AGGAATTTTCGTC**AACT**ACCC**TTCAA**ACGAAGCTTATTGGCCAA**GGTAC**CCCCATCT  
GTGGGTGAAACTGTAT**GTACTG**GAGCTC**TA**CTGCATCATTTTAGGACTTCCACCCTGTTTAAAAATTTTAAAGAA  
GAAAGCAACCTCAACTCACGTTTTTCACAATTACTC**TTCAA**AC**CTGCCATT**ACCAAAGGATACCACCCCATCTC  
CTTTGGGCTACAGGGTTGAAA

19 683 2.8

>A3\_N\_MUS

ATGCAACCCCAGCGCCTGGGGCCAGAGCTGGGATGGGA**CCATTC**TGTCTGGGATGCAG**CCCAT**CGCAAATGCTA  
**TTCA**CCGATCAGAAACCTGATATCT**CAAGAA****CATTCAAG**TTCCACTTTAAGAACCTAGGCTAT**TGCCA**AAGGCC  
GGA**AAGAT**ACCTTCTTGTGCT**ATGAA**GTGACTAGAAAGG**ACTGCGATTCA**CCCGTCTCCCTTACCATGGGGTC  
TTTAAGAACAAGGACAACATCCACGCTGAAATCTGCTTTTTA**TACTG**TTCCATGACAAA**GTA**CTGAAAGTGCT  
GTCTCCGAGAGAAGAG**TTCAAGAT**CACC**TGGTA**TATGTCCTGGAGCCCCTGTTTC**GAATG**TGCAGAGCAGATAG  
TAAGTTTCTGGCTACACACCACAACCTGAGCCTGGACATCTTCAGCTCCCCTCTACAACGTACAGGACCCA  
GAAACCCAGCAGAATCTTTGCAGGCTGGTTTCAGGAAGGAGCCCAGGTGG**CTGCCAT**GGACCTATACGAATTTAA  
AAAGTGTGGAAAGAAGTTTGTGGAC**AATGGT**GGCAGGCG**ATTCAG**GCCTTGAAAAG**ACTGC**TTACAAATTTTA  
GATACCAGGATTCTAAGCTTCAGGAGATTCTGAGG

31 623 5

>A3\_C\_MUS

C**GAATGG**ACCCGCTAAGTGAAGAGGAATTTTACTCGCAGTTTTACAACCAACGAG**TCAAG**CATCTCTGCTACTA  
CCACCG**ATGA**AAGCCCTATCTATGCTACCAGCTGGAGCAG**TTCAAT**GGCCAAGCGCCACTCAAAGG**CTGCC**TGC  
TAAGCGAGAAAGGCAAACAGCATGCAGAAATCCTCTTCTTGAT**AAGAT**TCGGTCCATGGAGCTGAGCCAAGTG  
ACAATCACCTGCTACCTCACCTGGAGCCC**CTGCC**CAAACCTGTGCCTGGCAACTGGCGG**CATTCAA**AAGGGATCG  
TCCAGATCTAATTCTGCATATCTACACCTCCCGCCTGTATTTCCACTGGAAGAGGCCCTTCCAGAAGGGGCTGT  
GTTCTCTGTGGCAATCAGGGATCCTGGTGGACGTCATGGACCTCCACAGTT**TACTGACTGC**TGGACAACTTT  
GTGAACCCGAAAAGGCCGTTTTG**GCCAT**GGAAAGGATTGGAGATAATCAGCAGGCGCACACAAAGGCGGCTCCG  
CAGGAT**CAAG**GAGTCTGGGGTCTG**CAAGAT**TTGGT**GAAATG**ACTTTGAAACCTACAGCTTGGACCCCCGATGT  
CT

20 590 3.4

>AID\_MUS

ATGGACAGCCTTCT**GATGA**AGCAAAGAAGTTTCTTTA**CCATT**TC**CAA**AAATGTCCGCTGGGCCAAGGGACGGCA  
TGAGACCTACCTCTGCTACGTGGTGAAGAGGAGAGATAG**TGCCA**CCTCCTGCTCACTGGACTTCGGCCACCTTC  
GCAACAAGTCTGG**CTGCCA**CGTGGAATTGTTGTTCTACGCTACATCTCAGACTGGGACCTGGACCCGGGCCGG  
TGTTACCGGCTCACCTGGTTCACCTCCTGGAGCCCGTGCTATGACTGTGCCCGGCACGTGGCTGAGTTTCT**GAG**  
**ATG**GAACCTAACCTCAGCCTGAGGATTTTACC CGCGCCTCTACTTCTGTGAAGACCGCAAGGCTGAGCCTG  
AGGGGCTGCGGAG**ACTGCA**CCGCGCTGGGGTCCAGATCGGGATCATGACC**TTCAA**AGACTATTTTT**TACTGC**TGG  
AATACATTTGTAGAAAATCGTGAAAGA**ACTTTCAA**AGCCTGGGAAGGGCTAC**ATGAA**AATTCTGTCCGGCTAAC  
CAGACA**ACTTCGG**CGCATCCTTTTGCCCTTGTACGAAGTC**GATGA**CTTGC**GAGATG**CATTTTCGTATGTTGGGAT  
TT

17 590 2.9

>A2\_RATTUS

ATGGCTCAGAAGGAAGAAGCCGCTGAGGCCGCGGCACCAGCCTCTCAGAACGGAGACGATTTGGAGAACCTAGA  
AGACCCTGAGAAGCTGAAGGAGCTGATCGAT**CTGCC**GCCCTTCGAGATCGTCACTGGGGTGCG**ACTGCC**TGTCA  
ACTTCTTTAAGTTTCAGTTCCGGAACGTGGAATACAGTTCCGGGCGGA**CAAGAC**CTTTCTCTGCTATGTGGTC  
GAAGCACAGAGCAAGGGAGGCCAAGTGCAGGCGACGCAGGGCTACCTGGAG**GATGAA**CACGCTGGTGGCCACGC  
TGAGGAGGCTTTCTTTAACACCATC**CTGCCA**GTTCGACCCGGCCCTGAAGTATAATGTCACCT**TGGTA**TGTGT  
CCTCCAGCCCCTGTGCAGCCTGCGCCGACCGAATT**TCAAGA**CCCTCAG**CAAGACCAAGA**ACCTGCGGCTGCTC  
ATCCTGGTGAGCCGGCTCTTTCATGTGGGAGGAGCCAGAGGTCCAGGCCGCCCTGAAGAAGCTGAAGGAGGCGGG  
CTGCAA**ACTTCGCAT**C**ATGAA**ACCCAGGACTTCGAGTACCTCTGGCAGAATTTTGTGGAG**CAAGA**AGAGGG**TG**  
**AAT**CCAAGGCCTTTGAGCCCTGGGAGGA**CATTCA**GGAGA**ACTTC**CTATACTATGAGGAGAAGTTGGCAGACATC  
CTGAAG

18 668 2.7

>A1\_RATTUS

ATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCACTCTGAGGAGAAGAATTGAGCCCCACGAGTTTGAAGT  
CTTCTTTGACCCCCGGGA**ACTTCG**GAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGAGGAAGGCACAGCA

TCTGGCGACACACGAGCCAAAACACCAACAAACACGTTGAAGTCAATTTTCATAGAAAAATTTACTACAGA**AAGA**  
**TACT**TTTTGTCCAAACACC**AGATGCTCCATT**ACCTGGTTCCTGTCTCTGGAGTCCCTGTGGGGAGTGCTCCAGGG**C**  
**CATT**ACAGAATTTTTGAGCCGATACCCCATGTAACCTCTGTTTATTTATATAGCACGGCTTTATCACCACGCAG  
ATCCTCGAAATCGGCAAGGACTCAGGGACCTTATTAGCAGCGGTGTTACTATCCAGATCATGACGGAG**CAAGAG**  
TCTGGCT**ACTGCT**TGGAGGAATTTGTCAACTACTCCCTT**CGAATG**AGCTCATTGGCCAAGGTACCCCATCT  
GTGGGTGAGGCTGTAC**GFACTG**GAACTC**TACTGC**ATCATTTTAGGACTTCCACCCTGTTTAAATATTTTAAGAA  
GAAAACAACCTCAACTCACGTTTTTTCACGATTGCTC**TTCAAAGCTGCCATT**ACCAAAGGCTACCACCCACATC  
CTGTGGGCCACAGGGTTGAAA

20 683 2.9

>A3\_N\_RATTUS

AGAAACCCGCTAAAGAAGTTATATCAACAAA**CATTC**TACTTTCATTTTAAAGAACGTACGCTATGCCTGGGGTCTG  
AAAGAATAACTTCTTGTGCT**ATGAAGTGAATGG**GATGG**ACTGCG**CTTTACCTGTCCCCCTTCGCCAAGGGGTCT  
TCGGAAACAGAGCCCTGCTCCTTCCCTCCA**TCAAGA**GTTCCCTTTGTACCAGGGCCACATCCACGCCGAA  
CTCTGCTTCATA**TACTG**GTTCCACGACAAAGTCCTGAGAGTGCTGTCCCGATGGAAGAG**TTCAAG**GTAC**GTG**  
**GTAC**ATGTCTGGAGCCCCTGCAGCAAGTGCGCGGAGCAGGTAGCCAGGTTCTGGCCGCACACCGCAACCTAA  
GCCT**GCCAT**CTTCAGCTCCCGCTT**GFACT**ACTACTTAAGGAACCCGAACTACCAGCAGAAGCTCTGCAGGCTG  
**ATTCAG**GAAGGAGTCCACGTGG**CTGCCAT**GGACCTACCAGAAATTTAAAAAGTGTTGGAACAAGTTTGTGGACAA  
TGACGGCCAA**CCATTCAG**GCCTTG**GATGA**GACTGAGAATAAATTTTAGTTTCTATGATTGCAAGCTTCAGGAGA  
TTTTCAGC

23 596 3.9

>A3\_C\_RATTUS

**CGAATGAAT**CTGCTAAGGG**AAGATG**TATTTTACTTTGCAATTTAACAACAGCCACCGGG**TCAAG**CCAGTCCAGAA  
TCGCTACTATCGCAGGAAGTCCATCTGTGTCTACCAACTGGAGCGGGCC**AATGGCAAGAG**CCACTCAAAGGCT  
ACCTGCTATA**CAAGA**AAGGTGAACAGCATGTAGAAATCCTCTTCCCTTGAG**AAGATG**CGGTCCATGGAGCTGAGC  
CAAGTGCGAATTACCTGCTACCTCACCTGGAGCCC**CTGCC**CAAACCTGTGCCCGGCAACTCGCTG**CATTCAA**AAA  
GGATCACCCAGACCTAATTCTGCGGATCTATACCTCCCGCTGTATTT**TACTG**GAGGAAGAAGTTCCAGAAGG  
GGCTGT**GFACT**CTGTGGCGATCAGGGATCCACGTGGACGTCATGGACCTCCCTCAGTTTGT**ACTGCT**TGGACA  
AACTTTGTGAACCCGCAAAGGCCCTTTAG**GCCATGGAATGAA**CTGGAGAAAACAGCTGGCGCATACAAAGGCG  
GCTCCGGAGGATC

21 527 4

>AID\_RATTUS

AGCCTCTT**GATGAA**GCAAAAAGAAGTTTCTTTACCAC**TTCAA**AAATGTCCGCTGGGCTAAGGGTTCGGCACGAGAC  
CTACCTGTGCTATGTGGTGAAGAGGAGAGATAG**TGCCA**CCTCCTTCTCACTGGACTTTGGCCACCTTCGCAACA  
AGTCGGG**CTGCCA**CGTGAATTGTTGTTCTACGCTACATCTCGGACTGGGACCTGGACCCCGCCGGTGTAC  
CGTGTACCTGGTTCACTTCTGGAGCCCCTGCTACGACTGTGCGCGGCACGTGGCTGAGTTTCTG**AGATG**GAA  
CCCTAACCTCAGCCTGAGGATTTTACCAGCGCGCCTTACTTCTGCGAAGACCAGGCTGAGCCTGAGGGGC  
TGCGGAGGCTGCACCGCGCCGGAGTCCAGATCGGGATCATGACC**TTCAA**AGACTATTTTT**TACTGCT**TGGAATACA  
TTTGTAGAAAATC**ATGAA**AGAACT**TTCAA**AGCCTGGGAAGGGCTGC**ATGAA**AACTCCGTCAGGCTAACAGACA  
GCTTCGGCGCATCCTTTTTGCCCTTGT**ATGAA**GT**GATGA**CTT**GAGAGATG**CGTTTCGTATTTTGGGACTT

16 584 2.7

>AID\_FUGU

ATGCTTT**TGCAAGA**AAAGAAGTTCATCTA**CCATTCAAGAATG**TGCGATGGGCACGGGGCCGGCACGAGACCTA  
CCTCTGCTTTGTGGTGAAGAGGCGAGTGGGCCAGACACGCTAACCTTTGACTTTGGACACCTCCGCAATCGCA  
GTGG**CTGCCA**CGTAGAGCTGCTTTTCTGCGTTACCTGGGAGCCTTATGCCCTGGGTTGTGGGGTTACGGAGCC  
GCCGGAGAGAAGAGGCTCAGTTACTCCGTTACCTGGTCTGCTCCTGGTCTCCCTGTGTCA**ACTGCT**CCATCCA  
ACT**CTGCCA**GTTTCTCAATAACACCCCAACCTTCGCCTCAGGATCTTTGTCTCTCGCCTTACTTTTGTGACC  
TGGAGGACAGCCTTGAAAGAGAAGGCTGAGGATGCTGACCAAAGCCGGCGTGAGGATCTCAGT**GATGA**GTAC  
AAAGACTATTTCTATTGCTGGCAGAAATTTGTGGATTGCAAAAAGAGCAACT**TTCAAG**GCCTGGGAAGAGCTGCA  
TCAAAACTCTGTACGCTCACCCGCAAACTCAATCG**CATTC**TG

14 557 2.5

>A2\_FUGU

GAAC**AGATGAGCCATTCTACTTCAAG**TTTCAGTTCAGGAACGTTGAGT**ATTCAT**TCAGGGAGGAA**CAAGACCCT**  
GCTGTGCTTCAGAGTGGATTTCGCCAGGAGGCAGCACCGAGCCACT**CAAGGGTTACATGGAAAGATGAACATGCCA**  
CAGCACATGCAGAGGAGGCCTTCT**TTCAACAAGTGCTGCC**TAATCCATCTCTGGAGTATGACATCAC**ATGGTAT**  
GTATCATCCAGTCCGTGTATCTCTTG**TGCCAACACGCTGGCAAGATGCT**CCAGCAACGCAAAAAAGTCCGCCT  
CTGTGTATTCTGCTCCCGGCTCTTTGAGTGGGACCAGCCAGAAGTAGTAGAG**GCCAT**CCGGACCCTGGTGAGGG  
CCGGATGCAAATAAG**GATGATGAA**GCCCTCTGACTTCCAGCATGTGTGGGAAACTTATGTGGAGAAGGAGGGC  
GAAAGCTTAC**GCCAT**GGGAGGATTGTCAGGACAACACTACAATTATTACAAGGAGATATTAGCTGATATCC**TCAA**  
**G**

26 515 5.1

>A2\*\_FUGU

TTTCAGTTTCAG**TTCAAGAATG**TAGAATACTCTTCGGGACGGAA**CAAGAC**CTTTCTTTGCTACCTGGTGGACAC  
GGGGAAAGCC**GATGAGGGGCTACAGAGGGGATATTTAGAGGATGAGCA****TACTG**GTTTACATGCTGAGGAGGCCT  
TCTTACACAGTGT**CTGCCA**CATTACGACCCCGCC**CAAGTACT**CAGTGACC**TGGTA**TATGTCTTCCAGTCCT  
TGCTGTGCTTGTGCAGCT**AAGATG**GCTGAAGCATTAAAGGCCAGGAGAAACATTAAGTTGAGCATCTTTGCTGC  
TCGTTTATTTGAGTGGGAGGAAGCAGAGATCCAGGCTGGACTGAAAGCTCTACACATTGCTGGCTG**CAAGAT**CA  
GAGT**GATGAA**GCCTCTGGACTTCTCTTACACCTGGGACACTTTTGTAGAG**GAATG**AGGATCAGCCACTGAACCTC  
TGGGCAGATTGCAAAGAAAACCTATGA**GTA**CTACCACGAGAGGTTGGCTGA**CATTC**TACAG

22 500 4.4

>AID\_DANIO

AGTGTGCTCATGACCCAGAAGAA**ATTCAT**CTTCCACTATAA**GAATG**TGCGCTGGGCTCGAGGGAGACACGAAAC  
CTACCTTTGTTTTGTAGTAAAGCGACGCATCGGCCCTGATTCCCTCTCTTTTACTTTGGACACCTGCGCAATC  
GCTCCGGAT**TGCCAT**GTAGAGCTTCTCTTTCTGCGTCACCTGGGTGCGTTGTGTCCGGGCCTGAGCGCTTCTAGT  
GTGGACGGTG**CAAGAT**TGTGTTACTCAGTGACCTGGTTCTGCTCCTGGTCGCCCTGCTCTAAATGTGCTCAACA  
GCTCGCCCACTTCTGTACAGACGCCCAATCTGAGGCTGAGGATCTTTGTGTACGCCT**GTA**CTTCTGT**GATG**  
**AAGAGGACAGCGTGGAGAGAGAAGGTCTGCGACACCTGAAGAGGGCAGGAGTTCAGATCTCGGTCATGACTTAT**  
**AAAGACTTTTCTACTGCTGGCAAACGTTTGTGCGAGGAGGGAGCGGAGTTTAAAGCCTGGGATGGACTTCA**  
**TGAA**AACTCTGTCCGGCTTGTTCGGAAACTCAATCGGATTCTG

12 557 2.2

>A2\_DANIO

GAAGTGCCGATGGAGAACGGAGAGGCCGGAGCTGCTGCTGCAGACGGAGAAAAACCAGAACCTATAGAG**CTGCC**  
**GCCATTC**GAGACCATCACAGGGGATCGCATGGACCCGTTTTTCT**TTCAAG**TTCCAG**TTCAAGA**ACGTGGAG**GTA**CT  
CGTCCGGCCGGAATAAAACCTTCTGTGTTACCTGGTGGACCACGGCGGCGAAGGCCT**GATGAGGGGCTACATT**  
GAGGACGAGCACGCTGGCGGACACGCAGAGGAAGCCTTTTTCCAGCAGATCCTGACCAACTACGACCCGGCCTG  
CAAATACACCATCACCTGGCACTCCTCCAGTCCCTGCGCCA**ACTGC**GCAACCAAGCTGGCGGAGATCCTGCGCT  
CCAGGAAGAACATCCGGCTG**GCCAT**CTTTTCGTGCGGCTGTTTCGAGTGGGAGGAGCCTGAGATCCAGGCAGGG  
CTGAAGCTACTAGCGTCCGTCCGCTGCAA**ACTGC**GCAT**GATGAA**GCCGCTGGATTTACCTACACCTGGGACAC  
GTTTCGTGGAGAGCGACGAACAGCAGTTCACACCCTGGGAGGACTGTCAGGAGAACTACGA**GTA**CTACCAGGACA  
AACTGGCCGACATCCTGCAG

17 608 2.8

>A2\*\_DANIO

**AATGG**GGAGATACCAGAAGAAGCCAAAGGCCAGTCAAAAATGAGGAGTTTGAGCTAGAGCCGATGGAG**CTGCC**  
**GCCAT**TTGAGATCATTGTAGGAGACCGAATAAACCAACCTTCT**TTCAAG**TTCCAATTTAAAAATGTGGAG**GTA**CT  
CGTCCGGGCGCAA**CAAGA**CCTTCTGTGCTACCAGGTCGACATAAAGGTGGAGCAGCAGAGACGGATGGCGTT  
CGTGGTTACCTGGAG**GATGAGCATTCA**GGCTCACATGCCGAAGAGGCCTTCT**TTCAA**CAGGTGCTGGCCTATTA  
TGACAAGTCCCTCCGCTATACAGTGACT**TGGTAC**ACATCGTCCAGCCCTTGTGTGGCCTGCGCTGCTAAGCTTG  
TGGAGATCTTGAAAGCACGCAAGGCGCTGCGCCTTAATATCC**ACTGC**TCCCGTCTCTT**TGAATGG**GAGGAACCG  
GAGATACAGGCCGGACTTCAGGCGCTAGTTAGGGCAGGGTGTAACCTACGCAT**GATGA**GGCCTGTGGATTTCTG

GTATGTCTGGTCTTTTGTGAGAACGAAG**AAGATA**AATTTTACACCCTGGGAAGACTG**TCAAGATA**ACTTTGAGT  
ATTAT**GATGAG**GAGATTGCAGGATATTTTAAAG

24 620 3.9

>A2\_GALLUS

ATGGCAGAAAAGCAGGAGGAGCCAGCAACACCCAGAACGGTGAACCCGACAACGCAGAGGAGGGTGAGGGGAA  
AAAGAAGAAGAAGTTGAAGAGGG**AAGATCTGCC**TCCGTTTGAAATCGTGACAGGGGAACGCCTCCCA**GCCAT**AT  
TT**TTCAAG**TTCCAGTTCAGGAACGTGGAATACAGCTCAGGCAGGAACAAAACCTTCCCTGTGTTATGTTGTAGAG  
ACCCAAGGCAAAGAGTCAAAGACCTCCCGGGTTACCTGG**AAGATGA**GCATGCAGCCTCCCATGCAGAAATTGC  
TTTCTTTAACACGATTT**TGCCA**AAGT**TGAATCAAG**CCTCCGCTACAACATCACCT**TGGTAC**GTCTCCTCCAGCC  
CCTGCGTGACCTGTGCTGATCGGATAAGTGAGACCCTAAGGAAGAA**CAAGA**ACCTGCGTCTGACAATC**ATGGTA**  
GGCCGCCTTTCATGTGGGAAGAGCCAGAAATGCAGG**CTGCC**CTGAAAAAA**ATGAA**GTCTGCTGGCTGCAAGCT  
GAGGATT**ATGAA**GC**TCAAGA**CTTTGAGTACGTCTGGCAGAACTTTGTGGAACAGGAAGAAGGAGAGGAAGCAA  
AGGCTTTCGTGCCGTGGGAGGA**CATTCAAGA**GAACTTCCAATACTATGAGGAAAAGCTGGCTGAGATTCTGCAG

26 662 3.9

>AID\_GALLUS

ATGGACAGCCTCTT**GATGAA**GAGGAAGCTCTTCCCTCTACAAT**TTCAAGA**ACCTGCGCTGGGCCAAAGGCCGTCG  
TGAAACCTACCTCTGTTATGTTGTGAAGCGCCGTGACAGTGCTACATCATGCTCCCTGGACTTTGGATACCTGC  
GTAA**CAAGATGGGT****TGCCAT**GTGGAGGTTCTCTTCCCTACGCTACATCTCAGCTTGGGACCTGGACCCAGGCCGC  
TGCTACCGCATCAC**ATGGT**TCACCTCCTGGAGCCCTGTTATGACTGTGCCGACATGTGGCTGACTTCCCTTCG  
TGCTACCCAACTTGACCCTCCGCATTTT**ACTGCC**CGCTCTACTTCTGTG**AAGAT**CGCAAGGCTGAGCCTG  
AGGGGCTGAGACGCCTGCACCGGGCTGGGGCCAAATC**GCCAT**CATGACT**TTCAAGAT**TTCTTCT**TACTGC**TGG  
AACACGTTTGTGGAGAACAGGGAAAAGA**CATTCAA**GCCTGGGAAGGGCTGC**ATGAA**AACTCTGTCCATCTGTC  
CAGGAACTCCGACGGATCCTT**CTGCCA**CTGT**ATGAA**GT**AGATGA**TTTACG**AGATG**CCTTTAAACTCTGGGAC  
TT

29 590 4.9

>A2\_XENOPUS

ATGGCACAGAGGCAAATA**ATTCA**CAGTCTTCCAAGGATCCTATCAATAACACAGAGAAAGCAGAGAAAGAAGA  
GAAAGCAGAA**AATGG**CGAGGGAAATTCCAAAAAGAAAGAACTAGAGGAGTTGCCTCCCTTTGAAATAGTGGAGG  
GGAGTCGCATCCCTGCTTCTTCTTTCATGTTT**CAGTTT**AAGAACGTGGAGTACAGTTCTGGACGTAA**CAAGACT**  
ATCTGTGCTACACTGTGGAAAGACCTGAGGCTCAAATATTCCATGGATATTTAGAG**GATGAA**CATGTCTCAGC  
TCATGCCGAGGAAGCTTTCTTTACATCTGTC**CTGCC**TCAGTTTCTCACCTCAGGGTCAGTCACTGTAACCTTGCT  
ATGTGTCATCTAGTCCATGCGTTACTTGTG**CTGCC**TCCATAGCTCAATGCCTGCGGAAAAACAAACTGTACGA  
ATACAGCTGGCAGTGGCCGGCTATTCC**AATGG**GAGGAGCCGAGATCCGAAGGGCCCTGAAAGGCCTCCGCTC  
CGCAGGT**TGCCA**AGTCA**GAATGATGA**GGGGTGTGACTATATTTATGTGTGGAAGAACTTTGTTGAGCCTGACA  
TCACCCTG**GATGA**TGAGG**GAATGG**AGGTAGTGAAG**GATGA**GCAT**CAAGA**CTTCATACCCTGGG**AAGAT**TTGGAG  
GAAACTCCAGGTATT**ATGAA**GAGAAGCTGGCAGAGATTCTAAGA

19 707 2.7

>A2\_ORYZIAS\_BJ502566

TCTGCTGCTGCAGTGGAAAACAGTGA**GAATGG**CACACTTCAGTTT**GAGCCC**ATGGAGCTACCCCTTTTGAAGT  
CATCACTGGAGACCGGATCGAC**CCATT**TGTCT**TTCAAG**TTCCAG**TTCA**AAATGTGGAG**GTACT**CGTCCGGCCGCA  
A**CAAGAC**CGCTGCTGTGTTACCTGGTTGA**TACTG**GGGGTGCAGGGGATGGGCTGCTAAGGGGATACCTGGAG**GAT**  
**GAA**CACGCTGGGATCCACGCTGAGGAGG**CATTC**TTACACACTGT**CTGCC**GAACATGACCCCAATGTT**CAGTA**  
TACAATAACC**TGGTAC**GTGTGCTCCAGCCCT**TGCCA**CGCTTGTTCCTTAAAGCTAGCAGAGGTGCTGAAGGCCA  
GGAAAAATG**TCAAG**CTCAGCATCTTTTCTGCTCGCCTGTTTCGAGTGGGAGGCAGAGGAGGTCCAGGCTGGTCTG  
AAGGCCCTGCACAGCGCCGGCTGCAAGCTTAGAGT**GATGAA**GCCTCTGGACTTCTCTTACACCTGGGACACCTT  
TGTGAAACTGAAGAGCAACCGCTGAACCTTTGGGAGG**ACTGCAAGATNA**CTATGAG**GTACT**ACCACGAGAAGC  
TGGACCAGATTCTGCAC

24 605 4



>A1\_MONODELPHIS

**ATGAAT**TCCTAAGACAGGTCCATCAGTAGGAGACGCCACTCTAAGGAGGAGGATCAAACCCTGGGAATTTGTGGC  
CTTCTTTAACCC**TCAAGA**ACTCCGTAAAGAGACATGTCTCCTCT**ATGAA**ATTA**AATGGGG**TAATCAAATATTT  
GGCGACTCCAATCAAATACATCTCAACATGCTGAAATCAACTTCATGGAAAA**ATCACTGC**AGAAAGACAT  
TTTAACTCTTCTGTT**AGATGCTCCATT**ACCTGGTTCCCTCTCTTGGAGCCCTTGCTGGGAGTGTTC**CAAGGCCAT**  
AAGAAAGTTTTTTGGATCATTACCCCAATGTGACTCT**TGCCATT**TTTTATTTCTCGGCTTTATTGGCACATGGATC  
AACAGCATCG**TCAAGGACTCAAG**GAACTTGTCCACTCTGGAGTGACTATCCAGATTATGAGCTACTCAGAGTAT  
CAT**TACTG**TTGGAGAAATTTTGTGACTATCCACAAGGGGAAG**AAGATTACTG**GGCCAAATACCCCTACCTTTG  
GATTATGTTATATGTCTGGAAC**TCCACTGC**ATAATTTTAGGGCTTCCACCTTGT**TTAAGAT**TT**CAGGGAGTC**  
**ATTC**CAATCAGCTTGCCTCTT**CAGCCTTGATTTGCAAGACTGCCATT**ACCAA**AAGAT**TCATATAATGTTCTT  
GTAGCCACAGGGCTGGTCC**AGCCATT**TGTGACCTGGAGA

30 701 4.3

>A1\_ORYCTOLAGUS

ATGGCTTCCGAGAAAGGTCC**TTCAA**ACAAGGATTACACGTTGAGGAGAAGAATTGAACCCTGGGAATTTGAAGT  
CTTCTTTGACCCC**CAAGA**ACTGC**GTA**AGAGGC**CTGCC**TGCTCTACGAA**TCAAG**TGGGGGGCAAGCTC**CAAGA**  
CCTGGCGCAGCTCGGG**CAAGA**ACACCACCAATCACGTGGAAGTCAACTTCCTAGAAAAGCTGACTTCAGAGGGG  
CGGCTTGGCCATCCACCTGCTGCTCCATCACCTGGTCTTGTCTTGGAGCCCTTGCTGG**GAATG**TTCCATGGC  
GATTAGAGAATTTTGGAGTCAACACCCGGGTGTGACTCTGATAATTTTGTAGCTCGGCTATTT**CAGCACATGG**  
ATCGCCGAACCGGCAAGGGC**TCAAG**GACCTCGTTACCAGCGGCGTAACTGTCCGGGTATGAGCGTCTCAGAG  
TATTGT**TACTGC**TGGGAGAATTTTGTCAACTACCCGCCGGGAAAGCAGCTC**AATGG**CCAA**GGTAC**CCACCTCG  
GTGGATGCTGATGATGCACTGGA**ACTTACTGC**AT**CATTC**TAGGTCTCCACCCCTGTTTA**AAGATTTCAAGA**  
GACATCAAAGCAGTTACATTTT**CAGCCTTACTCCTCAATATTGCCATT**ATAAAATGATTCACCATA**CATC**  
CTTTTAGCTACAGGATTGTTACAACCTTCTGTGCCTTGGAGA

23 704 3.3

>A3A\_EQUUS

ATGGAGGCCAGTGCAGCGCCCATGGCCAGACGCCTGATGCCTGAAGACACCTTCACTGAGAACT**TTCAA**CAATGT  
GGTTTGGCAGAAGGCGACCTATCTGTGCTATGAGGTGGAGCTCCCGGATGGCGACTACAGGGTCCC**GCCAGGCC**  
AGGACAAGGGCTTCC**TGCACA**ACCAGGGTGGT**CACATGGCCGGGCCACCCCGCCACGCAGAGATG**CGCTTCC**TG**  
GACCTGATCTCTTCTTGGAACTTGGACCAGGATCTGTGCTACAGGGT**CACCTGTTTCATCTCCTGGAGCCCCTG**  
CGCTGAGTGTGCTCAGAGACTGGCTGAGTTCTTGC**GGGAGA**ACAGCCACGTGAGCCTGCGCATCTT**CGCTTCCC**  
GCATCTACACCAAGGGAGAG**GAATGG**AGACTACAGGGAGGG**ACTGCG**CACCCTGCAGGAGGCTGGGGCCCAAATC  
**GCCAT**CATGACCTCAAAGAGTTT**GAGC**ACT**GC**TGGAA**ACTT**CGTGGACAACCAGGGAAGGACCTTCCAGGC  
CTGG**GATGAG**GCTGGATGCTGAGAGT**CGTACTG**GTCCATGGAGCTGCAGCG**CATTC**TACAGCC**ACTGC**TCCTC  
CCTCC**CTGCC**TCCCCTTACTTTCTCCCTCC**CATCCTGTCCCTGGCCCTTCTTT**CGCTCTCTCAGAGGCTCC  
TCTGGG

12 668 1.8

>A3B\_EQUUS

ATGGAGGCCAGGTGCCTCATG**GATGAA**GACACCTTCACTGAGAACT**TTCAA**CAATGTGAACTGGCCACG**CAAGAC**  
CTACCTGTGCTATGAGGTGGAGCTCCCGGATGGCGACTCCAGGGTCCC**GCCAGGCCGGGACA**AGGGCTTCC**TGT**  
GCAACAAGCCTGTTGATGTGGCTGGG**CCACGCCGACGCAGAGATG**CGCTTCC**TGGACCTGATCTTCTTGG**  
AACCTGGACCAGAAGCTGTGCTACAGGGT**CACCTGTTTCATCTCCTGGAGCCCCTGCGCTGACTGTGCTCAGAG**  
ATTGGCTGAGTTCTTGC**GGGAGA**ACAGCCACCTGAGCCTGCGCATATTCGCTTCCC**GCATCTACACCAAGAGAA**  
AATATGGAGACTACGGGGCGGG**ACTGCG**CACCCTACAGGCGGCTGGGGCCCAAATC**GCCAT**CATGGCCTCAGAA  
GAGTTT**GAGC**ACT**GC**TGGGACACCTT**CGTGGACC**ACCAGGGAAGACCCTTCCAGCC**TGGTAT**GC**CGCTGGATGC**  
**CAAGAG**CCGTTACCTATCTGAGCAGCTGCAGCG**CATTC**TACAGCC**ACTGC**TCCTCCCTCC**CTGCC**TCCCCTCT  
ACTTCTCCCTCC**CTGCC**TCTCCCTTACTTTCTCCCTCC**CATCCTGTCCCTGGCCCTTCTTT**CGCTCTCT

16 662 2.4

>A3A\_BOS

GCCGCGGCTCGAGCAACAA**TC**AAGGGTTCGAGGTTCCAAAACCTACCTGTGCTACGAGGTGGAGAGGCTGGACGG  
AAACGCTACGATCCCTCTGGACGAGTACAAGGGCTTTGTGCGCAACAAGGGTTTGGATCAACCGGAGGAACC**CT**  
**GCCAT**GCAGAGCTCTACTTCCCTGGGA**AAGAT**CCGTTCTTGAATCTGGACCGGAATCAGCACTACAGGCTCACC  
TGTTTCATCTCCTGGAGCCCCTGTTACGATTGTGCCAGAACTGACTA**CATTCT**TGAAGGAGAACCGCCACAT  
AAGCCTGCACATCCTTGCCTCCCGCATCTATACCCGCAA**CCATTTT**GGGT**GCCAT**CAGAGTGGCCTGTGCGAA**C**  
**TGC**AAGCGGCTGGGGCCCGAATCACCATCATGACCTTTGAGGACTTTAAGC**ACTGC**TGGGAAACCTTCGTGGAC  
CACAAGGGAAAACCTTTTCAGCCCTGGGAGGGGCTAAATGTAAAGAGTCAGGCGCTCTGTGCGGA**ACTGCAGGC**  
**CATTCTCAAGACT**CAGCAAAAC

17 536 3.2

>A3\_C\_SUS

AACACA**ATGAAT**CTACTAAAAGAAAATAT**ATTCAA**ACAACAGTTTGGCAACCAGCCCCGGGTCTCGGCGCCCTA  
CTATCTGAGGAAGACCTACTTGTGCTACCAAGTGAAGGGGCC**GATGACTCCATTCT**TGACAAAGGCTGCTTCC  
AAAA**CAAGA**AAAAGCGACATGCAGAAATTCGTTTTATTGA**CAAGAT**CAATTCCT**TGAAT**CTGGACCAGAACCAG  
TGCTACAGAATCATCTGCTATGTCACCTGGAGCCCTTGTCACAACTG**TGCCA**AGGAGCTGGTTGATTTTCATCAG  
CAATCGGCACCACCTGAGCCTGCAGCTCTTTGCCTCCCGCT**GTACT**TCCACTGGGTCAAGTGTATCAGAGGG  
GTCTGCAGCGTCTGCAGGCAAATCGGGTTTCCGTGGCTGTC**ATGAA**AGGCCAGAG**TTCAA**AG**ACTGC**TGGGAA  
AAGTTTGTGACCCCCAGGGGGAGAGCTTCCCATCCTGGGAGAAGCTGGAGCAATACAGCGAGAGCATAAGCCG  
ACGGCTC**TC**AAGGATCCTGCGTTTCGCGAATCAGAATAATTTAGAGGATTCCTTCAGAGATTTGCGACTCGGTT  
CCCCATCACCTTCATCA**TC**AAGAAGTGACTCGAGA

19 623 3

>Sus scrofa cytidine deaminase (CDA), mRNA NCBI Reference Sequence:  
NM\_001244385.1

CACACTGGCTGTGAGGTCTCTTTCCGATCCT**CTGCCA**GCCCCAGGTGGACCAACATGGCC**CAAGAG**CGTC  
CCAGTCCCCTTGGAGCCCGAGCACGTCCAGCAGCTGCTCTTCTC**CTGCC**GGGAGGCCAAGAAGTCCGC  
**CTACTGCC**CCTACAGTCGCTTCCCGGTGGGGCGGCCCTCCTCACCGGGACGGGAGGATTTTCTCCGGG  
TGCAACATAGAAAATGCCTGCTACCCGCTGGGCGTCTGTGCTGAACGGACC**GCCAT**CCAGAAG**GCCAT**CT  
CAGAAGGATACAAAGAGTTTCAGGGCAAT**TGCCAT**CGCCAGTGATCTG**CAAGATGA**TTTTATTT**GCCATG**  
CGGGCCTGCAGGCAAGTCATGAGAGAGTTTGGCACCAACTGGGCCGTCTACATGACCAAGCCGGATGGC  
ACATACGTTGTCAGGACCATCCAGGAGCTGCTAACGGCCTCCTTTGGGCCCGAGGACCTGCAG**AAGAT**CC  
AGTGAAAGGTG

17 497 3.4

>CDA\_FUGU

ATGGCGAACAAAACGAGGCGGCGCAATTATCGGCAGAAACCA**TC**AAGATGCTGATCGACCAATCTCACGAAGT  
GAAAAAGCACTCACACTGTCCTTACAGTAA**ATTCA**GAGTGGGGTCTGCTCTTCTGACCGTCGACAACCGCGTGA  
TTACAGGCTGCAATGTGGGA**GAATG**CATGTTTCTACCTGGGATTTTGTGCTGAAAGAAACGCTATAGCCAAGGCA  
GTGTCAGAAGGGAGCAAAAA**ATTCAAGGCCATT**GCGAT**TGCCA**GTGACATGACTGACCAGTTCATCTCCCCGTG  
TGGTGCCTGCAGGCAGTACATGCGGGAGTTTGGGACAACTGGAGAGTGTTCCTGTCAAACCCGACGGTTCTT  
ACCTGGAGAGAACACTAGAGGACCTGCTTCCCGTTTCTTCCGGGCCG**AAGAT**

13 419 3.1

>CDA\_HUMAN\_978

ATGGCCCAGAAGCGTC**CTGCC**TGCACCCTGAAGCCTGAGTGTGTCCAGCAGCTGCTGGTTTGGCTCCAGGAGGC  
**CAAGA**AGTCAGCC**TA**CT**GCC**CCTACAGTCACTTTCCTGTGGGG**CTGCC**CTGCTCACCCAGGAGGGGAGAAT**CT**  
**TC**AAAGGGTGCAACATAGAAAATGCCTGCTACCCGCTGGGCATCTGTGCTGAACGGACCGCTATCCAGAAGGCC  
GTCTCAGAAG**GGTAC**AAGGATTTTCAGGGCAATTGCTATCGCCAGTGACATG**CAAGATGA**TTTTATCTCTCCATG  
TGGGGCCTGCAGGCAAGTCATGAGAGAGTTTGGCACCAACTGGCCCGTGTACATGACCAAGCCGG**ATGGTAC**GT  
ATATTGTCATGACGGTCCAGGAGCTG**CTGCC**CCTCCTCCTTTGGGCTGAGGACCTGCAGAAGACTCAG

16 434 3.7

>DCMPDA\_FUGU\_Q9BVD8

GTCTCTACCGGGAAAAGAGAAGACTACCTGGAGTGGCCCGAATACTTCATGGCTGTAGCCTTCCTGTCTGCACA  
AAGGAGTA**AAGAT**CCAAGTTCACAGGTGGGAGCATGTATCGTGAAC**CAAGA**GAAAAAATAGTCGGTATTGGTT  
AT**AATGGTA**TGCC**GAATGG**CTGC**GATGA**CGACCTGTTGCCTTGGTGCCGTTCCGCCGAAGACCATCTCGACACC  
AAATATCCTTATGTGTGTACGCTG**AGATGAA****GCCAT****ATGAACAAGA**ACAGTGCCGACGTGAAAGGCTG**CGC**  
**CAT**GTACGTGGCTTTGTTCCCTGCAACGAGTGCGCCAAGCTCATCATCCAGGCAGGTCTGAAAGAGGTGATCT  
ACCTCTCCGATAAGTACCACAACACCCAG**GAATG**ATCGCTTCCAGAAAGTTGCTTACTTGGCTGGG**ATT**CAG  
TACAGG

17 446 3.8

>DCMPDA\_HOMO\_1635

ATGAGTGAAGTTTCCTG**CAAGA**AACGGGACGACTATTTG**GAATGG**CCAGAGTATTTTATGGCTGTGGCCTTCTT  
ATCAGCACAGAGAAGCA**AAGAT**CAAATTCCCAGGTCGGCGCCTGCATCG**TGAATTC**AGAAAA**CAAGAT**TGTCG  
GGATTG**GGTACAAT****GGGA****TGCCA****AATGG**GTGCAGT**GATGA**CGTGTTCCTTGGAGAAGGACAGCAGAGAATAAG  
CTGGACACCAAATACCCGTACGTG**TGCCAT**GCGGAGCT**TGAATGCCAT****CATGAA**CAAAAATTCGACCGATGTGAA  
AGGCTGTAGTATGTATGTGCCTTGTTCCTTGT**ATGAATG**CGCTAAGCTCATCATCCAGGCAGGTATAAAAG  
AAGTGATTTTCACGTCTGATAAATACCATGATAGTGACGAGGCA**ACTGC**TGCGAGGCTCCTGTTTAATATGGCC  
GGGGTGA**CATTC**CGGAA**ATTCA**TACCGAAGTGCAG**CAAGAT**TGTCATTGACTTTG**ATTCAA**TTAACAGCAGACC  
GAGTCAAAGCTTCAG

30 530 5.7

>ADAT2\_HOMO\_NM\_182503

**ATGGT**CTACAACA**ATGAA**GTTGTAGGGAAGGGGAGAA**ATGAA**GTTAACCAAACCAAAAATGCTACTCGACATGC  
AGA**AATGGT****GCCAT**CGATCAGGTCCTCGATTGGTGTCTCAAAGTGG**CAAGA**GTCCCTCTGAAGTATTTGAAC  
ACACTGTGTTGTATGTCACTGTGGAGCCGTGCATTATGTGTGCAGCTGCTCTCCGCCT**GATGAA**AATCCCGCTG  
GTTGTATATGGCTGTCA**GAATGAA**CGATTTGGTGGTTGTGGCTCTGTTCTAAATATTGCCTCTGCTGACCTACC  
AAACACTGGGAGA**CCATT**TCAGTGTATCCCTGGATATCGGGCTGAGGAAGCAGTGGAAATGTTAAAGACCTTCT  
ACAAA**CAAGA**AAATCCAAATGCACCAAAATCGAAAGTTCGGAAAAAG**GAATG**TCAGAAATCT

14 428 3.3

>ADAT2\_FUGU\_Q9CX14

GCCA**AAGATG**CTCTGCAG**GAATGG**AGAAG**TGCCA**GTTGGGTGTCTG**ATGGT**TTACAACAGTGAAGTTGTTGGCAA  
AGGGAG**GAATGAA**GTCAACGAGACGAAAAACGCCACCCGTATGCTG**AGATGGT**GGCCCTG**GATGA**GCTCCTTA  
ACTGGTGTACAGCAGCAACTTGGACGTGAGCAGAGTGAGCAGGAACACGGTCTGTACGTCACCGTTGAACCG  
TGTGTAATGTGTGCTGCTGC**ACTGC**GTCGTTGAACATCCCCACAGTCGTGTACGGCTGCAGGAACGAGCGGTT  
TGGAGGCTGTGGTT**CCATTC**TGGATGTTTCCCTCTGCAGATTTACCTCAGACTGGAAACACCTTTAAG

13 359 3.6

>Homo sapiens cytidine deaminase (CDA), mRNA NCBI Reference Sequence:  
NM\_001785.3

CCGCTGCTCTG**CTGCC****TGCC**CGGGGTACCAACATGGCCAGAAGCGTC**CTGCC**TGCACCCTGAAGCCTG  
AGTGTGTCCAGCAGCTGCTGGTTTGTCTCCAGGAGGC**CAAGA**AGTCAGCC**TACTGCC**CCTACAGTCACTT  
TCCTGTGGGGG**CTGCC**CTGCTCACCCAGGAGGGGAGAATCT**TTCAA**AGGGTGCAACATAGAAAATGCCTGC  
TACCCGCTGGGCATCTGTGCTGAACGGACCGCTATCCAGAAGGCCGTCTCAGAAG**GGTACA**AGGATTTCA  
GGGCAATTGCTATCGCCAGTGACATG**CAAGATGA**TTTTATCTCTCCATGTGGGGCCTGCAGGCAAGTCAT  
GAGAGAGTTTGGCACCAACTGGCCCGTGTACATGACCAAGCCGG**ATGGTAC**GTATATTGTCATGACGGTC  
CAGGAGCTG**CTGCC**CTCCTCCTTTGGGCCTGAGGACCTGCAGAAGACCCAGTGACAGCCAGAG**GAATG**CCC  
**ACTGCC**TGTAACAGCCACCTGGAGAACTTCATA**AAGATG**TCTCACAGCCCTGGGGACAC**CTGCC**CAGTGG  
GCCCCAGCCCTACAGGGACTGGGCA**AAGATGA**TGTTTCCAGATTACACTCCAGCCTGAGTCAGCACCCCT  
CCTAGCAACCTGCCTTGGGACTTAGAACACCGCCGCC**CTGCC**CCACCTTTCCTTTCCTTCTGTGGGC  
CCTCT**TTCAA**AGTCCAGCCTAGTCTGG**ACTGC**TTCCCCATCAGCCTTCCAAGGTTCTATCCTGTTCCGA  
GCAACTTTTCTAATTATAAACATCACAGAACATCCTGGA

30 805 3.7

>Saccharomyces cerevisiae DNA, contig: contig11, whole genome shotgun sequence GenBank: BEGW0100011.1: 423374-423802 cytidine deaminase  
**ATGAA**AGTAGGTGGCATAGAAGACAGACAAC**TGGA**AGCCTTAAAAAGGGCGGCACTGAAGGCGTGTGAAC  
TCTCCTATAGCCCCTATTCTCACTTCCGTGTCGGTTGTTCCATATTAACAAATAACGATGTTATCTTCAC  
TGGTGCTAATGTCGAG**GAATG**CAAGCTACAGTAATTGTATATGTGCAGAACGATCTGCTATGATACAAGTT  
CT**AATGG**CAG**GCCAT**CGCTCAGGATGGAAATGC**ATGGT**CATTTGTGGCGATTCTGAGGACCAGTGTGTTT  
CCCCTTGCGGCGTTTGCAGGCAGTTTATCA**ATGAAT**TTGTGG**TCAAG**GACTTTCCAATAGTCATGCTTAA  
TTCCACAGGTTCTCGTTCTAAAGTCATGAC**AATGG**GAGAA**ACTGCTGCC**CATGGCTTTTGGTCCATCTCAT  
TTAAACTAG

11 425 **2.6**

**Observed AL pentamer frequency: 4.49**

**Expected:  $22/1024 \pm 1.65 \times (22 \times 1002 / 1024^2 \times 100)^{1/2} \approx 2.15\% \pm 0.25\%$ \***

**$4.49\% \approx 2.15\% + 15 \sigma$ , where  $\sigma \approx 0.15\%$**

**$P(\text{Observed} \geq 4.45\%) \leq \exp(-225) \approx (e^{-2.25})^{100} \approx 10^{100}$**

*Caulobacter crescentus* CB15 strain CB15 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075118.1

GACCTGGTGGCTATGCCGGGGTTCCCCACCCGATC**CCATTC**CGAACTCGGTTCGTTAAGTCCCCCTGGGC  
**CAATGGTACT**TTCGTCT**CAAG**GC GCGGGAGAGTAGGTGCGCCGCCAGG

8 112 7.2

*Thermotoga maritima* MSB8 strain MSB8 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_103224.1

ATCCCCGGGTGCCG**TACTG**GGGGCGGAAACACCCGGTT**CCATTC**CGAACCCGGCCGTAAAGCCGCCCTG  
GGCC**ATGGT**AGTATGGGGGCAGCCCCATGCGAGAGTAGGTAGCGCCCGGGGATTTGT

5 124 4

*Helicobacter pylori* 26695 strain 26695 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075103.1

TCCCTTTTCTTGTGCCTTTAGAGAAGAGGAACTACCCAGTTAA**CCATTC**CGAACCTGGAAG**TCAAG**CTC  
TTCATCGCTGATAA**TACTGCT**CTTT**TTCAAG**AGTGG**GAATG**TAGGTTCGGTGCAGGGATAGGGAAAT

9 131 6.9

*Pseudomonas aeruginosa* PAO1 strain PAO1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075116.1

TGCTTGACGATCATAGAGCGTTGGAACCACCTGATCCCTTCCCGAACTCAGAAGTGAAACGACGCATCGC  
CG**ATGGT**AGTGTGGGGTCTCCCCATGTGAGAGTAGGTTCATCG**TCAAG**CTC

3 116 2.6

*Mycobacterium smegmatis* MC2 155 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075427.1

GTTACGGCGGTCCATAGCGGCAGGGAAACGCCCGGTCCCATCCCGAACCCGGAAGCTAAGC**CTGCCA**GCG  
CC**GATGA**TACTACCCTTCCGGGTGGAAAAGTAGGACACCGCCGAACAC

3 114 2.6

*Methanosarcina* sp. 795 genome GenBank: CP011449.1: c670055-669934 5S rRNA

TGTGGCG**GCCAT**AGCGGCAGGGCAACTCCTGTACCCATCCCGAACACAG**AAGAT**AAGC**CTGCC**CACGTTTC  
CTGACT**GTA**CTGAAGTGC GCGAGCCTTCGGGAACTCAGGATCG**CTGCCAT**GC

8 118 6.8

*Escherichia coli* str. K-12 substr. MG1655 strain K-12 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_103249.1

TGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGC  
CG**ATGGT**AGTGTGGGGTCTCCCCATGCGAGAGTAGGGAA**ACTGCCA**GGCAT

5 116 4.3

*Bacillus subtilis* subsp. *subtilis* str. 168 strain 168 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_103286.1

TCTGGT**GATGA**TGGCGAAGAGGTACACCCGTTCCCATGCCGAAACCGGAAGTTAAGCTCTTCAGCGCCG  
**ATGGT**AGTCGGGGTTTCCCCCTGTGAGAGTAGGACATCGCCAGGC

3 112 2.7

*Mycobacterium tuberculosis* H37Rv strain H37Rv 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_103329.1

TTACGGCGGCCACAGCGGCAGGGAAACGCCCGGT**CCATTC**CGAACCCGGAAGCTAAGC**CTGCCA**GCGCC  
**GATGATACTG**CCCTCCGGGTGGAAAAGTAGGACACCGCCGAACA

8 111 7.2

*Salmonella enterica* subsp. *Enterica* serovar *Typhimurium* str. LT2 strain LT2; SGSC 1412; ATCC 700720 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075120

TGCCTGGCGGCACTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCGTAGCGC  
CG**ATGGT**AGTGTGGGGTCTCCCCATGCGAGAGTAGGGAA**ACTGCC**AGGCAT

5 116 4.3

*Pseudomonas putida* KT2440 strain KT2440 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075145.1

TTCTTGACGACCATAGAGCATTGGAACCACCTGATCCCATCCCGAACTCAGTAGTGAAACGATGCATCGC  
CG**ATGGT**AGTGTGGGGTTTTCCCCATGTGAGAGTAGGTTCATCG**TCAAGAT**T

5 116 4.3

*Agrobacterium fabrum* str. C58 strain C58 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075124.1

GACCTGGTGGTTATCGCGGGGCGGCTGCACCCGTTCC**CATTTC**CGAACACGGCCGTGAAACGCCCCAGCGC  
C**AATGGTACT**CCGTC**TCAAGA**CGCGGGAGAGTAGGTTCG**CTGCCA**AGGTCT

11 115 9.6

*Bradyrhizobium japonicum* USDA 110 strain USDA110 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075299.1

GGCCTGGTGGTTTTAGCGAAGAGCCTCAACCCGATCCCATCCCGAACTCGGCCGTTAAACTCTTCAGCGC  
C**AATGGTACT**ATGGCTTAAGCCCTGGGAGAGTAGGTTCG**CTGCCA**AGGCCT

7 115 6.1

*Corynebacterium glutamicum* ATCC 13032 strain ATCC 13032 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075297.1

TGTGTCGGTGGTGATAGTAGCAGGGAAACGCCCGGTCCCTTTCCGAACCCGGAAGCTAAGCCTGGTTACG  
CTG**ATGGTACTGC**ACTCGGGAGGGTGTGGGAGAGTAGGTTACCGCCGACC

6 116 5.2

*Archaeoglobus profundus* DSM 5631 strain DSM 5631 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075715.1

GCCGACG**GCCAT**AGGGGTGCGGGAAACACCCGGACT**CATTTC**CGAACCCGGAAGTTAAGCCCGACCCCGTT  
CCGCG**TGGTACTGT**GTTCCGAGAGGGCACGGGAAGCGCGGAAGCCGTCCGGCA

6 118 5.1

*Burkholderia pseudomallei* K96243 strain K96243 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075304.1

TGCCT**GATGA**CCATAGCGAGTCGGTCCCACCCCTTCCCATCCCGAACAGGACCGTGAAACGACTCCACGC  
C**GATGA**TAGTGCGGATTGCCCGTGTGAAAGTAGGTAATCGTCAGGC

2 112 1.8

*Bacillus cereus* ATCC 14579 strain ATCC 14579 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075158.1

TTTGGT**GATGA**TGGCAGAGAGGTACACCCCGTTCCCATACCGAACACGGAAGTTAAGCTCTCTAGCGCCG  
**ATGGTA**GTTGGGACCTTGTCCCTGTGAGAGTAGGACGTCGCCAAG

3 111 2.7

*Streptococcus pneumoniae* R6 strain R6 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075123.1

TAAGTGACGATAGCCTAGGAGATACACCTGTACCCATGCCGAACACAGAAGTTAAGCCCTAGAACGCCGG  
AAGTAGTTGGGGTTGCCCCCTGTGAGATAGGGAAGTCGCTTAGC

0 0

*Shewanella oneidensis* MR-1 strain MR-1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075142.1

TGCTTGGTGACAATAGCATTGTGGTCCCACCTGATCCCATCCCGAACTCAGAAGTGAAACGCAATCGCGC  
CG**ATGGTA**GTGTGGGGTCTCCCCATGTGAGAGTAGGTCATCGCCAAGCGC

2 116 1.7

*Aster altaicus* chloroplast, complete genome NCBI Reference Sequence: NC\_034996.1: c129302-129182 5S ribosomal RNA

TATTCTGGTGTCTTAGGCGTAGAGGAACACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGA**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Datura stramonium* chloroplast, complete genome NCBI Reference Sequence: NC\_018117.1: c132806-132686 5S ribosomal RNA

TATTCTGGTGTCTTAGGCGTAGAGGAACACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGA**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Rubus takesimensis* chloroplast, complete genome NCBI Reference Sequence: NC\_037991.1: c132382-132262 5S ribosomal RNA

TATTCTGGTGTCTTAGGCGTAGAGGAACACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGA**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Cicuta virosa* chloroplast, complete genome NCBI Reference Sequence: NC\_037711.1: c131571-131451 5S ribosomal RNA

TATTCTGGTGTCTTAGGCGTAGAGGAACACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGA**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Gallus gallus* breed Red Jungle Fowl isolate RJF #256 unplaced genomic scaffold, GRCg6a CHRUN\_242 NCBI Reference Sequence: NW\_020109910.1: c2060-1941 5S ribosomal RNA

AAGCCTACAGCACCCGGTATTCCCAGGAGGTCTCCCATCCAA**GTA**CTAACCAGGCCCGACCCCTGCTTAGC  
TTCCGAGATCAGACGAGATCGGGCGTTACCAGGGTGGGATGGCCGTAGGC

1 116 0.9

*Cricetulus griseus* unplaced genomic scaffold, CriGri\_1.0 scaffold884, whole genome shotgun sequence NCBI Reference Sequence: NW\_003613683.1: c1840240-1840122 5S ribosomal RNA

ATC**TACTGCCAT**ACCACCCTGAACGTGCCGGATCTCATCTGATCTCAGAAGCTAAGCAGGGTTCGGCCCTG  
GCTA**GTA**CTTGGATGGGAGACCACCTGGGAA**TACTG**GGTGCTGTAGGCA

7 115 6.1

*Cephus cinctus* unplaced genomic scaffold, Ccin1 Ccin1\_scaffold0299, whole genome shotgun sequence NCBI Reference Sequence: NW\_014333002.1: 26981-27099 5S

GGCAAGGTCCATACCACGT**TGAAT**TACACCAGTTCTCGTCCGATCACTGAAGTTAAGCAACATTGGGCGCG  
GTC**GTA**CTTGGATGGGTGACCGCTTGGGAACACCGCGTGTCTTGCCT

2 115 1.7

*Saccharomyces cerevisiae* S288C chromosome XII, complete sequence NCBI Reference Sequence: NC\_001144.5 :459676-459796 5S

GGTTGCG**GCCAT**ATCTACCAGAAAGCACCGTTTCCCGTCCGATCAACTGTAGTTAAGCT**TGGTA**AAGACCT  
GACCGAGTAGTGTAGTGGGTGACCATACGCGAAACTCAGGTGCTGCAATCT

2 117 1.7

*Glaucocestis incrassata* strain SAG 229-2 plastid, complete genome NCBI Reference Sequence: NC\_040152.1:c27253-27129 5S

CCTGGTGCCTATGGCGTGGTGGACCCACACCGAAACCATCCCGAACTC**GGTACT**AGAAGTGAAACACTAC  
AGCGGTG**AAGAT**ACTAGAGGGGCAGCTCTCTGGGAAAATAGCTCGG**TGCC**AGGAT

4 121 3.3

*Callitropsis vietnamensis* plastid, complete genome NCBI Reference Sequence: NC\_026298.1: 60600-60720 5S ribosomal RNA

AATTCTGGTGTCTCAGGCGTAGGGGAACCCACACCAATCCATCCCGAACTTGGTGGTTAAA CTT**TACTGCG**  
GTGACGA**TACTGC**AGGGGAAGCCCTGTGGAAAAATAGCTCGA**TGCC**AGAAT

5 117 4.3

*Hesperocyparis glabra* plastid, complete genome NCBI Reference Sequence: NC\_026297.1: 60712-60832 5S ribosomal RNA

AATTCTGGTGTCTCAGGCGTAGGGGAACCCACACCAATCCATCCCGAACTTGGTGGTTAAACTT**TACTGCG**  
GTGACGA**TACTGC**AGGGGAAGCCCTGTGGAAAAATAGCTCGA**TGCC**AGAAT

5 117 4.3

*Cupressus sempervirens* plastid, complete genome NCBI Reference Sequence: NC\_026296.1: 62043-62163 5S ribosomal RNA

AATTCTGGTGTCTCAGGCGTAGGGGAACCCACACCAATCCATCCCGAACTTGGTGGTTAAACTT**TACTGCG**  
GTGACGA**TACTGC**AGGGGAAGCCCTGTGGAAAAATAGCTCGA**TGCC**AGAAT

5 117 4.3

*Callitropsis nootkatensis* plastid, complete genome NCBI Reference Sequence: NC\_026295.1: 60515-60635 5S ribosomal RNA

AATTCTGGTGTCTCAGGCGTAGGGGAACCCACACCAATCCATCCCGAACTTGGTGGTTAAACTT**TACTGCG**  
GTGACGA**TACTGC**AGGGGAAGCCCTGTGGAAAAATAGCTCGA**TGCC**AGAAT

5 117 4.3

*Cavia porcellus* strain inbred line 2N unplaced genomic scaffold, Cavpor3.0 supercont2\_13, whole genome shotgun sequence NCBI Reference Sequence: NT\_176406.1: c36908410-36908290 5S ribosomal RNA

GTCTACG**GCCA**TACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTG  
GTTA**GTA**CTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCTTT

2 117 1.7

*Malus domestica* cultivar Golden Delicious isolate X9273 #13 chromosome 9, ASM211411v1, whole genome shotgun sequence NCBI Reference Sequence: NC\_041797.1: c9964992-9964874 5S ribosomal RNA

**GAATG**CGATCATACCAACTAATATGCACCGGATTCCATCAGAACTCTGAAGTTAAGGGAGTTTGGG**CAAG**  
**ATTAGTACT**AGGATGGGTGACCTCCTAGGAAGTCCTCGTGTTCATCCC

4 115 3.5



Antrostomus carolinensis isolate BGI\_N321 unplaced genomic scaffold, ASM70074v1 scaffold4047, whole genome shotgun sequence  
NCBI Reference Sequence: NW\_010326526.1: c34865-34747 5S ribosomal RNA

GCCTACG**GCCAT**ACCACCCTGAGAACTCCCAATCTCATCTGATCTCGGAAGCTAAGCAGGGTTGGGCCCG  
GTTA**GTA**CTTGGATGGGAGACCGCCTTGGGAATACCAGGTGCTGTAGGCT

2 115 1.7

Ornithorhynchus anatinus isolate Pmale09 chromosome 3, mOrnAna1.p.v1, whole genome shotgun sequence NCBI Reference  
Sequence: NC\_041730.1: c53600548-53600430 5S ribosomal RNA

ATCTGCG**GCCAT**ACCACCCTGAACGTGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGAGTTGGGCCTG  
GTTA**GTA**CTTGGATGGGAG**GAATG**CCTGGGAATACCTGGTGCTGTAGGTT

3 115 2.6

Denticeps clupeioides chromosome 1, fDenClu1.1, whole genome shotgun sequence NCBI Reference Sequence: NC\_041707.1:  
30970836-30970954 5S ribosomal RNA

GCTTACG**GCCAT**ACCACCCTGAGAACGCCCGATCTCGTCCGATCTCGGAAGCTAAGCAGGGTCGGGCCTG  
GTTA**GTA**CTTGGATGGGAGACCGCCTGGGAATACCAGGTGCTGTAAAGCT

2 115 1.7

Carassius auratus strain Wakin unplaced genomic scaffold, ASM336829v1 scaf\_tig00037271, whole genome shotgun sequence NCBI  
Reference Sequence: NW\_020526371.1 :48919-49037 5S ribosomal RNA

GCTTACG**GCCAT**ACCAACCTGGCTATGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGTTTGGGCCTG  
GTTA**GTA**CTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAAAGCT

2 115 1.7

Pteropus vampyrus isolate Shadow unplaced genomic scaffold, Pvam\_2.0 Scaffold66, whole genome shotgun sequence NCBI  
Reference Sequence: NW\_011888847.1: c8268243-8268125 5S ribosomal RNA

GTCTACAG**GCCAT**GGCACCCCTGAACATGCCTGATCTCATCTGATCTCAGAAAGCTAAGCAGGGTCAGGCCTG  
GTTACTACTTGGATGGCAGACCACGTAGGAATACCAGGTGCTGTAGGCA

1 115 0.9

Haloferax volcanii DS2, complete sequence NCBI Reference Sequence: NC\_013967.1: 1603076-1603197 5S ribosomal RNA

TAAGGCGGCCAGAGCGGTGAGGTTCCACCCGTACCCATCCCGAACACGGAAGTTAAGCTCACCTGCGTTC  
TGGTCAG**GTA**CTG**GAGT**GAGCGATCCTCTGGGAAATCCAGTTCGCCGCCCTA

2 118 1.7

Halomicrobium sp. ZP60 chromosome, complete genome NCBI Reference Sequence: NZ\_CP039375.1: 2602863-2602984 5S  
ribosomal RNA

AAAGGCG**GCCAT**AGCGGCAGGGAAACACCCGTACCCATCCCGAACACGGACGTTAAGC**CTGCCA**GCGTTG  
CGGTGA**GTA**CTG**GGGT**TGTGCGAACCCCTGGGAAAGCCGGTTCGCCGC**CTGCC**

6 118 5.1

Haloterrigena jeotgali strain A29 plasmid unnamed6, complete sequence NCBI Reference Sequence: NZ\_CP031304.1: c666418-  
666297 5S ribosomal RNA

TACGGCG**GCCAT**AGCGGCGAGGTGCCTCCCGTACCCATCCCGAACACGG**AAGAT**AAGCTCGCCTGCGTTT  
CGGTCA**GTA**CTG**GAGT**GGGCGACCCCTCTGGGAAATCTGATTTCGCCGCCACCA

4 118 3.4

Halobellus limi strain CGMCC 1.10331 chromosome, complete genome NCBI Reference Sequence: NZ\_CP031311.1: 1536989-  
1537110 5S ribosomal RNA

TAAGGCGGCCAGAGCGGCAGGGCGACACCCGTACCCATCCCGAACACGGAAGTTAAGC**CTGCCA**GCGTGT  
CAGCAA**GTACTG**GAGTGAGCGATCCTCTGGGAACACTGGCTCGCCGCCTCCC

4 118 3.4

*Halobacterium salinarum* strain 91-R6 chromosome, complete genome NCBI Reference Sequence: NZ\_CP038631.1: 75841-75962 5S ribosomal RNA

TAAGGCG**GCCAT**AGCGGTGGGGTTACTCCCCTACCCATCCCGAACACGG**AAGAT**AAGCCCGCCTGCGTTC  
CGGTCA**GTACTG**GAGTGCGCGAGCCTCTGGGAAATCCGGTTCGCCGCCTACT

4 118 3.4

*Natronorubrum bangense* strain JCM 10635 plasmid unnamed4, complete sequence NCBI Reference Sequence: NZ\_CP031309.1: c108047-107926 5S ribosomal RNA

TACGGCGGCCACAGCGGCGGGGTTCTCCCCTACCCATCCCGAACACGG**AAGAT**AAGC**CTGCC**TGCGTTC  
CGGCGA**GTACTG**GAGTGCGCGAGCCTCTGGGAAATCCGGTTTCGCCGCCTCCA

4 118 3.4

*Halapricum salinum* strain CBA1105 chromosome, complete genome NCBI Reference Sequence: NZ\_CP031310.1: 1480608-1480729 5S ribosomal RNA

TAAGGCG**GCCAT**AGCGGCGGGGCAACTCCCCTACCCATCCCGAACACGG**AAGAT**AAGCCCGCCTGCGTTC  
CGGTGA**GTACTG**GAGTGCGCGAGCCTCTGGGAAACTCGGTTTCGCCGCCTACT

4 118 3.4

*Candidatus Nitrosocosmicus franklandus* isolate NFRAN1 genome assembly, chromosome: NFRAN NCBI Reference Sequence: NZ\_LR216287.1 :c2458839-2458721 5S ribosomal RNA

GTAGCC**GCCAT**AGCGTTAGGGCCCGACCCGGTCC**CCATTC**CGAACCCGGAAGTCAAACCTAACGTGCTG  
CTGG**A****GTACTG**ACGTGCGAGAGCCGTTGGGAAACAGCAGTGCTGGCACCT

4 115 3.5

*Halorubrum ezzemoulense* strain Fb21 chromosome, complete genome NCBI Reference Sequence: NZ\_CP034940.1: c1789778-1789657 5S ribosomal RNA

AAAGGCGGCCAGAGCGGTGGGGACACACCCGTAC**CCATTC**CGAACACGGAAGTTAAGCCCACCAGCGTAC  
CGTGAA**GTACTG**GAGTGAGCGATCCTCTGGGAACCACGGGTCGCCGC**CTGCC**

5 118 4.2

*Halorubrum* sp. BOL3-1 chromosome, complete genome NCBI Reference Sequence: NZ\_CP034692.1: 2178713-2178834 5S ribosomal RNA

GAAGGCGGCCAAAGCGGTGGAGACACACCCGTAC**CCATTC**CGAACACGGAAGTTAAGCCCACCAGCGTAT  
CGTGAA**GTACTG**GAGTGAGCGATCCTCTGGGAACCACGAGTCGCCGCCTACC

4 118 3.4

*Sulfodiicoccus acidiphilus* HS-1 DNA, complete genome NCBI Reference Sequence: NZ\_AP018553.1: c2264953-2264834 5S ribosomal RNA

CCAACCG**GCCATGGT**GAACGGGAAACACCCGGACTCATTTTCGAACCCGGAAGTTAAGCCGTTACAGTTCG  
GCT**TGGT**AGTGGGACTCCGTGAGGACCCGCAAC**CTGCC**GGAGCCGGTATGG

4 116 3.5

*Haloarcula marismortui* ATCC 43049 strain ATCC 43049 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075280.1

TTAGGCGGCCACAGCGGTGGGGTTGCCTCCCGTACCCATCCCGAACACGG**AAGATAAGCCCACCAGCGTT**  
CCAGGGAG**TACTG**GAGTGC GCGAGCCTCTGGGAAATCCGGTTCGCCGCCACC

3 118 2.6

*Acorus calamus* plastid, complete genome NCBI Reference Sequence: NC\_007407.1: c130993-130873 5S rRNA

TATTCTGGTGTCTTAGGCGTAGAGGAACCACACCAACCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGAT**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Ammopiptanthus mongolicus* chloroplast, complete genome NCBI Reference Sequence: NC\_034742.1: c129838-129718 5S rRNA

TATTCTGGTGTCTTAGGCGTAGAGGAACCACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGAT**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGGCGCCAGGAT

3 117 2.6

*Ammopiptanthus nanus* chloroplast, complete genome NCBI Reference Sequence: NC\_034743.1: 108087-108207 5S rRNA

TATTCTGGTGTCTTAGGCGTAGAGGAACCACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGAT**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGGCGCCAGGAT

3 117 2.6

*Ammopiptanthus nanus* chloroplast, complete genome NCBI Reference Sequence: NC\_034743.1: 108087-108207 5S rRNA

TATTCTGGTGTCTTAGGCGTAGAGGAACCACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGAT**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGGCGCCAGGAT

3 117 2.6

*Arabis alpina* complete chloroplast genome NCBI Reference Sequence: NC\_023367.1: c129079-128959 5S rRNA

TATTCTGGTGTCTTAGGCGTAGAGGAACAACACCAATCCATCCCGAACTTGGTGGTTAAACTC**TACTGCG**  
GTGACGAT**TACTG**TAGGGGAGGTCTGCGGAAAAATAGCTCGACGCCAGGAT

3 117 2.6

*Methanoregula boonei* 6A8 strain 6A8 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075507.1

GATAGCG**GCCAT**AGCTGTTCGGGAAACACCTGGTCT**CATT**CCGAACCCAGCAGTTAAGCTGACACACGTAG  
TATGCT**GTACTTGGTAC**GCGAGTCCCTGGGAACCATACCAAGCTGCTATCA

4 118 3.4

*Ignicoccus hospitalis* KIN4/I strain KIN4/I 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075517.1

CCCGGCCCGGCCACAGCGACCGGGTAACACCCGGAC**CCATT**TCGAACCCGGAAGTTAAGCCGGTTCGCGCG  
GGGTCCGGCAGTGGCGTCCGAGAGGCGCCGAGCCGCCCGTGTGGGACCGGGC

1 121 0.8

*Thermococcus onnurineus* NA1 strain NA1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075529.1

GGTACGGCGGTCATAGCGCGGGGTACACCCGGTCTCGTTTTGACCCCGGAAGTTAAGCCCGCCAGCGA  
TCCCGGGT**GTACTGCC**CTCCGAGAGGGGGCGGGAAACCGGGAACGCCCGCCGCCAC

4 122 3.3

*Methanococcus aeolicus* Nankai-3 strain Nankai-3 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075496.1

TAACGGTCATAGCGAAGGTGTTATATCCGATC**CCATTC**CGATCTCGGAAA**TCAAG**CCCTTCAGCGATTCC  
CTAA**GTA****CTG****CCATTC**GGTGGAAACAGGGAGACGCTGTTAGT

11 108 10.2

*Methanococcus vannielii* SB strain SB 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075495.1

TAACGGTCAAAGCGGAGGTGTAACATCCGATCCCATCCCGATCTCGGAAATTAAGCCCTCCAGCGATTCT  
TTAA**GTA****CTG**CTATCTAGTGGGAACAAAGTGACGCCGTTAGTCA

2 110 1.8

*Metallosphaera sedula* DSM 5348 strain DSM 5348 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075478.1

TCCACCCG**GCCAT**AGCGGGGAGCAACACCCGGACTCATCTCGAACCCGGAAGTAAAGTCCCCTGCGTTG  
GGGATGCGGTGGGATCCGAAAGGGCCCGCAGCTACCCCAAGCTGGGATGGA

1 117 0.9

*Methanobrevibacter smithii* ATCC 35061 strain ATCC 35061; PS; DSMZ 861 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075474.1

GTTTTGCGGTCATAGCATGGGGTTATACCTGGTCTCGTTTCGATCCCAGTAGTGAAGTCCTTTTGTGTT  
TTGTTTTGT**GTA****CTAT****GGT**TTTTCTATGGGAATTTTCATTTTCGCTGCAAG

2 113 1.8

*Staphylothermus marinus* F1 strain F1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075459.1

CCCG**GCCAT**AGTGGCCGGGCAACACCCGGTCTCATATCGAACCCGGAAGTTAAGCCGGCCACGTTGGGGT  
GGGCCGTGAGGTCCGAGAGGCCCTCGCAGCCGCCCAAGCTGGGAT

1 111 0.9

*Methanoculleus marisnigri* JR1 strain JR1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075455.1

TAGCG**GCCAT**AGCAGAGGGGAAACACCTGGAC**CCATTC**CGAACCCAGCAGTTAAGCCCTCTCACG**TGCCA**  
**TG****TGGT****ACTG**AGGTGCGCGAGCCCTCGGGAAGCACGGCTCGCTGCTATTAC

9 111 8.1

*Pyrobaculum calidifontis* JCM 11548 strain JCM 11548 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075454.1

CCCG**GCCAT**AGGCGCCGGTGCTACACCCGGTCTCATCAGAACCCGGAAGTTAAGCCGGCGCCGCGCTCG  
GG**GTA****CTG**GGCTCCGCGAGGGCCCGGAAACCGGCGTGCTGGGAG

3 112 2.7

*Methanocorpusculum labreanum* Z strain Z 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075452.1

TCATGCG**GCCAT**AGCAGAGAGGTTCCACCCGGAC**CCATTC**CGAACCCGGCAGTTAAGCTCTCTTACGTCTG  
GCTATT**GTA**CTGAGATATGCGAATTCTCGGGAAGTTGTCAACGCTGCACCCC

5 118 4.2

*Pyrobaculum islandicum* DSM 4184 strain DSM 4184 5S ribosomal RNA, complete sequence NCBI Reference Sequence:  
NR\_075435.1

CCCG**GCCAT**AGGCGCCGGTGCTACACCCGGACTCATCAGAACCCGGAAGTTAAGCCGGCGCCGCGGACG  
GGA**GTA**CTGGGCTCCGCGAGGGCCCGGAAACCGCCGTGCT

3 107 2.8

*Rahnella aquatilis* HX2 strain HX2 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075987.1

CCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCGTAGCGCCG  
**ATGGT**AGTGTGGGTCTCCCCATGCGAGAGTAGGGA**ACTGCC**AGGC

5 112 4.5

*Frateuria aurantia* DSM 6220 strain DSM 6220 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075981.1

CCTGGCGGCTATAGCGCTGTGGTCCCACCCGATCCCATCCCGAACTCGGAAGTGAAACGCAGTTGCGCCG  
**ATGGT**AGTGTAGCTTGCTATG**CAAGA**GTAGGTCACCGCCAGGG

3 109 2.8

*Solitalea canadensis* DSM 3403 strain DSM 3403 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075980.1

TCAGGTGCCTATATCGGCGGTGTCCACCTCTTC**CCATTC**CGAACAGAGAAGTTAAGCCCGCCAGAGCCGA  
**TGGTACTGCC**GTAAACAGGTGGGAGAGTAGGTCGGTGCCGATT

9 108 8.3

*Flavobacterium columnare* ATCC 49512 strain ATCC 49512 5S ribosomal RNA, complete sequence NCBI Reference Sequence:  
NR\_075965.1

TAAGGTGGTTATTGCGCGGGGGCTCACCTCTTC**CCATT**TCGAACAGAGAAGTTAAGCCCGCCAGCGTCGA  
**TGGTACTGC**ATTTTTGTGGGAGAGTAGATCGCCGCCTTTC

7 108 6.5

*Desulfovibrio africanus* str. Walvis Bay strain Walvis Bay 5S ribosomal RNA, complete sequence NCBI Reference Sequence:  
NR\_075964.1

CCTGGCG**GCCAT**AGAGGTGGGGCCACACCCGATC**CCATTC**CGAACTCGGCAGTTAAGACCACCCCTCGCCG  
**ATGA**TACTTGGAGTCCACTCCTGGGAAAGTAGGTCGCCGCCAGGT

4 111 3.6

*Sulfobacillus acidophilus* DSM 10332 strain DSM 10332 5S ribosomal RNA, complete sequence NCBI Reference Sequence:  
NR\_075961.1

GTCGGTGCCTTGCGCCGGGTGAGTTGACCCGTTCCCATGCCGAACACGGTCGTCCCCCGCCCGTACGCTG  
AGAC**TACTG**TCCCCTCGGGGACTGGGATCATCAGGCCGTGCCGACC

1 113 0.9

*Niastella koreensis* GR20-10 strain GR20-10 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075960.1

**TATGGT**GATTATACCGAGGGTGTTCACCTCTTC**CCATTC**CGAACAGAGAAGTTAAGCCCCTCATGGCCGA  
**TGGTACT**ACACAACAATGTGGGAGAGTAGGTAGT**TGCCATAT**

9 108 8.3

Owenweeksia hongkongensis DSM 17368 strain DSM 17368 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075957.1

**TATGGT**GACTATAGCGGTGGGGATCACCTCTTC**CCATTC**CGAACAGAGCAGTTAAGCCCACCAGCGCAGA  
**TGGTACTG**GAGTAATATCCGGGAGAGTATGTCGT**TGCCACTT**

9 108 8.3

Streptococcus parasanguinis FW213 strain FW213 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075949.1

TAAGTGACGATAGCCTAGGAGATACACCTGTACCCATGCCGAACACAGAAGTTAAGCCCTAGAACGCCGG  
AAGTAGTTGGGGGTTGCCCCCTGTGAGATAAGGAAGTCGCTTAGC

0

Gordonia polyisoprenivorans VH2 strain VH2 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075948.1

CGGCGGCTATAGCGGTGGGGAAACGCCCGGTC**CCATTC**CGAACCCGGAAGCTAAGCCCACCAGCGCC**AAT**  
**GGTACTGC**ACCCTAATCAGTGTGGGAGAGTAGGACACCGCCGAACA

9 112 8

Paenibacillus terrae HPL-003 strain HPL-003 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075946.1

TTTGGTGGCGATAGCGGAGGGGTTCCACACGTACCCATCCCGAACACGACCGTTAAGCCCTCCAGCGCCG  
**ATGGTACT**TGGACCGCAGGGTCCTGGGAGAGTAGGACGTCGCCAAGC

4 113 3.5

Thermovirga lienii DSM 17291 strain DSM 17291 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075945.1

TCTGGTGGTCATAGCGGAGGGGAAACACCCGGATCCATGCCGAACCCGGCAGTTAAGCCCTCCAGCGCCG  
**ATGGTACTGC**GATGGGTAATCGTGGGAGAGTAGGTGCCCGCCAGAA

6 112 5.4

Exiguobacterium antarcticum B7 strain Eab7 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075943.1

TGGTGACGATAGCCAAGTGGTCACACCCGTTCCCATGCCGAACACGGAAGTTAAGCACTTGAACGCCGAA  
AGTAGTTGGGGGCTTCTCCCTGTGAGGATAGGACGT**TGCCAGGCC**

1 111 0.9

Muricauda ruestringensis DSM 13258 strain DSM 13258 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075936.1

CTAGGT**TGCCAT**TGGCGACGGGGCCACCCCTTT**CCATTC**CGAACAGGGAAGTTAAGCCCGTTTGCGCCGA  
**TGGTACTGCCA**CACCAGGTGGGAGAGTAGGTGCCCGCCTTCT

11 108 10.2

*Acetobacterium woodii* DSM 1030 strain DSM 1030 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075932.1

CGTAATGATGGCGGAAGGGCCACACCTGTTCCCATACCGAACACAGAAGTAAAGTCTTCCAGCGCCGAAA  
**GTACTTGGTGGGTAACTGCTTGGGAGGATAGGACATTGCGGGACT**

2 111 1.8

*Emticicia oligotrophica* DSM 17448 strain DSM 17448 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075929.1

**CATGGT**GTTAATAGGGCGGGTGTTCACCTCTTCCCATCTCGAACAGAGAAGTTAAGCCCCGCACCGCCGA  
**TGGTACT**AATGTCAAATTTGGGAGAGTAGGTAGACGCCAGGA

5 108 4.6

*Turneriella parva* DSM 21527 strain DSM 21527 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075928.1

TCTGTGCGCAATAGAGCAAGGGTCACACCCGTTCCCATCCCGAACACGGAAGTTAAGACTTGCATCGCCA  
**ATGGTACT**CGGTTCTTCGGAGCCCCGGGAGAGTAGGACGCGCACAGGT

5 113 4.4

*Cyclobacterium marinum* DSM 745 strain DSM 745 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075927.1

TTAGGCGGCCCGGCGCAGGGGACCCACCTCTTCCCATCCCGAACAGAGAAGTTAAGCCCTGCAGCGCCGA  
**TGGTACTG**GGGTTACACCTGGGAGAGTAGGTCGCCGCTTCT

5 108 4.6

*Spirochaeta thermophila* DSM 6578 strain DSM 6578 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075923.1

CCTGGTGAICTGAGCGGAGGGGCAACACCCGTTCC**CCATTC**CGAACACGGAAGTTAAGCCCTCCAGCGCCG  
**ATGGTACTGCC**CCATGAGGGGTGGGAGAGTAGGCCGTCGCCAGGC

9 111 8.1

*Runella slithyformis* DSM 19594 strain DSM 19594 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075917.1

**GATGGT**TGGTTACAGGGCGGGTGTTCACCTCTTCC**CCATTC**CGAACAGAGAAGTTAAGCCCCGCTCCGCCGA  
TG**ATACTG**AGATAAACTTGGAAAAGTAGGTAGCCGCCACAA

4 108 3.7

*Flexistipes sinusarabici* DSM 4947 strain DSM 4947 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075916.1

ACTGGTGGCTATGGCGGAGGTGCAACACCCGTAC**CCATTC**CGAACACGGAAGTTAAGACCTCCAGCGCCG  
**AAGAT**ACTCGGGCTAACCAAGCCCCGGAAAAATAGGTTCGCCAGTG

3 114 2.6

*Mycococcus fulvus* HW-1 strain HW-1 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075913.1

TCCGGTGGCTATGTTCGGAGGGGTCCCACCCGTTCCCATCCCGAACACGGAAGTTAAGCCCTCCAGAGCCG  
**ATGGTACT**TTCGCCGAAACCGCGCGGGAGAGTAGGTCG**CTGCC**GGAT

5 113 4.4

*Oligotropha carboxidovorans* OM5 strain OM5 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075911.1

TGGTGGTTCTAGCGAGGAGCCTCAACCCGATC**CCATTCC**GAACTCGGCCGTTAAACTCCTCAGCGCCG**ATGGTACT**ATGGCT**CAAG**CCCTGGGAGAGTAGGTCG**CTGCCA**GGCC

9 110 **8.2**

*Corynebacterium ulcerans* BR-AD22 strain BR-AD22 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075906.1

TGGTGGTTATAGCGTCGGGGGTACGCCCGGTCCCTTTCCGAACCCGGAAGCTAAGCCCCGATTGCGCTG**ATGGTACTGC**ACCTGGGAGGGTGTGGGAGAGTAGGTCGCCGCCAACCA

6 114 **5.3**

*Lacinutrix* sp. 5H-3-7-4 strain 5H-3-7-4 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075910.1

TAAGGTGATTATAGCGATAGGGCTCACCTCTTA**CCATTCC**GAAACAGAGAAGTTAAGCCTATTAGCGCCG**ATGGTACT**TACATTTGTGGGAGAGTAGGTCGTCGCCTTTT

6 104 **5.8**

*Treponema brennaborensis* DSM 12168 strain DSM 12168 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075894.1

CCTGGT**GCCAT**AGTGGAGAGGTAATACCCGTTCCCATCCCGAACACGGAAG**TCAAG**CTCTCTTACGCC**GATGATACTG**CTATTGCGGGAAAGTAGGTTG**CTGCCA**GGC

7 105 **6.7**

*Thermodesulfobium narugense* DSM 14796 strain DSM 14796 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075892.1

CCCGGTGACCATAGCAAAGGTGCCCCACCCGTT**CCATTCC**GAAACACGGAAGTGAAAGCCTTTAGCGCCGACGATACTATTGGGGAAAATAGGACGTTGCCGGGA

2 101 **2**

*Porphyromonas asaccharolytica* DSM 20707 strain DSM 20707 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075891.1

TAAGGTGGTTATAGCGTTGGGGATCCACCTCTT**CCATTCC**GAAACAGAGCAGTTAAGCCCAATAGCGCCG**ATGGTACTG**AGTAAAACCGGGAGAGTAGGTAACCGCCTTTC

7 107 **6.5**

*Thermodesulfator indicus* DSM 15286 strain DSM 15286 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075890.1

CCGGGTGCCCATACCGGAGGGGCCACACCCGTT**CCATTCC**GAAACACGGTCGTTAAGCCCTCCAGGGCCG**ATGGTACTGC**TGGGTTACCCCGGTGGGAGAGTAGGTCGGCGCCCGGG

8 113 **7.1**

*Sphaerochaeta coccoides* DSM 17374 strain DSM 17374 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075885.1

CCCGGTGGTCATAGCAGAGTGGTCATACCCGTTCCCATCCCGAACACGGCAG**TCAAG**CACTCTCGCGCCG**ATGGTACTGC**GTTTTTCGCGGGAGAGTAGGTCGCCCGCCGGC



7 108 6.5

Verrucosipora maris AB-18-032 strain AB-18-032 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075884.1

TACGGCGGTTATGGCGGAAGGGAAACGCCCGGTTA**CATTC**CGAACCCGGAAGCTAAGCCTTCCAGCGCCG  
**ATGGTACTGC**ACTCGGGAGGGTGTGGGAGAGTAGGACACCCGCGGAC

7 113 6.2

Selenomonas sputigena ATCC 35185 strain ATCC 35185 5S ribosomal RNA, complete sequence NCBI Reference Sequence: NR\_075883.1

TCCAGTGGCGATAGCTGAGTGGTTCCACCTGTTCCCATCCCGAACACAGTCGTTAAGCACTCATACGTCTG  
AAA**GTACT**TGGTTGGAGACGACCTGGGAGGATAGAAAGCTGCTGGTT

1 113 8.9

**Observed AL pentamer frequency: 401.2 /100  $\approx$  4%**

**Expected: 22/1024  $\pm$  1.65x(22x1002/1024<sup>2</sup>x100)<sup>1/2</sup>  $\approx$  2.15%  $\pm$  0.25%\***

**4%  $\approx$  2.15% + 12  $\sigma$ , where  $\sigma \approx$  0.15%**

**P(Observed  $\geq$  4%)  $\leq$  exp(-144)  $\approx$  (e<sup>-1.44</sup>)<sup>100</sup>  $\approx$  3 10<sup>-63</sup>**

## AL-pentamer proximity for 100 SASP proteins

*Clostridioides difficile* strain 630 delta erm, complete genome GenBank: CP016318.1: 1503870-1504070 SASP

ATGGATGATGTTTCAAGACAAAATGCTATAAAAAGCTTTAAAAACAAACAAAATGGAAAATAGCAGGAGAGT  
ATGGAATGAATTATGAAGATGCTTTTGAATAATAGAAAATGCGTCTAATAAAGGTGTTTTAGAAGGTTA  
CTTTAAAAAGCTTGAAAAGAAAAAATTTAGGTCAAGGTATATCTAGGCATTTAGAATAA

13 197 6.6

*Pseudoflavonifractor capillosus* ATCC 29799 B\_capillosus-2.0.1\_Cont39, whole genome shotgun sequence GenBank: AAXG0200007.1: c199825-199622 SASP

TGGAGGCCGCTTCCGAGGTTGGCGTCAACCTGAAGCAGGGCTACAACGGCGACCTGACCTCCCGTGAGGC  
CGGCTCCGTGGGCGGCCAGATGGTCAAGAAGATGATCCAGGCCTACGAGAACGGAATGCAGTAA

8 130 6.2

*Eubacterium ventriosum* ATCC 27560 E\_ventriosum-MSIQ\_Cont10.2, whole genome shotgun sequence GenBank: AAVL02000028.1: 64678-64878 Small, acid-soluble spore protein, alpha/beta type SASP

ATGAGTAACACTAACAAAGTTAATGTTTCTGAAGCAAAGGGTGCTATGAAACAGATTCAAATGGAAATGTG  
CATCAGAGGTAGGCGTTTACAGCTTACAGATGGTTATAACGAAACTTAAACATCAGCCAGGCAGGTTCTGT  
GGGTGGACAGATGGTTAAGAAGATGATTGAAGCTCAGGAAAGACAGATGACGGTAAATAA

14 197 8.2

*Dorea longicatena* DSM 13814 D\_longicatena-MSIQ\_Cont14.1, whole genome shotgun sequence GenBank: AAXB02000016.1: 61524-61730 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCAAATCGTTCATCAAACAAAGCAGCCGTACCTGAGGCAAAGGGTGCATTAGACAAAATCAAGTATG  
AGGTAGCAAGCGAAGCTTGGAGTTCCGTTATCAGACGGCTACAACGGAGACCTGACATCAAGACAAAACGG  
ATCTGTCCGGCGTTATATGGTTAAGAAAATGATTGAACAGCAGGAAAAACAGATGCAGGCAAAATAG

7 203 3.5

*Intestinibacter bartlettii* DSM 16795 C\_bartlettii-2.0.1\_Cont622, whole genome shotgun sequence GenBank: ABEZ02000022.1: 280387-280689 Small, acid-soluble spore protein, beta type SASP

TTATTGTTGTCCTTTTACCTGCCATTTGTTGTTCTGCCATTTCGACTAATTTTTTTAGTCATGTATCCACCA  
ACATAAACCATTTTGTCTAGCTGTTAAGTTTCTTTTATCAACTTGTGATAATCAGCCATACCTAGTTTCAT  
TAGCTATTTCTAATTTTCAATTTGATTTAATGCTTGTGTTAGCTTGTGGAAGTCTGTTTTGTTATTGTTGTT  
GTTTGACATATCTAATACCTCTTTTCTATGATTTTTGTTTCTGGTAGTATTATATGAAATATTTGTTTT  
ATAATACATACAATTTGTGACAA

13 299 4.4

[*Eubacterium*] *siraenum* DSM 15702 E\_siraenum-3.0.1\_Cont322, whole genome shotgun sequence GenBank: ABCA03000045.1: c37855-37667 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCTAGTACCAATAAAAAGAACAAGGCAGCATTAAACAGCATCAAGATGCAGGCTGCAAGCGAAGTAG  
GCGTACCCCTTAAGGATAACGGCTACAACGGCGACCTTACTTCAAAGCAGGCTGGTTCAAATCGGCGGTCA  
GATGGTCAAGAAGATGATCGAATCCTACGAAATGTCAAACAAGCAGTAA

13 185 7

[*Eubacterium*] *hallii* DSM 3353 E\_hallii-1.0\_Cont439.1, whole genome shotgun sequence GenBank: ACEP01000143.1: 6905-7126 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCAAGTAATAAAAATTTCTTCATCTAACAGAGTAGAAGTACCGGAAGCAAGAGACGCAATGGACCGTT  
TTAAATGGAAAGTAGCCAACGAAGTTGGTGTGAATTTAAAACAGGGATATAACGGAGATATTAATGCAAG  
AGATGCCGGCTCCATCGGAGGAGAATGGTTCGTAAAGATGATCAAGAGACAGGAAGAGGAATGGCTGGT  
CAGAGCCGCTAA

16 218 7.3

[Clostridium] scindens ATCC 35704 C\_scindens-2.0.1\_Cont260, whole genome shotgun sequence GenBank: ABFY02000013.1: 132098-132310 Small, acid-soluble spore protein, alpha/beta type SASP

ATGAC**AATGG**CAAGTCGTTTCATCAAACAGAGCAGCCGTACCTGAAGCAAAGGGTGCATTGGACAA**ATTCA**  
**AATAT**GAGGTTGCAAACGAATTAGGAGTACCGTTAACAGACGGATACAACGGAGACCTGACATCTAGACA  
**GAATGGT**TCTGTGCGCGGATAT**ATGGT**TAAGAAAATGATCGAACAG**CAAGA**GAAAC**AGATG**GCTGGCAAATAA

9 199 **4.5**

Coprococcus eutactus ATCC 27759 C\_eutactus-2.0.1\_Cont262, whole genome shotgun sequence GenBank: ABEY02000018.1: c65614-65408 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCAAG**TTCAA**ACTCTAACAAAGGCAAGTGTCCAGAGGCAAAGAGTGCT**ATGAA**CAAG**TTCAAGATGG**  
AGGTAGCTAAGGAGATCGGTGTAGATCTTAAGGACGGCTAC**AATGG**AGATTTAACTTCAGCTCAGTCTGG  
ATCGATCGGAGAG**AGATGGT**TCGTC**AGATGATCAAGA**GACAGGAAGAGC**AGATG**TCAGGCAGATAA

15 203 **7.4**

Blautia obeum ATCC 29174 R\_obeam-MSIQ\_Cont262, whole genome shotgun sequence GenBank: AAVO02000003.1: 172483-172710 Small, acid-soluble spore protein, alpha/beta type SASP

ATGACAAAAAACAGTTCTTCTAACAGAGCAGTTGTACCTGAAGCCAAAGGAGCCCTGGATCGTTTTAAAT  
TCGAAGTTGCAA**ATGAA**CTGGGCGTTCCGCTTACCGATGGCTACAACGGAAACCTGACATCCAAACAGAA  
CGGAAGTGTGCGGCGGATAT**ATGGT**CAAAAAAATGATTGAACAGCAGGAAAGAC**AGATG**TCCGGCAATCT  
C**GAATGAA**CA**GTACTGA**

7 224 **3.1**

Blautia hydrogenotrophica DSM 10507 R\_hydrogenotrophicus-1.0\_Cont1.23, whole genome shotgun sequence GenBank: ACBZ01000017.1: 52449-52661 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCTAGAAATTC**TTCAA**ATTCCGCAGCAGTGCCGGAAGCAAAGAGCGCATTAGACCGTTTTAAATTTG  
AGGTTGCAAACGAGTTGGGAGTACCTCTGACTGACGGTTATAACGGAAACCTGACTTCCAAACAGAACGG  
AAGCGTAGGCGGTTAT**ATGGT**TAAG**AAGATGAT**TGAAGCCAGGAAAGAC**AGATGA**CTGGCGGACA**GTACTAA**

8 209 **3.8**

[Ruminococcus] torques ATCC 27756 Ruminococcus\_torques\_2.0\_Cont64, whole genome shotgun sequence GenBank: AAVP02000003.1: 123659-123898 Small, acid-soluble spore protein, alpha/beta type SASP

ATGTCGTTAAATGCAAAGGAGGAAATTGATCATGGCAAATCGTTTCATCTAACAGAGCAGCAGTACCGG  
AGGCAAAGGGCGCTCTGGACAA**ATTCAAGTATGAA**TGGCAAAGTGAGTTAGGAGTACCGTTGACAGACGG  
TTATAACGGAGACCTCACATCGAAACAGAACGGATCTGTTGGCGGATAT**ATGGTCAAGA**AAATGATCGAG  
CAGCAGGAAAAAC**AGATGGC**AGGCAAATAA

8 236 **3.4**

Flavonifractor plautii ATCC 29863 F\_plautiiATCC29863-1.0\_Cont510.3, whole genome shotgun sequence GenBank: AGCK01000277.1: 763-972 small, acid-soluble spore protein C2 SASP

ATGTCTAGCTCTAACAAATTTCCAATTTCCAACCAGCCCGTGGTTCCAGCGCCCGTGAGGCCCTGAACAAG**T**  
**TCAAGATG**GAGGCCCGCCAGGAGGTTGGCGTCAACCTGAAGCAGGGCTAC**AATGG**CGATCTGACCAGCCG  
TGAGGCCGTTCCGTTGGCGGCC**AGATGGTCAAGAAGATGA**TCCAGGCGTACGAGAACAGCATGCAGTAA

13 207 **6.3**

Desulfitobacterium hafniense DP7 D\_hafnienseDSM13498-1.0\_Cont144.1, whole genome shotgun sequence GenBank: AFZX01000039.1: c7693-7475 Small, acid-soluble spore protein, alpha/beta type SASP

ATGAGCAATAAAACGCC**AATGGT**TCC**TGAAT**CCGAAAACAAGTTGGATCAACTTA**AATGG**GAAGTTGCTG  
AGGA**ACTCCATT**TAGACGACGACATCCAGGAAAGAGGTTTTGAAAACATGACCACCCGGGAAGTCGGACA  
AATCGGCGGCAAT**ATGGT**CAAAAA**AATGGT**GGCATTTGCCGAAGAGC**AGATG**GGTAAAGGGGCCGACAT  
AAGGATTA

9 215 **4.2**

*Clostridium celatum* DSM 1785 C\_celatum-1.0\_Cont521.1, whole genome shotgun sequence GenBank: AMEZ01000133.1: c61494-61312 Small, acid-soluble spore protein, alpha/beta type SASP

ATGTCAAAGCATTAGTTCCAGAAGCAAAACAAGGATTAAGCG**CATTCAAGA**ACGAAGTAGCAGCGGAAA  
**TGGGAGTACCATTACAGATTACAATGGTA**ACCTAACATCTAGACAATGCGGTTCCGTAGGTGGTGA**AAAT**  
**GGT**TAAGAG**AAATGGTA**GAAACAATACGAAAGTGGATTAAAGTAA

18 179 10

*Marvinbryantia formatexigens* DSM 14469 B\_formatexigens-1.0.1\_Cont7.1, whole genome shotgun sequence GenBank: ACCL02000008.1: 188424-188837 Small, acid-soluble spore protein, alpha/beta type SASP

TTTACATTA**ACTGCCA**GTTTATATTCCGGAGAATTTTCATATACTATC**ACTGC**AAATCGAAAAACACACAAA  
AAAACAAACGAA**AGAT**TGCAGATAAAATGCAAAGGAGGCGTAAATTTTATG**TCAAG**TAAATCTTCTAAC  
ACAGCAGCAGTACCGGAGGCGAAGGGAGCACTGGACCGT**TTCAA**ATTTGAAGTTGCGAGCGAGCTCGGCG

6 206 2.9

[*Bacteroides*] *pectinophilus* ATCC 43243 B\_pectinophilus-1.0.1\_Cont749, whole genome shotgun sequence GenBank: ABVQ01000035.1: c417970-417746 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGAGGTGGACAGT**ATGAACAGTTCAA**ATTCTAATTCTAAGACAAAGAAACAGGCACAGGATGCTCTTG  
ACAGATTT**AGATG**GAAAGTTGCATCAGAGATTGGTGTACCTCTTAAGGAAGGATACAACGGAGATCTTAC  
AACAAGCCAGGCCGGTTCAGTCGGTGGT**AGATG**TAAGA**AGATGA**TTAAGAGACAGGAAGAGC**AGATG**  
TCTTCAGGCAACTAG

11 221 5

*Ruminococcus lactaris* ATCC 29176 R\_lactaris-2.0.1\_Cont378, whole genome shotgun sequence GenBank: ABOU02000017.1: 17538-17744 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGCAAGTCGTTTCATCTAACAAAGCAGCAGTGCCTGAAGCAAAAGGCGCACTCGACAAG**TTCAAG**TACG  
AGGTAGCAACAGAGCTGGGAGTACCGTTGACAGACGGATACAACGGCGACCTGACATCCAAACAGAACGG  
ATCTGTCCGGTGGATAT**ATGGTA**AGAAAATGATCGAGCAGCAGGAAAAGC**AGATG**GCAGGAAAATAA

5 203 2.5

*Coprococcus comes* ATCC 27758 C\_comes-1.0.1\_Cont2253, whole genome shotgun sequence GenBank: ABVR01000041.1: 516733-516945 Small, acid-soluble spore protein, alpha/beta type SASP

**ATGAAT**ATGGCAAGTCGTTTCATCAACAGAGCAGCCG**TGCCA**GAAAGCAAAAGGTGCACTGGACAAATTTA  
AATACGAGGTTGCAAGCGAACTGGGAGTACCGTTATCAGACG**GGTACA**ACGGAGACTTAACT**TTCAAGACA**  
GAACGGATCTGTCCGAGGTTAT**ATGGTCAAGA**AAATGATCGAAGAACAGGAAAAAC**AGATG**TCCGGTTCGATAG

11 209 5.3

*Clostridium sporogenes* ATCC 15579 C\_sporogenes-2.0.1\_Cont478, whole genome shotgun sequence GenBank: ABKW02000003.1: 1295759-1295986 Small, acid-soluble spore protein, alpha/beta type SASP

**ATGAAAAATAAACAGATGAT**AGAAGAGATAATGTAGATAAA**ATTCAA**TATAATATTACTAAGACT**ATTC**  
**AAA**ACTGTGAGCTTGC**AGATGAA**ATGATTGCAAAAAC**AGATGATGAA**AAAAACGAAAAAACTCTAATAGA  
AAAAAACCAAAGAAGAAGAGAAGCACTTG**ATGGTAT**GAGAGAAGAAATTA**AGATGAAGCAAGAGATAAG**  
AAA**ATGG**ATATATGT

22 222 10

*Clostridium* sp. M62/1 C\_sp\_M621-1.0.1\_Cont5.1, whole genome shotgun sequence GenBank: ACFX02000031.1: 96192-96377 Small, acid-soluble spore protein, alpha/beta type SASP

**ATGAATG**TACCAGAGGCTAAGGAAGCA**ATGAACAGATTCAAGATG**GAGGTT**TGCCA**ATGAGCTTGGTGTTC  
CGTTAAACAGCAACGGATACAACGGCAACTTAACTTCTGCTCAGAACGGTCTGTAGGCGGATAT**ATGGT**  
TAAG**AGATGA**TCGAGGCTCAGGAAAGAC**AGATG**GCCGGCAAATAA

16 182 8.8

Anaerofustis stercorihominis DSM 17244 A\_stercorihominis-2.0.1\_Cont258, whole genome shotgun sequence GenBank: ABIL02000005.1: 428678-428869 Small, acid-soluble spore protein, alpha/beta type SASP

TTATTTTCAT**TGAAT**TTTTTCGT**AAGAT****CAACCATT**CTTTTAA**CCATT**TCACCACCGATAGAACCGGCTTCT  
TTACTAGTTAAATTA**CCATT**TATATCCATCTT**AAGAT**TAATGCTAAGTTTCGCTAGCAACTTGTTCCTTAA  
ATTGGTCTAAGCCTTTTTTAGCTTCTCTGTTGTTTCATCTGTTTCACCTCCAT

9 188 4.8

Clostridium sp. SS21/ Clostridium\_sp\_SS21-1.0.1\_Cont14.1, whole genome shotgun sequence GenBank: ABGC03000004.1: c56363-56142 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGAAGAACGAACCTTCAG**CAAGA**AAGCAAAGGAACAAACAGAGTCGAG**GGTAC**CAGAAGCGAGAGGACCT  
TAGACAGATTTAAGTTTCGAAGTTGCAAATGAGATCGGTGTGGATTTAAAAGAAGGATATAACGGACACTT  
AACAA**ACTGCT**CAGGCAGGATCTGTCCGAGGAG**AGATGGT**TCGTAAAATGATCCGCAGACAGGAAGAAGAA  
ATGTCTCGCTAA

5 218 2.3

Anaerotruncus colihominis DSM 17241 A\_colihominis-0.2.1\_Cont303, whole genome shotgun sequence GenBank: ABGD02000027.1: 56025-56219 Small, acid-soluble spore protein, alpha/beta type SASP

ATGACCAGTTCCAACAATCCTGTTGTCCCGCAGGCCCGCGAGGCTCTTGATAAG**TTCAA****AATGG**AAGCCG  
CCCGTGAAGTCGGCGTCAAC**CAAG**CAGGGCTACAACGGCGATCTGACCTCTCGTGAAGCCGGTTCCTGT  
CGGCGGCC**AGATGGTCAAGA**AGATGATCGAGGCGTACGAGAACTCC**ATGAA**GTAA

11 191 5.8

Anaerostipes caccae DSM 14662 A\_caccae-4.0.1\_Cont13, whole genome shotgun sequence GenBank: ABAX03000002.1: c118067-117843 Small, acid-soluble spore protein, alpha/beta type SASP

ATGGAAGAAC**GTA**CTTCAGCAAACAGCAAAGGAAACAACAGAGTGGAAGTTCCAGAGGCTAGGGGAAGCAC  
TGGACAG**ATTCAA**ATTCGAGGTTGCTA**ATGAA**ATCGGTGTAGACTTAAAAGAAGGATACAACGGACATTT  
AACAA**ACTGCT**CAGGCAGGTTCTGTTGGAGGAGAA**AATGGT**TCGTAAAATGATCCGCAGACAGGAAGAGGAA  
ATGTCTGGCCGTTAG

7 221 3.2

Pseudobacteroides cellulosolvens ATCC 35603 = DSM 2933 ctg1, whole genome shotgun sequence GenBank: LGTC01000001.1: 124072-1240946 Small, acid-soluble spore protein, alpha/beta type SASP

TTACTCAGAAATCATACTTT**CTGCCATT****TGAATCATT**CTTTTACCATCTGACCGCCTATAGGGCCGCCT  
TCCAAACCGTTTACTTTGGCCGGAACATCTCCCTTGTAGTGGTCATTGTTTCCCTTTATATATTGAAGGT  
GTCCAATTTCTTGTGCACACTCTATTT**TGAAT**CTTGTAAGTGCTTCTCTGCTTTCAGGTACCAGTGGTFT  
TCTTGGATTTGACAT

7 221 3.2

Bacillus mycoides strain AH603 chromosome, whole genome shotgun sequence GenBank: CM000737.1: 444037-444333 Small, acid-soluble spore protein gamma-type SASP

GTGGAGGTGAGA**AAGAT**GAGTAAAAAACAACAAGGTTATAATAAGGCAACTTCTGGTGCTAG**CATTCAA**  
GTACAAATGCTAGCT**ATGGTAC**AGAGTTTGCAACTGAAAC**GAATG**TACAAGCAGTAAAAACAAGCGAACGC  
ACAATCAGAGGCACAGAAAGCACAAGCTTCTGCTGCTCAAAGTGCAAATGCTAGTT**ATGGTAC**AGAGTTT  
GCAACTGAAAC**AGATG**TGCATGCAGTGAAAAAACAATGCACAATCAGCTGCAAAAACAATCACAATCTT  
CTAG**TTCAA**ACGAGTAA

14 293 4.8

*Bacillus thuringiensis* serovar *kurstaki* str. T03a001 chromosome, whole genome shotgun sequence GenBank: CM000751.1: c349955-349659 Small, acid-soluble spore protein gamma-type SASP

GTGGAGGTGAGAAAGATGAGTAAAAACAACAAGGTTATAACAAAGCAACTTCTGGTGCTAGCATTCAA  
GTACAAATGCTAGTTATGGTACAGAGTTTCAACTGAAACAGATGTACAAGCTGTAAAAACAAGCAAACGC  
ACAATCAGAAGCAAAGAAAGCACAAGCTTCTGGTGCACAAAGTGCAAACGCTAGTTATGGTACAGAATTT  
GCAACTGAAACAGACGTGCATTCTGTGAAAAACAATAATGCTAAGTCAGCTGCAAACAATCACAATCTT  
CTAGCTCAAATCAGTAA

14 293 4.8

*Clostridium botulinum* A str. ATCC 3502 complete genome GenBank: AM412317.1: 1907279-1907482 small, acid-soluble protein alpha SASP

ATGCCAAGTAATAAAAAATTCTAATAATTTAGTAGTGCCAGAAGCACAACAAGGATTAACCAACTAAAA  
TGGAAAGTTGCAAACGAAGTAGGAATAGCAAACATGATTCATGGATAAAGGAACTTAACATCAAGACA  
AATGGATATGTAGGCGGAAACATGGTTAAAAAGATGGTAGAAGCTTACGAAAGAAATCTATAA

15 200 7.5

*Clostridium botulinum* A str. ATCC 3502 complete genome GenBank: AM412317.1: 1907543-1907743 small, acid-soluble spore protein beta SASP

ATGGCAAATAGAAATTCTAATAATTTAGTAGTGCCAGAGGCAAAACATGGATTAAATCAATTAATAATGG  
AAGTTGCAAACGAAGTAGGAATAGCAAACATGATGCTGTGGATAAAGGAAATTTAACATCAAGACAAA  
CGGATATGTAGGCGGAAATATGGTTAGAAAGATGGTAGAAGCTTACGAAAGAAATCTATAA

10 197 5.1

*Clostridium saccharoperbutylacetonicum* NI-4(HMT), complete genome GenBank: CP004121.1: 3458976-3459185 small, acid-soluble spore proteins, alpha/beta type SASP

ATGTCATCAACAATAGTGAAGAAACAGAACATTAGTACCAGAAGCAAAGCAGGGATTAAGCAGATTAA  
AAACTGAGGTTGCTTCAGAAGTTGGTTTTAAGTGACTATGAAAAATCAAGATAAAGGTAACCTTTTCATCAAG  
ACAAATGGTAGCGTTGGCGGTTACATGGTAAACACATGATCGAAAGCTACGAACAAGGCTTAAAGTAA

13 206 6.3

*Bacillus subtilis* subsp. *subtilis* strain SRCM101441 chromosome, complete genome GenBank: CP021507.1: c3351111-3350971 Small, acid-soluble spore protein J SASP

ATGGGTTTTCTTTAATAAAGATAAAGGAAAACGTTCCGAAAAAGAAAAAACGTAATCCAAGGAGCTCTTG  
AAGATGCTGGTTTCAGCTCTAAAGATGATCCGCTTCAAGAAGCTGTGCAAAAAAGAAAAATAATCGATAA

9 137 6.6

*Paenibacillus* sp. JDR-2, complete genome GenBank: CP001656.1: 1350163-1350387 SASP

ATGGCTAACCAATCCAACAATTTAGTTGTACCGCAAGCGAAAGAAGCTTTGAACCGCATGAAATACGAAA  
TTGCTACTGAACACGGCATTCCGCTCTCCGAAGGTTATAACGGCAATCTGATGACTCGGAAGCGGGTAC  
GATTGGCGGCAACATTACGAGACGCCTTGTTCAAATGGCAGAACAAATCTTTAAGCGGAGCACAATCTCTC  
AATAAAGAAGATAG

8 221 3.6

*Paenibacillus polymyxa* M1 main chromosome, complete genome GenBank: HE577054.1: 1395028-1395207 SASP H

ATGAAACGTGGAACGAGCAAAAAGCCATTATGATTCAATCCGAAACGATTGCCGTCCAATTTGGATGGCAAAC  
CGATTTGGATTGAGCATGTAGATGAAGCGAATGGTATGGCAACGGTACAGATCGGCAGCAGACCGGGTGA  
TACACAAACCGTTCTGTGTAGATCGATTGGAAGAACAATAG

12 176 6.8

Butyrate-producing bacterium L2-50 GenBank: AY357288.1: 3119-3301 small acid-soluble spore protein (SASP) gene

ATGAGTGTTCGGAAGCTAAGGACGCT**ATGAA**CAAG**TTCAA****ATGG**AGGTTGCTAAGGAAATTGGTGTA  
ACCT**CAAGGATGGT**TAC**AATGGT**GACTTAACATCTGCTCAGAGCGGATCAATCGGTGGCG**AGATG****GT**TCG  
TC**AGATGATCAAG**AGACAGGAAGAAC**AGATG**TCAGGCAGATAA

14 179 7.8

Bacillus cereus Q1, complete genome : CP000227.1: c2872211-2871999 SASP C2

ATGGCAAACAATAATAGTGGAAGTCGTA**ATGAAT**TGTTAGTTCGAGGTGCTGAACAAGCTCTTGATCAA**A**  
**TGAAATATGAA**ATTGCA**CAAGA**GTTTGGTGTACAACCTTGGTGCAGATACAACAGCTCG**TTCAA**ACGGATC  
TGTTGGTGGTGAATTACAAAACGTTTGTAGTAG**AATGG**CTGAACAACAACCTTGG**TGGTA**GAGCGAACCGTTAA

8 209 3.8

Oceanobacillus iheyensis HTE831 DNA, complete genome GenBank: BA000028.3: 1185411-1185635 SASP

GTGCAGACAATAGTCTTACAGGAGGTGATATT**AATGG**CTAACAACAG**TTCAA**ATCAATTAGTAGTTCAG  
GAGTACAACAAGCTCTAGATCAA**ATGAA**ATATGAGATCGCT**CAAGA**ATTTGGTGTGTCTTTAGGTGCTGA  
TACAACCTTCTCGTGCAAACGTTTCTGTAGGTGGAGAGATTACTAAACGCTTTG**TTCAA****ACTGC**TCACAA  
CAATTGCATGGCTAA

7 221 3.2

Clostridium saccharoperbutylacetonicum N1-4(HMT), complete genome GenBank: CP004121.1: 3865590-3865796 SASP

TTAGAGTCCTTGTTTCGTA**ACTTTCAA****TCATTC**TTTTAA**CCATT**TCTCCACCAACACTT**CCATT**TTGTCTT  
GAAGAAAGGTTTCTTTATCAGTATTTTCATAATTGCTTAATCCTATTTCTGAAGCAACCTCAGTT**TTCA**  
**ATCTGTTTAAACCTGC**CTTTGCCT**TGGTACT**AAAGTTCTATTGCTTCCACTATTACTTGAT**TGCCAT**

11 206 5.3

Gottschalkia acidurici 9a chromosome, complete genome GenBank: CP003326.1: c1744362-1744150 SASP

**ATGCCA**AGTACAAACAGTAGTAATAAAATAGTTGTTCCCTGAAGCTCG**TCAAG**CTTTAAATCAA**ATGAAA**  
CTGAGATTGCAAGTGAGTTAGGATTAAGTAACTATGAGTCAATAGACAAGGGAAACCTATCTGCTAGACA  
**AAATGGT**TACGTTGGTGGATACATGACAAAAAGACTAGTAGAGCAAGCTCA**AAGAT**CATTATCTGGAAAAATAG  
6 209 2.9

Anoxybacillus flavithermus TNO-09.006 genomic scaffold chrAF6, whole genome shotgun sequence GenBank: KB205935.1: 356441-356587 SASP

**ATGAA**ACCAAACAAAACAGCAGCAGGAACAAA**CATTCAA**CATGTAAAACAACAAAACGCTGGCCAGTATG  
GGACAGAGTTTGCAGCTGAAACTGATG**TTCAA**CACGTGAAGCAACAAAATGCAAAATCAGAAGCAAAGAA  
AAGCTAA

4 143 2.8

Parageobacillus thermoglucosidasius TG4 DNA, scaffold01, whole genome shotgun sequence GenBank: BHZK0100001.1: 2683000-2683209 SASP

ATGG**CAAGATCAAG**CAATAAA**ACTGC**TTGTTCCGGGAATCGAACAGGCATTGGAACAAATAAAAT**ATGAAA**  
TCGCC**CAAGA**ATTTGGCGTACAATTAGGAGCAAATACCGTTTTCCCGGGCAAACGGCTCTGTCGGCGGCGA  
AATTACGAAACGGCTCGTTTCCCAAGCGCAAAGCCAGCTGGCCGGATCTGCTT**TTCAA**AAAAACAGAATAA

7 206 3.4

*Bacillus megaterium* DNA, transposon TnMERI1, complete sequence, strain: MBI GenBank: LC152290.1: c684-519 SASP

AT**ATGGT**GCTCAACAAGCAATTGACCAA**ATGAAATATGAA**ATTGCTAGTGAGTTTGGTGTAAACCTAGGT  
CCTGATGCA**ACTGC**TTCGTGCAAACGGATCTGTAGGTGGAGAGATTACAAAACGCT**TGGTACAATGG**CTG  
AGCAACA**ACTTGGCGGCGGACGCTTC**

7 162 4.3

*Thermoactinomyces* sp. CDF CotJC protein gene, partial cds; and CotJB protein, CotJA protein, alpha/beta-type SASP, GenBank: GQ202245.1: 1287-1499

ATGGCACAA**CAAGAT**CGCACTCGCAGTTCTAGTCAATTGCTGGTGGCTGGTGTGCT**TCAAG**CAATCGATC  
AA**ATGAA**ATATGAGATTGCT**CAAGA**GTTTGGCG**TTCAA**TTGGGAGCTGACACCACTTCCC**CGCAAACGG**  
CTCTGT**CGGGGAGAAATCACGAAACGTCTGGTTTCCATGGCTCAACAACAGCTTGGAGGAAGCAACTTCTAA**

7 209 3.4

*B.megaterium* SASP B gene encoding small, acid-soluble spore protein B GenBank: M14161.1

AGCTAGTCGTATGGAGGAACCAAAAGAC**ATGGTAGATG**CAGTTATGTTCTTAGTGTCCGAT**TCAAG**CAAAT  
ATGATTTGCGGTCAAACGCTTATTATTGATGGAGGGCGTTCACTTTTAAATGTAAGGTAAATGTATTTTGG  
GCTTGCCTTAAAA**TACTG**AAAAAAGTTTGCATACCTATTTACCTCCC**GGGCATATTAATTTTCGTGG**  
AGGTGATTAC**ATGG**CAAAGCAAACAAAACAGCAT**TGGTACAAGCACACAACACGTGAAACAACA**  
AAATGC**TCAAG**CATCTAAAAACA**ACTTCCGGTACTGAATTTGGTAGCGAAACAATGTA**CAAGAAGTGAAA  
CAACAAAATGCACAAGCAGCGGCAAACAAAAGCCAAAATGC**TCAAG**CATCTAAAAACA**ACTTCCGGTACTG**  
**AAATTTGCTAGCGAAACAAGTGTCAAGA**AGTGAGACAACAAAACGCACAAG**TCAAG**CAAAGAAAACCA  
AAACTC**TGGTA**AATACCAAGTTAATATCAAAAGCGAACTTAA**CGTCTTTTCATACGGCGATGTTAAGTTC**  
GCTTTTTCTTATATAAAAAGCAACAAAGTATGACAAA**ACTTCTTGAAACTTTACACA**ACTCGTCAAAGGAATCT

24 630 3.8

*B.megaterium* small, acid-soluble spore protein (SASP) C-1 gene GenBank: M10919.1

ATATCC**ATGAA**ATTTGCAGAGGTTTACAATCAGGTGAAATCAGCG**TGAATG**ATTTGAAAGAAAAAGACC  
TTTTGTTGTAATGCAG**TTCAAGAAGCAATGG**ATAGAATAACTAAACACAATCTTAAAAATAGCTATAAA  
AAGCAATCATTATTTAAGCCAAAGACTAACTCTTACAGAGTTAGTCC**TTTGTCAATCAA**ATAGTATA  
CC**ATGAA**CGCGAAAAAA**GAATG**ACAAACGAGTTGT**CATTCT**CTCTCCTACACATGGCAAGCGATTAGTTT  
TTAGTCACAAGGAA**ATGAA**ACATCTAATCCACTTTTTACAGTGT**CGGTTGTTTCTAAAAA**TTAGTATGT  
CATCCTAACAGGAAATATACAT**CCATT**ATTTACCACGAATACATTTCTTCCACAATAGGAACTTAAGA  
AAAACAAGGAGGTGACACAACATGGCAAACAACAACAGCAGCAACAACAATGAGTTATTAGTATACGGTG  
CTGAACAAGCAATCGATCAA**ATGAAATATGAA**ATCGCTAGCGAATTTGGTGTAAACCTTGGTGTGATAC  
**ACTGC**ACGTGCA**ATGG**ATCAGTAGGTGGCGAAATCACAAAGCGTCTTGTGCAATTAGCTGAGCAACAA  
CTTGGCGGCGGACGTTTCTAAGACA**ACTTTATAATTGATGGCTGAAGGAA****GAATGG**TCC**CCATTCT**TCC  
TTTTTGTACGAATAAAAAGGAAGGAGAGATATATGCAGCAACATCAATTAGCT

23 761 3

[*Bacillus thuringiensis*] serovar konkukian str. 97-27 chromosome, complete genome NCBI Reference Sequence: NC\_005957.1: c2943911-2943699 SASP

ATGGCAAA**CAAGA**ACAGTGGAAGTCGTA**ATGAAT**TGTTAGTGCGAGGCGCTGAACAAGCTCTTGATCAA**A**  
**TGAAATATGAA**ATTGCA**CAAGA**GTTTGGTGTACA**ACTTGGTGCAGATACAACAGCTCGTTCAA**ACGGATC  
TGTTGGTGGTGA**ATTACAAAACGTTT**AGTAGCA**ATGG**CTGAACAACA**ACTTGGTGGTA**GAGCTAATCGCTAA

9 209 4.3

*Bacillus amyloliquefaciens* subsp. plantarum YAU B9601-Y2 complete genome GenBank: HE774679.1 SASP

TTGGGCAGACGTCGTGGAGTTATGTC**AGATGAGTTCAA**ATATGAGCTGGCGAAGGACCTTGGTTTTTACG  
ATACGGTGA**AAAACGAGGGCTGGGGAGG****CATTCC**GGGCGCGTGACGCCGGGAAT**ATGGTA**AAGCGCGGAT  
TGAGATTGCCGAACAGC**AGATG**CGGCTCAGAACC**AAAATAACCGATAA**

7 185 3.8



*Bacillus velezensis* strain SRCM101413 chromosome, complete genome GenBank: CP021890.1: 229419-229670 SASP

ATGATGGACAGGACACTCGGATATTTGCGCGAGTCGCTTTTCGAACCACTTGGAGCATGATGTGTGCCGAA  
GGATTTATAAAAAGCTTCAGGTGAAGCATT**ATGAAA**ATGAGGGCGAGTTTGTGA**AAGAT**CTAAATGATTT  
GGA**AATGG**ATGCGCT**TGAAT**CAGGTGCTTGAAAAGGAAAATCAAATATGCTGAGAACGAGCAG**GATGA**GAAAG  
CGCACGATGGAAT**TGAATGAA**GTATATGAGCTCCTGTTTTAA

8 248 3.2

*Bacillus licheniformis* strain SRCM100027 chromosome, complete genome GenBank: CP021677.1: 213006-213254 SASP

ATGGACACAACATTGGGCTACCTCCGTGAGTCACTGTCTGAATCACCTTGAACACGGCATAGGACAAAACA  
TTTACAGAAAAATCGTTTCCGGCGATAT**TGCCA**ATGAGGAAGAGTTTGTCTGAGCACTTGGAGGAGCGGGA  
**AATGGA**ATTTCT**TGAATCAAG**TGCTTGAACATG**AGATGAA**ATATGCGCT**TGAATGAA**CAGGACCATAAACGG  
**CAAGA**GAAATTAACGAAGTGT**ATGAACTGCT**GTTTTTGA

13 245 5.3

*Bacillus atrophaeus* strain SRCM101359 chromosome, complete genome GenBank: CP021500.1: 227782-228030 SASP

ATGGATAAAACCTTGGGTTATTTACGGGAATCGCTTTCCAATCATCTTGATAACGATGTA**TGCCA**GCGAA  
TTT**CAAGAAGAT**TTAACTCGGCATTATACGGGGGAAGGCAACTTTGTAAAGGATTTAAAT**GATGATGA**  
**AATGTCTT**ATCTGAACACTATATTAGAAAAAGAAAT**CAAG**TACGCTCAAA**ATGAA****CAAG**CAACACGCG  
GCAAGCGAGTTAAACGAGGTTTT**ATGAAT**TGCTGTTCTGA

12 245 4.9

*Bacillus subtilis* subsp. *subtilis* strain SRCM100757 chromosome, complete genome GenBank: CP021499.1: 221169-221417 SASP

ATGGATGTTACACTCGGTTATTTGCGCGAATCACTTTTCGAATCATCTAGAAA**ATGAA**GT**CTGCCA**GCGTA  
TTT**CAAGAAGAT**GCTGGCGAAGCGTTATGCAA**ATGAA**GGGGAATTTGTAAAAGACTT**AGATGA****AATGA**  
**AATGTCTTTTT**TGAACCATGTATTAGAAAAAGAAATTAAGTATGCCCA**GAATGAA**CAGGATCAAAAACGG  
GCCAAGGAATTAACGAGGTGT**ATGAAT****ACTGCT**TTTTGA

17 245 6.9

*Bacillus subtilis* subsp. *subtilis* strain D12-5 chromosome, complete genome GenBank: CP014858.1: 2244206-2244352 SASP P

ATGACCAATAAGAATACAAGTAA**AAGAT**ATGCATAAAAACGCCCTAAAGGACACAATCCCGGCCAACCTG  
AGCCTCTAAGCGGAAGCAAAAAGTAAAGAACCGAAACCATA**CAAGA**CAAAAGCACAAC**TCAAGCCAT**GA  
TATGTAA

4 143 2.8

*Bacillus coagulans* 2-6, complete genome GenBank: CP002472.1: 31304-31471 SASP

ATGTCTGAACGTTTTTAAAGAAGAAGTGGCGAA**AAGAT**CTCGGCTTTTTATGATGTCTGTCAGGAAGAAGGCT  
GGGGCGGCATTAAGCCCGGGATGCCGGGAAT**ATGGTAA**AGCG**TGCCATT**GAACTGGCTGAAGCGCACCT  
TGCTTCGGAAAAACAATCGGGCAACTGA

6 164 3.7

*Geobacillus* sp. Y412MC61, complete genome GenBank: CP001794.1: 1747465-1747665 SASP

TTATTCCGTTTTTTCGGC**TGCCA**ATTCGCTTTGTGCTTGAGCAATGAGCCGTTTCGTAATTTCTCCGCCG  
ACGGAGCCGTTGGCGCGCGAAACCG**TGCCA**GCCCCGAGCTGTACGCCAAACTCTTGGGCGATTTCT**GTA**  
TAATTTGTTCCAACGCTTGTTTCGATGCCTGGAACAAGCAGTTTGTGCTTGAACG**GCCAT**

5 197 2.5

*Thermoanaerobacter tengcongensis* MB4, complete genome GenBank: AE008691.1: 792422-792740 SASP

TTAAGACCTTCTTTGCATATACTCTG**ATTCACTG**CC**TG**CAAGGCCTTGCTGACCGCCTTGTGCAGGGGA  
GAACCATATATGCTGATGCCTCTAG**GCATT**GCCGATTCTGCGAA**ATGAA**CC**ATT**CGCCTTAACATTTGTC  
CTCCTACTTTTTCCGCAGTCTC**TACTG**CTTATATGTCCCCAGTAATCTTCAGACCCCTTGTFTTACAGGCAA  
ACCTAATTCTGAGGCGATTTTCGTATTTTCATGTT**GTCAAG**AGCTTTATGAGCCTC**TGGTAC**CA**TT**TTGGC  
CTGT**CATT**CCAGCTTCCTCT**TGCCATT**CAATCCCTCCT

20 315 6.4

*Bacillus methylotrophicus* strain B25 genome assembly, chromosome : BAMMD1 GenBank: LN999829.1 SASP

TTGGGCAGACGTCGTGGAGTTATGTC**AGATGAGTTCAA**ATATGAGCTGGCGAAGGACCTTGGTTTTTACG  
ATACGGTGAAAAACGAGGGCTGGGGAGG**CATT**CGGGCGCGTGACGCCGGAAT**ATGGTA**AAGCGCGGAT  
CGAGATTGCCGAACAGC**AGATGG**CGGCTCAGAACCAAAATAACCGATAA

8 185 4.3

*Bacillus bombysepticus* str. Wang, complete genome GenBank: CP007512.1: c3335004-3334870 SASP N

ATGGGTAATCCGAAAAAA**ATTCAA**AAGACTTTGCACCGAATCATATTGGAACACAATCAAAAAAGCTG  
GTGGAAATAAAGGAAAGCAAATG**CAAGAT**CAAACAGGTAAACAACCGATTGTTGAT**AATGGT**TAA

6 131 4.6

*Bacillus halodurans* C-125 DNA, complete genome GenBank: BA000004.3: 77286-77471 SASP

ATGAGTCGAAGACGTGGAATCATGTTCAGATCGCTTAAAGGAAGAAATTGCGAAAGAGTTAGGGTTCTACG  
ATACCGTCCAA**CAAGA**AGGTTGGGGAGG**CATT**CGTGCACGTGATGCAGGAAAC**ATGGT**GAAGCG**TGCCAT**  
**TGAG**CTTGCTGAGCAGCAATTAGCAGAGAGAGAATCTTCACGTTAA

7 182 3.9

*Bacillus anthracis* strain PR09-4, complete genome GenBank: CP012727.1: 3383781-3383930 SASP

ATGGGTAAAAGAAAAGCAAATCATACTATTTTCAG**GAATGAATG**CGGCATCTGCACAAGGACAAGGTGCTG  
GTTATAACGAAGAGTTTGC**AAATGAA**AACTTAACTCCTGCAGAACGACAAAAATAATAAGAAACGCAAAAA  
GAACCGTAA

5 146 3.4

*Clostridium beijerinckii* strain NCIMB 14988, complete genome GenBank: CP010086.2: 4894410-4894598 SASP

ATGGATAAAGAAAGAGCACAGGAAATTTGTTTCATCATCAGT**AATGG**CAAATGTAACACTTAAACGGAACCTC  
CAATTTACATCGAAAGCGTAAATACTAATAACAGAA**ACTGCT**TAATAT**CCATT**TTTTTAAAAATCCAAAAA  
**TTCA**CGCGAAGTGCCTTTAACTAGTTT**CAAGA**ACATAATAAAAAATAG

5 184 2.7

*Bacillus mycoides* strain M2E\_15 contig\_238, whole genome shotgun sequence GenBank: LLWA01000150.1: 1332088-1332480 SASP

**ATGAA**CAAAAAACAACAAGGTTACAATAAGGCAACTTCTGGTGCTAGCATCCAAAGCACAAATGCTAGCT  
**ATGGTAC**AGAATTCGCAACTGAAACAAATGTACAAGCAGTAAAACAAGC**GAATG**CACAAGCAGAGGCAAA  
GAAAGCACAAGCTTCAGGTGCACAAAGTGCAAATGCTAGTT**ATGGTAC**AGAATTCGCAACTGAAACAAAT  
GTACAAGCAGTAAAACAAGC**GAATG**CACAAGCAGAGGCAAAGAAAGCACAAGCTTCAGGTGCACAAAGTG  
CAAATGCTAGTT**ATGGTAC**AGAATTTGCAACTGAAACGGATGTGCATGCAGTGAAAAACAAATGCACA  
ATCAGCTGCAAAAAATCACAATCTTTTAGCTCTAATCAATAA

12 389 3.1

*Geobacillus* sp. WCH70, complete genome GenBank: CP001638.1: c900227-900018 SASP

ATGG**CAAGATCAAG**CAATAAAATTGCTTGTTCCTGGAATTGAACAAGCGTTAGATCAAATAAAAT**ATGAAA**  
TTGCC**CAAGA**ATTTGGCGTACAATTAGGAGCAAATACGGTTTTCTCGCGCA**AATGG**CTCTGTTCGGCGGCGA  
AATTACAAAACGGTTAGTCGCACAAGCACAAAGCCAACCTTTCCGGATCTGCTTTCCAAAAAACCGAATAA

6 206 2.9

*Parageobacillus toebii* NBRC 107807 strain DSM 14590 Ga0265171\_11, whole genome shotgun sequence GenBank: RCWX01000001.1: 279784-279993 SASP F

TTAAAATTGTTGTTGAGTACCGCCTAATTGTTGTTGC**GCCAT**AGCAACAAGGCGTTTTGTGATTTCTCCC  
CCAACAGAACCGTTAGCGCGAGAAGTAGTATCAGCGCCTAATTTTACGCCAAATTTCTTGAGCGATTTTCGT  
ATTTCAATTTG**TTCAAG**AGCTTGTGTGCTCCTGGAACAAGCAATTGATTCGAATTGTTGTTGCG**TGCCAT**

6 206 2.9

*Parageobacillus toebii* NBRC 107807 strain DSM 14590 Ga0265171\_11, whole genome shotgun sequence GenBank: RCWX01000001.1: 1356834-1357019 SASP

ATGGCAAAACAACAACCAACAAACAAACAGCTGCAGGCACTAA**CATTCAAGA**AGTAAGACAACAAAACGCAC  
AATCTGCACAAGCAGGTCAATTCGGCAC**TGAAT**TTGCTGCAGAACTAATGTGCAACAAGTAAAAACAACA  
AAACGCGCAAGCAGAAGCTCGCAAAGCGCAAA**ATTCAAG**CAAATAA

9 182 5

*Geobacillus thermodenitrificans* NG80-2, complete genome GenBank: CP000557.1: 895067-895318 SASP

TTATTCTGTTTTTTCGGC**CTGCC**AGCTGGCTTTGGGCTTGAGCAACAAGCCGTTTTGTAAATTTCCCCGCCG  
ACGGAACCGTTGGCGCGCAAACAG**TGCC**AGCCCCGAGCTGTACGCCAAATTTCTTGAGCGATTTTCGTAT  
TAATTTGTTCCAATGCTT**TTCAAT**GCCTGGAACAAGCAGTTTGTGCTTCCACG**TGCCATCATTC**TTGC  
CTCCTTTGTCATGTTTTAAGTGTGGCATCATTAGTATGAGCAA

7 258 2.7

*Carboxydotherrnus hydrogenoformans* Z-2901, complete genome GenBank: CP000141.1: c1047458-1047186 SASP

GTGAAATC**CCATT**GGAAAACTTAAATATGAGGTTGCCGGGAACTGGGCATCGGCACCGACGACACCA  
CCTACAGACAAAACCTTAGAAAA**ATGAAA**ATCGAAGCCGCCAGGGAAATCGGTATTTATGACCAAGTTAA  
AGACGGC**TACTG**GGGTGAAGTTCCATCCCGG**GAAATG**CGGCCGGGTTGGTGGAAAGCTTGGTGGAAAAAT  
GG**TGGTAACATGGTA**AAAAAATTAATTGCCTTGGCAGAACAGCAATTACAACAGAAG**TGGTAG**

8 269 3

*Paenibacillus alvei* isolate *Paenibacillus* B-LR1 genome assembly, chromosome: PBLR GenBank: LS992241.1: 460538-460756 SASP

ATGTCTAATCGTTCTGGTTCCAACAACCTTGGTTGTTCC**TCAAGCTACTGC**TGCATTGGAGCAA**ATGAA**AT  
TCGAAATTGCTCAGGAATTGGGCGTTCCAATTTCA**CAAGA**CGGCTACTACGGAACTACACTTCCCCTGA  
TACCGGTTCTATCGGGGATACATGACGAAGCGTTTGGTGC**AATGG**CACAACAATCGCTTTCGGCGGTT  
TCCCCTAA

6 215 2.8

*Paenibacillus alvei* isolate *Paenibacillus* B-LR1 genome assembly, chromosome: PBLR GenBank: LS992241.1: c457500-457273 SASP

**ATGCCA**AATCAACAATCAGGAAGCCGCTCCAATAATCTAGTTGTTCC**TCAAG**CAACAGCAGCATTGCAAC  
AA**ATGAA**GTTTGGAGATTGCT**CAAGA**ATTGGGTATTCCAATTTCT**CAAGATG**GTTACTACGGTAACCTATAC  
TTCTCGGATATGGGCTCCATCGGAGGCTACATCACGAAGCGTCTTGTACAAATGCTGAACAGCAATTG  
GCTGGCCGCGCTCGCTAA

9 224 4

*Bacillus filamentosus* strain Hbe603, complete genome GenBank: CP011974.1 :2734256-2734435 SASP H

ATGGATGCGCAAAGAGCA**CAAGAA**ATTTCTA**ATTCAA**CGGACATGGCAAATATAATGTATAATGGAAAA  
GTGTTTATATTGAACATGTAGAC**CAAGAAATGG**GACAGCTACA**ATTCA**TAATATTGACGAACCAAATAA  
TAAACAAAGTGTCTATATCAGATTTAACAGAACAAATAA

6 176 3.4

*Bacillus anthracis* strain Ames chromosome, complete genome GenBank: CP009981.1: 3625253-3625456 SASP C

ATGGCAAACCAAATTC**TTCAA**ATCAATTAGTAGTACCAGGAGCAACAGCTGCAATCGACCAA**ATGAA**GT  
ACGAAATCGCT**CAAGA**ATTTGGTGTACAATTAGGAGCAGATTCTACAGCTCGTGCTAACGGTCTGTGG  
TGGCGAAATCACAAAACGTCTAGTTGC**AATGG**CTGAGCAAAGCCTTGGCGGATTCCACAAATAA

5 200 2.5

*Bacillus paralicheniformis* strain BL-09, complete genome GenBank: CP010524.1: 57263-57448 SASP

ATGGGCAGACGTGCAAGCATCATGTTCAGAGGA**ATTCAAAATGAAT**TGGCTAAGGATCTTGGATTTTACG  
ACACGGTAAAAACCGAGGGCTGGGGCGGAATTCGGGCCAGAGACGCCGGTAAC**ATGGT**GAAGAGG**GCCAT**  
**TGAAT**TAGCCGAACAGCATATGGCTCAAAACCAGCAAAATCATTAA

7 182 3.9

*Bacillus paralicheniformis* strain BL-09, complete genome GenBank: CP010524.1: 227024-227272 SASP

ATGGACACAACATTGGGCTACCTCCGTGAGTCACTGTTCGAATCACCTTGAACACGGCATAGGACAAAACA  
TTTACAGAAAAATCGTTTTCCGGGCGATA**TGCCAATGAA**GAAAGAGTTTGTTCGAGCACTTGGAGGAGCGGGA  
**AATGGAATTTCTGAATCAAG**TGCTTGAACATGAGATGAAATATGCGCT**TGAATGA**ACAGGACCATAAACGA  
**ACAAGA**GAATTAACGAAGTGT**ATGAA**CTGTTGTTTTGA

13 245 5.3

*Sporosarcina globispora* strain DSM 4 scaffold1, whole genome shotgun sequence GenBank: LGUF0100007.1: 3756751-3756900 SASP O

ATGGCTAAAAGAAAGGCTAATCATATTATCGAGG**GAATGAATG**CTGCAAAAAGGCCAGGGCATGGGCGCGG  
GCTACAATGAGGAATA**TGCCAATGAGCCATT**AACTGAAGCACAAAAACAGAACAATAAAAAAGCGGAAAAA  
GAATCAGTAA

7 146 4.8

*Bacillus infantis* NRRL B-14911, complete genome GenBank: CP006643.1: c2178064-2177921 SASP P

**ATGAAC**AAAAACGACAGCA**AAGAT**ATGCACAAAAATGCCCAAAAAGGAAATAACCCGGGTCAGCCTGAAC  
CGTTAAGCGGCTC**CAAGAA**AGTAAAAAACCGCAACCATA**CAAGA**CAAAAGCATA**ATTCA**GGGCACGACATGTAA

5 140 3.6

*Bacillus firmus* strain NCTC10335 genome assembly, contig: ERS1247816SCcontig000001, whole genome shotgun sequence GenBank: UFTC01000001.1: 1286847-1287047 SASP A

CTAGCGTGCTCCGCCTAGCTGCTGCTCC**GCCATT**TGCACAAGGCGCTTCGTAATTTCTCCGCC**TACTGAT**  
**CCATT**TGCTCTTGAAGATGTTTCCGGGCCAAGGTTACACCCGAATTCTGTTGCAATTTCACTTCATTT  
GATCAATTGCCTGACTGACTCCTGATACTAAAAGCTGATTAGAATTGCTGTTTCT**TGCCAT**

8 197 4.2

*Bacillus krulwichiae* strain AM31D chromosome, complete genome GenBank: CP020814.1: 642664-642843 SASP B

ATGGATAGACAACAACAACAAGCGTCTAAAACAAAACGCACAACAAGTTAGAAGACAAAACGCTGCTTCTG  
CACAAGGTGCAGATACAGAATTCGCGAGCGAAACTGATGTA**CAAGA**AGTACGTCAGCAAAATGCTCAGTC  
**TCAAG**CTCGCAAGCAGCAAAACTCT**TTCAAGA**CAACAGTAA

5 176 2.8

*Bacillus krulwichiae* strain AM31D chromosome, complete genome GenBank: CP020814.1: 1113182-1113388 SASP C1

**ATGGTA**AACAACAACAACAACAGTAATCAAT**TACTG**GTTCCAGGAGTACAGCAAGCATTAGACCAA**ATGA**  
**AATAC**GAGATTGCTT**TGAAT**TCGGTGTGAACTTGGTGC**AGATG**CAACAGCACGTGCAAAACGGTTCTGT  
TGGCGGAGAAATCACCAAGCGTCTCG**TTCAACA**AGCT**TGAAT**CTCAGCTTGGAGGTTACAATTTTAA

8 203 4

*Bacillus krulwichiae* strain AM31D chromosome, complete genome GenBank: CP020814.1: 2324595-2324801 SASP C2

ATGGCAAGCAACAACAATAG**TTCAA**ATCAATTACTAGTTTCTGGAGTACAACAAGCTTTAGATCAA**ATGA**  
**AGTAT**GAGATTGCA**CAAGA**ATTTGGT**TTCAA**CTTGGCGCTGATACAACCTCTCG**TGCCA**ACGGTTCAGT  
TGGTGGCGAAATCACAAAACGTCTTG**TTCAA****ATGG**CTGAACAACAATTAGGCGGTTACAATTTCTAA

7 203 3.5

*Bacillus krulwichiae* strain AM31D chromosome, complete genome GenBank: CP020814.1: c3432677-3432468 C3 SASP

ATGGCAAACAACAATAGCAACA**ATTCAA**ATCAACTTCTTGTCTCTGGTGTACAACAAGCACTAGATCAA**A**  
**TGAAG**TATGAGATCGCTTCTGAGTTTGGT**TTCAA**CTTGGTCTGACGCAACTTCTCGTGCTAACGGTTC  
TGTTGGTGGAGAAATCACTAAGCGTTTAGTACA**AATGG**CTGAACAACA**AATGG**GTGGTT**TTCAA**CAATAA

7 207 3.4

*Geobacillus subterraneus* strain KCTC 3922, complete genome GenBank: CP014342.1: 1252984-1253166 SASP H

TTAC**GAATG**CTCTTCTAAGCTGGCGACCGGCACGCTTTGTTTTTGATTTCGGATTG**TCAAG**CGGATGGATG  
GT**TGCCAT**CCCTTTTTGCTGATCGACATGCTCAATGTAAATGCGCTGGCCGTGTACGTGA**CATTCGCCA**  
**TC**ACTGGC**GATG**AGGCGATTTCTTGCGCCCCGACGAGCATCCAT

7 179 3.9

*Bacillus toyonensis* BCT-7112, complete genome GenBank: CP006863.1: 3120802-3121083 SASP

ATGAGTAAAAACAACAAGGTTATAACAAGGCAACTTCTGGTGTAG**CATTCAA**AGTACAATGCTAGTT  
**ATGGTAC**AGAGTTTGCAGCTGAAACAAATGTACAAGCAGTAAAAACAAGCAAACGCACAATCAGAAGCAAA  
GAAAGCACAAGCTTCTGGTGCACAAAGTGCAAACGCTAGTT**ATGGTAC**AGAATTTGCAACTGAAAC**AGAT**  
**GTGCATGCTGTG**AAAAACAATAATGCAAAATCAGCTGCAAAACAATACAATCTTCTAG**TTCAA**ACGAGTAA

11 278 4

*Thermoactinomyces vulgaris* strain DSM 43016 Ga0070019\_102, whole genome shotgun sequence GenBank: REFP01000008.1: 146200-146412 SASP

ATGGCACAA**CAAGAT**CGCACTCGCAGTTCTAATCAATTGCTGGTGGCTGGTGTGCT**TCAAG**CAATCGATC  
AA**ATGAA**TATGAGATTGCT**CAAGA**GTTTGGCG**TTCAA**TTGGGAGCTGACACCACTTCCCAGCAAACGG  
TTCTGTCCGGGGAGAAATCACGAAACGTCTGGTTTCCATGGCTCAACAACAGCTTGGAGGAAGCAACTTCTAA

7 209 3.4

*Thermoactinomyces vulgaris* strain DSM 43016 Ga0070019\_102, whole genome shotgun sequence GenBank: REFP01000008.1: c152962-152672 SASP

ATGATGAACAACAAAACTTTGGACCGCAAGAATCCAAAACCAATGCACAACAAGTTCGTACGCAAAATC  
AGCAATCCGCTCAACGTCAAACCGGTAAATATGGTACCGAGTTTGCTTCTCAAGCTGGTGCAGCTCAACA  
ACAACAACAGCAACAACAAGCTCGCCAAGCTCAACGCCAAGGGCAACAAGCTGGTAAATTCGGAACCTGAG  
TTTGCTTCCGAAACCAATGCACAACAAGTTCGTCAACAAAATCAGCAATCTCAAGCCAAAAAAGCCAAC  
AACAACAATAA

12 287 4.2

*Thermoactinomyces vulgaris* strain DSM 43016 Ga0070019\_102, whole genome shotgun sequence GenBank: REFP01000008.1: 290944-291132 SASP

GTGGCCAACCGTTCCAACCGATTGCTCGTTCGCCGGTGCSCCGGAGGCTTTAAACCAGTTCAAAGAAGAAA  
TTGCTTCGGAATTCGGAGTGAAGTTGGGTTCTGACACGGCGGCCCGTGCCAATGGTTCCGTCGGCGGTGA  
GATCACGAAACGGCTCATTTCGCAAGCGCAAAAACAATGAATGAATAA

8 185 4.3

*Bacillus subtilis* subsp. *spizizenii* strain SW83 chromosome GenBank: CP030925.1: 192171-192419 SASP

ATGGATGTTACACTCGGTTATTTGCGCGAATCACTCTCGAATCATCTAGAAAATGAACTGCGCAAGCGTA  
TTTGCAAGAAGATGCTGGCGAAGCGTTATGCAAAATGAAGAGGAATTTGTAAAAGACTTAGATGATAATGA  
AATGTCTTTTTTGAACCATGTATTAGAAAAAGAAATTAAGTATGCCCAGAATGAAACAGGATCAAAAACGG  
GCCAAGGAATTAACGAGGTGTATGAATTAAGCTTTGA

16 245 6.5

*Bacillus simplex* strain SH-B26, complete genome GenBank: CP011008.1: 2256208-2256360 SASP O

TCATTGGTTTTTTCTTTCTTTTTTTTATTATTTTGCCTTTGCATTTCCGATAATGGCGGTTTCATTTCCAGT  
TCATCATTATACCCAGCTCCATTACCTTGCCTTTTGCAGCATTCAATACCTGGTATAACATGATTTGCCT  
TTTGTTTTGTTCAT

5 149 3.4

*Aneurinibacillus soli* DNA, complete genome GenBank: AP017312.1: 1462925-1463125 SASP A2

TTATGCACGACCGCCAAGCTGTTGTTCAGCCATTTGTACAAGGCGTTTTGTAAATCTCGCCACCTACAGAA  
CCGTTCTGGCGAGAAGTTGTGTCCGGCCCGAGTGTACACCGAATTCAGAAAGCGATCTCTGACTTCATCT  
GATCCAGCGCGTTTGTCTGCTTGTTGGTACGACAAGCTTGTGTTGTGTTGCAATTTTGAAGCCAT

7 197 3.6

*Aneurinibacillus soli* DNA, complete genome GenBank: AP017312.1: 2346698-2346907 SASP C1

ATGTCACAGCAACAAAATAGAAATCGTAATAACAATGAACTGGTTGCTCCACAAGCACTGGATATGATGA  
AGTATGAAATAGCTTCTGAATTCGGTGTACAACCTCGGACCGGATACAACCTTCTCGCCAGAACGGTCTGT  
CGGAGGCGAAATTACAAAGCGCCTGGTTCAGATAGCAGAGCAACAATTAAGTGGTAAATACGTGCAATAA

6 206 2.9

*Aneurinibacillus soli* DNA, complete genome GenBank: AP017312.1: c3048746-3048531 SASP A3

ATGCCACAACAAAATCAAGGCCGTAATACTAACAACCTGGTAGTACCACAAGCGACTAATGCACTCGATC  
AAATGAAATGAGATTGCGGCGGAATTCAGTGTACAACCTGGACCAGATACAACCTTCTCGCCAGAACGG  
TTCTGTCCGGGGCGAGATTACAAAGCGTCTTGTTCAGATGGCTGAACAACAACCTAAGTGGTAAATACGTG  
CAATAA

7 212 3.3

*Aneurinibacillus soli* DNA, complete genome GenBank: AP017312.1: c3353234-3353034 SASP C3

ATGGCTAACAAACAACAGCTCAAACAAGCTGGCAGTACCACAAGCGGCTAACCGCGCTCGATCAA**ATGAA**GT  
**ATGAA**ATTGCTTCTGAGTTTCGGCGTACAGCTTGGACCGGATTCTACATCTCGTGCG**GAATGGT**TCTGTAGG  
CGGCGAGATCACGAAGCGTCTTGT**TTCAAATGG**CTGAACAACAGCTCGGTGGCCGTGCATAA

7 197 3.6

*Lysinibacillus sphaericus* C3-41, complete genome GenBank: CP000817.1: 70145-70306 SASP

ATGTCACCTCGATTAAAGAAGAGATCGAAAAGAACTTGGATTTTATGATGTTGTGCAAAAAGGAAGGCT  
GGGGCGGCATAAAAGCTCG**AGATGCTGGTAATATGGTCAAG**CG**TGCCATTGA****AATGG**CTGAGCGAGCAAG  
TAATGAG**CAAGAG**CGTAAGTAG

9 158 5.7

*Lysinibacillus sphaericus* C3-41, complete genome GenBank: CP000817.1: c4019787-4019569 SASP

ATGGCT**TTCAA**ACAACAACCGCAG**TTCAA**ACCAGCTTGCAGTACCTGGTGTACAACAAGCACTTGATCAA**A**  
**TGAA**ATACGAAATTGCA**CAAGA**ATTTGGTGT**TTCAA**ATTAGGGGCTGACGCTTCAGCTCGTGCTAACGGTTC  
CGTTGGTGGTGAATCACTAAACGTCTTGTACA**AATGG**CAGAATCTCAATTAAAGGGTA**TGCCA**ACAAC  
AACCAATAA

7 215 3.3

*Edaphobacillus lindanitolerans* strain MNA4 genome assembly, contig: Ga0116976\_13, whole genome shotgun sequence GenBank: FTPL01000003.1: c332939-332733 SASP

**ATGCCA**AACAACGACAACAACACTCGAACAAGCTTCTCGTTCCCGGCGTCCGCCAGGCACTCGACC**AGATGA**  
**AGGA**AAGAGATCGCCCGCGACTTCGGCGTCCAGCTTGGCGAAAACACGTCAGCGCGTGCCAACGGATCCGT  
CGGCGGAGAAATCACAAAGCGCCTCGTTCGCCAAGCCCAGCAGC**AGATGAA**GGGCCCG**CAAGA**ATAA

8 203 3.5

*Edaphobacillus lindanitolerans* strain MNA4 genome assembly, contig: Ga0116976\_13, whole genome shotgun sequence GenBank: FTPL01000003.1: c332939-332733 SASP

**ATGCCA**AACAACGACAACAACACTCGAACAAGCTTCTCGTTCCCGGCGTCCGCCAGGCACTCGACC**AGATGA**  
**AGGA**AAGAGATCGCCCGCGACTTCGGCGTCCAGCTTGGCGAAAACACGTCAGCGCGTGCCAACGGATCCGT  
CGGCGGAGAAATCACAAAGCGCCTCGTTCGCCAAGCCCAGCAGC**AGATGAA**GGGCCCG**CAAGA**ATAA

8 203 4

*Parageobacillus thermoglucosidasius* strain DSM 2542 chromosome, complete genome GenBank: CP012712.1: 1972688-1972837 SASP O

CTATTGATTCTTTTTTCTTTTCTTGTGTTTTGGCGCTGCGCTTCTGTCA**AATGG**CTCGTTGGAAAATTCT  
TCGTTATAGCCGGCTCCCTTTTCTTGGCGTCCGCCG**CATTCA**TGCCGG**GAATG**ACATGATTGGCTTTCC  
GTTTGCCCAT

4 146 2.7

*Bacillus aryabhattai* strain AB211 scaffold11, whole genome shotgun sequence GenBank: MCAN01000003.1: 1288271-1288450 SASP H

ATGGATGCACAAAGAGCA**CAAGAA**ATTGCTG**ATTCA**GTGACATGGCCAATGTAATACATA**AATGG**GAAAA  
GCATTTACATTGAGCATGTAGACCAACA**AATGGG**GATAGCTACA**ATTCA**TAACCTC**GATGAA**CCAAGTAA  
TAAGCAAAGTGTCTATATCTGAAGTAGCAGAACAATAA

7 176 4

**Bacillus loiseleuriae strain FJAT-27997 Scaffold1, whole genome shotgun sequence GenBank: LFZW01000001.1: 1923819-1923965 SASP O**

T**TACTG**GTTTTTCTTTTCGCTTTTTATTATTTTGTTCCTCTGTTAAAGGTTGAGAAAATTCCTCG  
TTATATCCTGCACCTTTCCCTTGTGCGCTAGCTG**CATTCA**TGCCAGGAATTAT**ATGGT**TTGCCTTACGTT  
TAG**CCAT**

6 143 **4.2**

**Geobacillus stearothermophilus 10, complete genome GenBank: CP008934.1: 325050-325199 SASP O**

ATGGCAAAGCGAAAAGCAAACCACGTAATCCCTGGC**ATGAA**CGCCGCCAAAGCCCAAG**GAATGG**GAGCCG  
GCTACA**ATGA**AGAGTTTTCCA**ATGAA**CCGCTGACAGAAGCACAGCGCCAAAACAACAAAAACGAAAAA  
GAACCAATAG

5 146 **3.4**

**Bacillus rhizosphaerae strain SC-N012,DSM 21911,NCCB 100267 genome assembly, contig: Ga0075195\_103, whole genome shotgun sequence GenBank: FRBS01000003.1: 254907-255107 SASP**

TTATCTTTGAAATCCGCCTAATTGTTGTTTCAGCTTGAGCAACCAAACGTTTCGTGATTTCCGCCACCTACA  
GAACCGTTTGCACGAGAAGTCGTGTTCAGCTCCAAGGTTTACGCCAA**ATTC**AGAAGCGATTTTCGTATTTCA  
TTTGAT**CAAG**TGCTTGTGTACGCC**GGTACT**TACCAAGTTATTTGAGTTTGAGTTA**CCAT**

5 197 **2.5**

**Virgibacillus chiguensis strain CGMCC 1.6496 genome assembly, contig: Ga0104465\_104, whole genome shotgun sequence GenBank: FQXD01000004.1: 159098-159298 SASP**

ATGGCTCGACGAAACCGTAATCAGTTAGAAG**TGCC**AGGTGCAGAAAAAGCAGTCGACTCCCTAAAGTATG  
AGATCGCTA**ATGA**ATTTAATGTCGATCTTGGCGCCGATACT**ACTGCA**AGAGAA**ATGGT**TCCGTGGGTGG  
CG**AGATGT**GAAAGCGTATGATAAAA**ATGG**CAGAA**CAAGAT**CTCGCAAATAGACAGAAGTAG

13 197 **6.9**

Observed AL pentamer frequency: 4.49

Expected:  $22/1024 \pm 1.65 \times (22 \times 1002 / 1024^2 \times 100)^{1/2} \approx 2.15\% \pm 0.25\%$ \*

$4.49\% \approx 2.15\% + 15 s$ , where  $s \approx 0.15\%$

$P(\text{Observed} \geq 4.45\%) \leq \exp(-225) \approx (e^{-2.25})^{100} \approx 10^{-100}$



## AL-pentamer proximity for the 20 aminoacid-tRNA ligases of *Deinococcus radiodurans* R1

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c651235-649715 glycine--tRNA ligase

ATGCCCGCAACCTCAATGGAAGAAGCTCGTCAGCCTGTGCAAACGCCGGGGCTTTATTTTTTCAGGGCAGTG  
AGATTTACGGCGCCTGCAAGGCTTTTACGACTACGGCCCCCTCGGCGTGAGCTGAAGAACAACATCAA  
GGCGGCGTGGTGGCGCTCCAACGTCTACGAGCGCGACGACATGGAAGGGCTGGACGCTCCATCATCATG  
CACCGCATGGTGTGCGACTCGGGCCACGAGGCGACGTTCTCCGACCCGATGATCGACAAACAAGA  
ACAACAAGCGCTACCGCTCGACCACCTTCTCAAAGACCAGAAAGCCGACGTACAGGCCAAAGTCGCCGA  
AATTATGGGTGAGAGTGCCGACAACCTTTGCCGCGCTGGTGGCTGCGCTGAACGCCAAGCCCGCCAGGCC  
TCCGCGCGATTCAAGGAAGCGGGCGTGCCTGACCCCTTCTCCGGCGAGGTAGGCGAGTGGACCGACCCCA  
AGCCGTTCAACATGATGTTCCGCACGACCATCGGCCCGGTGGCCGACGAGGAAAGTTACGGCTACCTGCG  
TCCCGAAACCGCGCAGGGCATCTTACCAACTTCAAGAACGTTGGTGGACTCGACCAGCCGCGCCTGCCC  
TTCGGCATCGCGCAGATCGCAAGGCGTTCGCAACGAGATCACGCCGCGCAACTTCATTTCCGGGTG  
GTGAGCTAGAGCAGATGAAATTGAGTTCTTCGTGACGCCCGGCACCGACGAGGAGTGGCAGCAGCAGCATG  
GCTGAAAAGCGCTGAATGGTGGGAAGACCAGGGCGTGCCGCGCGAGAAGATCGAGATTCTGGATGTG  
CCGAAAGAAGACCTCGCGCACTACTCCAAGCGCACCTACGACCTGATGTACGACTACCCACGCTGGGGC  
ACGAGGAAATCGAGGGCATCGCCAACCGCTCCGACTACGACCTCGGCTCGCACACCAAGTCGAGAGCGA  
GCTGGGCTGGTGGCGAAGGTGGAGGAAAACAACGACTCCATCGCAAGCTGACCATCCCCACCCGAA  
ACGAACAAGCCGGTCTGCGCTTCTGTGATCGAGCCGTCGGCGGGGTAGACCGGGCGATGCTGGCGGTG  
TGAGCGAAGCGTTTACCAAGGAAACGCTGGAACCGCAACGAGCGGATTGTGCTGAAGCTGAAGCCGCA  
CCTCGCGCCGATTAAAGTGGCGGTCAATTCGCTGGCCCCGAACCGGGAAGAAATCACGGACGTGGCCAAA  
GCCATCAAGGCCGAGCTTCAGGGCCTCGGCCTGGGCCGGGTGCTGTACGAGGACTCCGGCAACATCGGCA  
AGGCGTACCGCCGCATGACGAGGTGGGCACGCCCTACTGCGTACCGTGGACTTCGACACCGTCGGGCT  
GGGTGAGAACACGGACGAGAGCCTGAAAGACACCGTGACCGTCCGTGACCGCGACACGCTCGCGCAGGAG  
CGCGTAAAATCAGCGAACTGGCGGGGTGGATTCAAGGCCAAGTTGCGCTGA

25 1517 1.6 observed, 2.15 ± 0.6\* expected (\* 90% confidence interval)

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1 :1441515-1442873 asparagine--tRNA ligase

ATGACTGCTTCATCCCCAGTGACCCGTTTCGAGTATTCAAGACCTGAAAACAGCACATCGGGCAGTCCGTCG  
AGGTGCTGGCCTGGCTGCAAGCAAGTCCGGCAAGGGCAAGATTCAATTTCTCAAGCTGCGCGACGGCTC  
GGGCTTCGTGCAGGGCACCCTGTTCAAGGGCGACGTGGCCGAGGACGTGTTTCGAGGCCGCAAGCGCTG  
ACCCAGGAACAGGCCGTGCGCGTGACCGGCGAAGTCCGCGCCGACGAGCGGGCTCCCGCGGCGTGGAAC  
TCAGCGTGCCTGACCTCAGCGTCATTTCCGAAAACGTGGGCGAGTACCCGATCACGCCAAGGACACGG  
CATCGAGTTCTGATGGACCACCGCCACCTCTGGCTGCGCCACCGCCGCCCTGGGCCATGCGGGTG  
CGCGACTGCTGACGCGCGGCATCGTCGAGTTTTTCCACGGCGAGGGCTTCGTGCGCTTCGACGCGCCGT  
TCTTACGCCCCAACGCCGCTGAAGGCACCACCGAACTTTTTGAAATCGATCTCTTCGGCGAGGACAAGGC  
GTACCTCTCGCAGACCGGGCAACTGCACGCCGAGGCGGGCGCCTTCGCCTTCGGCAAGGTCTACACCTTC  
GGCCCCACCTTCCGCGCCGAGAAGTCAAGACCGCGCGTACCTGCTCGAATTCTGGATGGTTCGAGCCG  
AAGTCGCGCCGAGCAGCCACGCCGAGAACATGGCGCTCAAGAGCGCATGGTTCGAGCCACTCGCCCGCAA  
AGTCTGAGCGAGTGCAGCCCGCAACTCGAGATGCTGGGCGCGACGTAAGCAAGCTCGTGGGCGCCG  
GAGGGAACTACCCCGCGTGACCTACACCGAGGCGCTGGAGATCAATTCGGCAGCACATCGAGGACAAG  
ACCTGCCGCCAACGTCAGGAGGACGTGACGCCGTCGAGTGGGGCGACGACTCGGGGCGCCGACGAA  
AACGATTCTGGGCCACCACTTCGACCGTCCGGTATGGTTCGAGAAGTACCCGGCGGCCATCAAGCGTTC  
TACATGCAGCCCCGACCCGAAAGACCCCGCGTGGCCCTGTGCGACGACATGATTGCGCCGAAAGGCTACG  
GCGAAATTATCGGCGGCTCGGAGCGTATTACGACTACGAGCTGCTCAAGTCGCGCATCGAAGAGCAGGG  
ACTGCCGCTCGAAGCCTTCGACTGGTACTTCGACCTGCGGCGGGTGGGCAGCATGCCGACGCGGGCTAC  
GGCATGGGCTGGAGCGCTTCATCGCCTGATGACCGGCATCGACCACATCCGTGAGGCGATTCCCTTTC  
CCCGGATGCTGACGCGCATGAAGCCCTAA

34 1353 2.5 observed, 2.15 ± 0.6\* expected

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: 80923-82152 tyrosine--tRNA ligase

TCAGCCTTCCAGCACCAGCCGGGCGAACTTGTCTTGCCTTCTGATGACACGCTTCGGTCACTTGC  
TCGCGGGTGGAGCTGGCCTTTCGGGTGCTGTAAAGTCTCGCCCGGAGCTTAAGACCCCGGTTCTGAATCA  
GTTTTCGGGCGCGCCGTTGCTGGGTTCCAGGCCCGCAGCACGACCAGTTTCGCCATGCTGACGTGGCC  
CTGTTGCTTTCAGCTCACTTCCGGGACGCTGACGCGGGGAATGTTGTCGGGAATGCCACCCTTCGCCACC  
GAGCGGAAGCGCGCCTCGGCAGCGTTCGAGGTGCGCATCGGGGTGGAAGGCTCGCACCACCTTCGCGGGCGA  
GTTTCGCGGTGGGCGGCGACCGGGTGCAGGCGAGCAGTTCTCGATGCGCTCACGCGGAGGTTCGGTTCAG  
CAGCGTGAAGTAGTTGTCCAGCAGCGGGTCCGGCACCTTCATCAGCCCGCGCAACATGGCGTGGGCTCG

TCGGTCAGGCCGATATAGTTGTCCAAGCTCTTGGACATCTTCTCGGTGCCGTCCAGGCCGACGAGCAGTG  
GCAGCGTCAGCACGACCTGCGCTTCTTGCCCGTAGTCGCGTTGCAGGGCGCGGCCACCAGGTTGTTGAA  
CAACTGGTCGGTGCSCCGGAGTTCCACGTGGCGTGCAGCGCCACCGAGTCGTAGCCCTGGGTGAGCGGG  
TAGAGCAGCTCGTGCATGGAAATCGGGGTGCCGGCGCTGAGGCGCT**TGGTAAAGTCGTGCGGTTCCAGAA  
TG**CGCGCCACGGT**GTA**CTTCGCCGCCAGGCC**GATGA**TGTCCTTGTAGCCAAGCTGTTTCGAGCCACTCGGA  
GTTGTACCTGATT**TTCAAGAG**CTTCGGGTTCTTGGCGCAGAATCAGGCGGC**ACTGC**GCGAGGTAGCTCTCG  
GCGTTGGCGCGCGCTTCTCCAGGCTCAGCGGGCGGGCTTGGATTTCGCCGTGGGGTCGCCG**ATGG**  
**TC**GGGTGAAGTCGCCCTCAGCAT**GATGAC**CTTGTGCGCCGAGGTCCTGAA**ACTGCC**GCAATTTGCGCAG  
**AATG**ACGGCGTGCCCGAGGTGCAGGTCGGGGCGGGTGGGGTCGGCGCCGAGCTTGACGCGCAGGGGCTCG  
CCTTTCTCAATCTTGCCTTGGAGTCTTCTTCGGACACGAGGTCCACCACGCCGCGCTTGAGAAAGCTGGA  
TTTGCTCATTGACGGGAACATTTCTGCGGATTTTCGGACAT

19 1226 **1.5 observed, 2.15 ± 0.6\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: 104401-106887 glutamine—tRNA ligase

ATGACCGATGCCCCCGCCCAACCGCCGGGGCCGACGCCCCAGCCCCGGCCCCCG**CTGCC**CCACTCGTCG  
CGCCCAACTTCATCACCGAAATCATCGAGCGCGACCTTGAGGCCGGCAAGTACCCCCGCGTCGTGACCCG  
TTTTCCGCCCGACCCAGCGGCTACGCGCACCTGGGGCACGTCTT**TGCCA**GCCTGCTCGAC**TTCAA**CACG  
GCGCGGCAGTACGGCGGGCAG**TTCAA**CTGCGCATGGACGACACCAACCCGAGCTGGCGCGGCAGGAGT  
ACGTGGACTCCATCGCCGACGACCTGAAGTGGTTGGGGCTCGACTGGGGCGAGCACTTCTACTACGCCAG  
CGACTATTTGACCGCTACTACGCCTATGCCGAGCAACT**CATT**CGCCAGGGCGACGCTACGTGGAAAAGT  
GTAAGCCCCGAGGAACCTGTCGCGGCTGCGCGGCAACGCAACCCAGCCACGCCCGGACGCGGAGTCTTACCGTG  
ACCGTTCGGTTGAAGAAAATCTCGACCTGCTGCGCCGC**ATGAA**GGCCGGCGAATTTGCCGAGCGTGAGCA  
CGTCTCGCGCGC**CAAGAT**TGACCTCACCGCCCCCAAT**ATGAA**GCTGCGCGACCCGGTGTGTACCGCATC  
GTCAACAAGCCGCACTTCCGCACCAGCGACGAGTGGCACATCTACCCCGCCTACGACTTCGAGCACCCCT  
TG**CAAGAGCCAT**TCGAGGGCGTGACCCACTCCATGTGCAGCCTGGAATTCGTGGACAACCGC**GCCAT**CTA  
CGACTGGCTGATGGAAAACTGAACTTCGACCCGCGCCCCACCAGT**ATGAA**TTCGGGCGCCGGGGCTG  
GAATACACCATCACCGCAAGCGCAAGCTGCGCGAACTGGTGCAGGCGGGCCGGGTGAGCGGCTGGGACG  
ACCCCGCATGCCACCCTGCGTGCCAGCGCCGCTGGGGGTCACGCCGAAGCGGTGCGGGCCTTCGC  
CGCCAGATCGGGGTGAGCCGACCAACCGCACGGTGGACATCGCCGTGTACGAAAACCGGTTGCGCGAC  
GACCTCAACCACCGTGCCCCCGCGTGATGGCGGTGCTGGACCCAGTGAAGGTCACGCTGACCAATTTGG  
ACGGCGAAAAGACGCTCTCGCTTCCCT**TACTG**CGCCGACGACGTGGTGCAGGACTCGCCGACGGTCTGGT  
CG**GAATG**CCCCGGCGGGCGGGTGGCCCCGAGGAAGCTGTGCGCGACGTGCCGCTCACCCGCGAGCTG  
TACATCGAGCGCGACGACTTCAGCCCCGCGCCGCCAAAGGGCT**TTCAAG**CGCTGACCCCGCGCGCACGG  
TCCGGCTGCGCGCGCCGCGCAT**CATT**CGCGCCGACGACTTCGGGACCGAGGCGGGGAGGTCACCCA  
CATCCGCGCCACCCTGCTGGGTGAGGACGCCAAAGCGCGGGCGTCCACTGGGTGAGCGCCGAGCGT  
GCC**CTGCC**GGCCGAGTTTTCGCTCTATGACCGCTGTTCCGGGTGCCGCACCCCGAGGGCGAAAACCGCG  
ACGTGGAGGACGACAGCGCCGGCCCCGCCGAACACGAGGCCGAGCCGGGCGCCGGGCGAGGAAAACGGCGCC  
GGTCAGCCAGGGCTTCATGCGCTACCTCACCCCGATAGCCTGCGCGTGTGCGCGGTTACGTGGAGCCG  
AGCGTGGCCGGGACCCGGCAGACACCCGCTACAGTTCGAGCGTCAGGGCTACTTCTGGCGGACCCCTG  
**TGAAT**TGGAACGGGTAGACAGCCGCGAGGACGCGCTGGTGTTCGGGCGCATCATCACCTGAAAGACAC  
CTGGGGCAAGCAGGGCGGGCACCAGCAGAAGGCGGAAGG**CAAGA**AGCGACCCCTCTACCAAGGGGAGA  
GGGCTGACGAAGTCAGGGGTGAGGGTCTTCTTCCCCAGCTAAAGCCATGCGCCCAAAGCCCAGCCCC  
TCACCCCGAACAGGACGCCGA**ATTCA**CCCGGCTGCTCGGCCCTCGGTGCCTCGGAGGGGGACGCCCGCAC  
CATCGCCCGTGACCCGGCGCTGCTCGCTTCTGTTGGGGGGGGCGGGCGCCGGGGGACACCTTCGCTCAGGTC  
GCCTCGTGGACAGTCAACGAACCTCGTCGCGGGCCTGCGTGCCGGGAGGTGAAGGTGCGGGCCGCCGACC  
TCGCGCCGCTCGCCGAGGGGTGGCGAGTGGGCAACTCAGCGCGCATTGCGCGTGAGGCCCTCGCCCG  
CGCCCGCCAGCGGGGACGCGCCGCTCACCATCATCGAGCGCGAGGGTC**TGAATG**CCGGCTGAGCGCC  
GAGGCCCTGCAACAGGTAGTGGCCAAAGTCATCGCCGCAATCCCGACAAGGCGGAGGCCCTACCGCGGCG  
**GCAAGA**CCGCGCTGCTCGGTTTTTTTACCAGGCGAGGTCATGCGCGGACCCCGGCAAGGCCGACCCGCA  
GGCGCTCGCGGGCGCTCAAAGACGCTCTCGCTGA

26 2483 **1 observed, 2.15 ± 0.5\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c156062-154233 arginine--tRNA ligase

ATGGATTTGAAGGCCCAACTCAAAGCAGCCGTCGAGCAGGCCCGCGCACCC**AGATGGGAATG**CCCGTCGACG  
CC**GCCATTCA**GGAAACGCCCGCAATAAACCCGGCGACTACGGCACGCCCGCCGCTTCC**AGATG**GCGAA  
GGCGGCCGGGGCAACCCCGCCAGATCGCCGCGCAGCTCGCGCAGACGGTGGTG**CTGCC**CGCCGGCATC  
CGCAGGGTGGAGGCCACGGGGCCGTTCTGAACTTTTTCTCGACGCGGGCGCTTTCGTGCGCGGCGTGG  
TCGAGCGTCCCTTCGAG**CTGCC**CAAGCGCGAGGGCAAGGTCGTTCATCGAGCACACCTCGGTCAACCCCAA  
CAAAGAGCTGCACGTGGGCACCTGCGTAACGTGGTGTGGGCGACTCCATGGCGCGGATTCGCGGGCG

GCGGGGCACACCGTTCGAG**GGTAC**AGA AACTACATCGACGACACCGGGCGACAGGCCCGGAGTCGCTGTTCG  
CCACCCAGCATTACGGCCGCTCTGGGACGGCGTGCAGAAGTACGACCAGTGGCTCGGTGAAGGCTACGT  
GCAGCTCAACGCCGACCCCGAGAAGCCCGAGCTGGAAAGCGGCATCATGGAGATCATGCACAAGCTCGAA  
GCCGGGGAGCTGCGGCCCTGGTTCGAGCAGACGGTGAAGGCGCAGCTTCAGACCTGCTTTTCGCTCGGCG  
CGCGCTACGACCTGCTGAACTGGGAATCCGACGTGGTGGGCGAGCGGGTTTCTGGCGCAG**GCCATGAA**CAT  
TCTGGAAGGCAGCAG**GGTAC**ACCTCGCGCCCCACCGAGGGCAAGTACGCCGGGGCGTTCATCATGGACGTG  
TCGGAATTTATGCCGGGGCTGGAAGAACCAGACGTGGTCCCTCGTGCCTCGGGCGGCACC**GCCAT**GTACG  
CCGCCAAGGACATCGGCTACCAGTTCTGGAAATTCGGCCTGTTTCGAGGG**GATGAAATCAAG**CCCTTCAT  
GCAGGACCCCGAAGGCAACACCATCTGGACGAGTGGCCCGGACGGCCAGCCCGAC**GATGA**GCGGCGCTTC  
GGGCACGCGCAGGAGGTCATCAACGTGATCGACTCGCGCCAGGACCACCCGACAGCGGTGGTGCCTTCGG  
CGCTGGGTGTGGCGGGCGAGCAGGAGAAGGAGGAGCGCAGCATCCACCTCTCCTACGCCTTCGTGACGCT  
GGAAGGCCAGACCATCAGCGGGCGCAAGGGCATCGCGGTGAGTCCGACGACGCGATGGACGAGGCCAG  
AAGCGCGCGCTGAGCGTCTTGCAGGGCATCAATCCCGACCTGGCCGCGCGGAGGACGCCGCCGAAATCG  
CCCGGCGCATCGGCTGGGC**GCCAT**CCGCTTC**GCCATGCTCAAG**GCCGAGCCGACCCGCAAAATCGACTT  
CCGCTGGGAGCAGGCGCTGGCGCTGAACGGGGACACCGCGCCCTACGTGCAGTACGCCCGCTGCGCGCC  
GCCAACATCCTGAAAAAGGCCGAGGAAGCCGGTTA**TGCCA**CCGACGGCACCGGGGCCGACTGGGACGCG**C**  
**TGCC**CGACATCGACCTCGTGCTCGCCAAGCAGATCGCCAAG**CTGCC**CGAAGTCGCCGCGCAGGCCGCCG  
CATCCACTCGCCGCACGTGGTGGCGCAGTACGCGCTCGATCTCGCCACATCC**TTCAA**CGCG**TGGTAC**AAC  
GCCAAAAGCAAGCAGGGCAAACCCGCCACCAACGTCTGCAATCGGAAGAAGGGTTGCGGGAAGCGCGG  
TGGCGCTGATCGTGGGCTGCGCAAAGCCTTCGAGGACACCCTGGACCTCATCGGCATCGAGATTCCGGC  
GGCGATGTAA

29 1826 **1.6 observed, 2.15 ± 0.55\* expected**

Deinococcus radiodurans R1 chr 1, complete sequence GenBank: CP015081.1:c371799-369337 phenylalanine--tRNA ligase subunit β

**ATGAA**ACTTCCCTACTCCTGGCTCAAAGA AACTGATTCCCGACCTCCCGCCGTCGCTGACCTCGAACCCA  
CCTTCGCGCATCTCGGG**CTGCC**GCTCGAAGGCGTGGAGGACGTGCCCGCGCCGGTGGGCGGCGTGTGCT  
GGTGGCCGTGAAAGCCGCCGAACCGATGGAGGGCACGCAGCTCACCAA AACTGACGCTCGACACCGGCGAG  
**AATGGT**GAAAAGACCATCGCCAGCGGTGCCCCCAACGCGGTGGGT**CTGCC**GGCGGGCAG**ATGGT**CGCG  
TGGTACGCCCCGGCACACGCTCGGCGGGATAACCTACGGCGTGCGGCCCATGCAGGGCGTCGAAAAGCTG  
GGGCATGGCCGCCAGCGCCAAGGA AACTCGGCATCGGCGAGAGCAACGCGGG**CATT**CTGACCTTCCCGGCA  
GGCACGGCGGGCGCCGGCACACCCCTGCGCGAGCTGTGGCCCGCCGACAGCGTGTGGACGTGGAAGTCA  
CGCCCAACCGCGCCGACGTGCTGAGCGCGCTGGCCCTCGCCCGGACCTCGCGGGCTACCTGAACCTGGA  
GCTGAAAGA AACCAGATTC**CTGCC**GCACCCACCCAGCCGGCGAGATCCGCGTCAG**CTGCC**TGACCG  
GGCCGGGTGCTGGACCGCACCCCGCAGGGCAA**ACTG**CGCTTCGGCTGCGA**CCATT**TCGCCCGCCGCGCC  
TCAGCGGCTGCAAACCGCCCCGCGCCGCTGTGGATGCAGCGCCGGGTGAGCCTCGCGGGGATGCGCTC  
CATCGACCTGATCGTGGACACCTCCA AACTACGTGATGCTCGAACTCGGTGAGCCACCGCGCTGTACGAC  
CGCCGCGACGTGGCCGGGGACGGCCTGGTGGTGGCCTTCGGTCTGCGCGAGGGCGAGACT**GTC**AAGGACC  
TGCTGGGAATACGCATCAGGTGCGCCCGGAGGACCTGCTGATTCTAGACGCGG**GAATG**AGCGACGAACC  
GGT**GATGAC**GGTAGCCGAGGCTTT**TGCCA**GCAGCCGGGAGCCCAAGGAAGGCTCGCACGTGCTGGGCATC  
GCGGGCATCATGGGGGGCGACCACGGCCACGTGCGTGCAGACACCCGCGACGTGGTGTGAGTTCGGCGC  
ACTTCGACCCCGTGTGCTGCGCCGACCTCCACCCGGCTGGGCC**TCAAGA**CCGACGCCGTGTACCGCTA  
CGAGCGCGGCGTGGACCCGCTGCTGGCGCCCAAAG**CTGCC**GTCGGGGTGGCCGAGCTGCTGCGCGCGGCG  
GGCGGACCCCGAAGCCGGGCGAGACGGTGGTGGGACCCCGAAGTCCGCGAGACCATCACACGACCG  
GCGAGCAGATTGCGCCCTGCTGGGGATGCACATCGGCACGGCGGAAATGCGCGAGAGCCTCACTCGCCT  
GGGCTGCACAGTGACGGGCGACGGCGACAGCCTGACCGTCAACCCGCTTCGTGGCGGGTGGAC**ATGGT**C  
ATCTGGCAGGACCTCGCCGAGGAAGTGC**ACTG**CACGGCTTTACCGAGTTGCCCGAAACG**CTGCC**CA  
CCCTGCGCGTGCACGAGAGCAACATCGGCGCCTCGGCCAGAGCGAAGCCCGCGCGGA**ACTG**CGGCGCAC  
CCTCGCCGGGCTGGGCTTTT**CAGG**AGGTCGTGACTTACACCTTACCAGCGACGAGGAAGCGCAGAAGGCG  
CGGGCCGAGGCGCCCGGCTGCGCCTGCGTAACCCCATGACCACCGACCGCACTGGGATGCGCACGGCGC  
TGTACCCCTCGCTGCTGCGGGCGGGCGGCGCACACCCGAAAGGCGAGCGGGCGC**TACTG**TTTGAATCGG  
GCGGATTTTCCCGCGGCGGGCGAGCAGGAACGCCTGGGCTGCTGATGCGCGGCGACCTCGCGGCCCGC  
ACCTATCAGGACGGCGTGCAGCGGATTT**CTCAAGG**ACTGGTGCAGGGCTTTGCCGGAGCGG  
TCGGCGCCAGCTTTGCGCTGGAGCA**ACTG**CGCGGCGACGACGTGCCCGCCGCCCTGCACCCCGGCGTCG  
GGGCGCGGTGGTCTGGAACGGCGAGCGGGTGGGCTGGTGGGCGCGCTGCATCCCGAAATTGCCAGGAA  
TTTGGCTTGAAAGGCGACACTTTCTGATGGAAGCCGCG**CTGCC****CTGCC**GGGCCGCGACTGGGCTTCC  
GCGACCCAGCCGTGCCCGCGCGCTGGCGTGACCTCGCGGTGATTGCGCCGAGGGCGTGAGCTACGG  
CGACATCGTCGGGTGCTGAAAGGCGCGGGCGGCGA**ACTG**CTGGAGAGCGTGAGCCCTTCGACGTGTTC  
ACCGGCAACAGGTGGGCGCGGGCAACCGCTCGGTGGCGGTGCGCCTGACCTACCGGGGCGCCAAAACCC  
TGACCGACGAGGAAGTGGACCCGGT**TTCAA**CGCGCAGATTGACGCGGTGAAGGCGCGGGGCTGGGCGAT  
TCGGGAGAAGTAA

32 2459 **1.3 observed, 2.15 ± 0.5\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: c373751-372732 phenylalanin--tRNA ligase subunit alpha

ATG**CAAGA**CACCGCG**ATTCAG**GAA**ATTCAG**GGGGCCGAGACGCTTGAGGCGCTGCAGGCCG**TCAAGA**CGA  
AGTACGTCGGCAAGTCCGGCCTGGTGACGAAAGAACTTGGGAGCCTCGGCAAG**CTGCC**GCCCCGAGGAACG  
CAAGGCGTGC GGCGCCGAAATCAACGTGGTGC GGCGCG**ATTCAG**GGCCGCGCTGGACGAGAAAGAGTCA  
GCC**TCAAG**CGCGCGCACTCGACGCTAAACT**TGCCAG**CGAAG**CCATCG**ACGTGACG**CTGCC**GGCG**CTGC**  
**CCCTGCC**GGCGGGCGGGCTGCATC**CCATT**TCCCGCGTGCTCGACGACCTGATCGGCATCTACCGCC**AGAT**  
**GGG**TACGCGGTGATCGAGGGGGCCGGAAGTCGAGGAAGAGCACTACAACCTTCGAGGCGCTCAACGTGCCG  
**TGGTAT**CACCCCGCGCGGACCTG**CAAGA**CACCTTCTGGCTGGAAGACGGACGCCTGCTGCGCACCCACA  
CCTCGCCCATGCAGATTTCGTAC**ATGGT**GGACCACGAGCCGCC**TTCAAGAT**CGTGGTGC GGCAAGGT  
CTACCGCTACGAAGCCACCGACGCCACCCACGAGGCGATGTTTACCAACTGGAAGGGCTGGTGGTGGGC  
GACGGCATCAGCATGAGCGACCT**TCAAG**GGGA**CCATT**GCCGA**AATGG**CGCGGGGCTGTACGGCGGAGCG  
CCAAGGCGCGTTTTAGCCAGCTATTACCCCTTCGTGGAACCCGGCGCCGACTTCGCCGT**GTA**CT**G**GGA  
AAACCCGCGCGGGCGAGAGCAAGTGGCTGGAAGTGGGCGGGCTGCGGC**ATGGT**GCATCCCAACGTG**TTCAAG**  
GCGGTGGACGACCTGCGCGAGGGCGGGCAAGGACCGCGTGTACGAGGGCAAAACCGGCTTCGCGTTCG  
GGCTGGGGCTGGAGCGCATC**GCCAT**GTGTAAGTACGGGATTCGGACATCCGCTACTTCTACGCCAACGA  
CCCGCGCTCATCGGGCAGTTCCGGGGGAACTGG**GATGA**

32 1016 **3.2 observed, 2.15 ± 0.75\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: c427261-424589 alanine--tRNA ligase

ATG**ACTGC**GCCGCTCACCACCGCCGAGATTCGCGAGA**GGTAC**CTGTATTTCTTCGAAACCAAGGGGCATC  
TGCGC**CTGCC**CTCGCACTCCA**CCATT**GCCCCGACCCACACGCTCTTTACGGTGCCTGGGATGCAGCC  
**TTCAAG**GAAACTTCATGGGCGCGCCCGCGTGTTCGATGGCCAACCCAGCAAGCGGTACCACCGCG  
CAAAAGTGC GTGCGGTGGGCGACATCGAGAACGTGGGCGCACCCGCGCCACCTCTCGCTGTTC**AGA**  
**TGAT**GGGCAACTTCTCCTTCGGGACTACT**TTCAAG**CGCGACGCG**ATTCAT**GGGCTGGGAGTTCCTGAC  
CGGCCCGGAGTGGATGGGCATGGACAAGGA**CAAGATG**TACGTCACCGTCTACAAGGACGACGACGAGGCG  
TTCGGC**TACTG**ACGCAGGACATCGGTTTTGCCCGCCGAGCA**CATTCA**CCGCTTCGACGCCGACGAGA  
TCTGGCCCCCAACGCGCCGCTCGAAGGCCCAACGGGCGGTGTGGCCCTGCTCGGAGATTTACTACGA  
CCGTGGCGAGAACTACGGCGACGACACCTGGGGGACTACTACCAGACCCGCGAGAGCGCCCGCTTCTG  
GAGGTCTGGAACCTGGTGTTCCCGAGTACGACCGCCAGGACTTGGACGCCAGCGGTACAGCCGTG**TCA**  
**AGG**ACTTGCT**TTCAAGA**ACATCGACACCCGGGATGGGCTGGAGCGCGTCCGACGCGTGGTGCAGGACGT  
GCCCCACTTCTATTCCAACGACGTGTTTAAGCCCATCGTTCGAGCGGGTGGCCGAGCTGAGCGGCAAGCCT  
TACGAGGGCGAAACGAGCGTCTCGCACCGCGTGGTTCGCCGAACACATCCGCTCGGTGAGCATGATCGTGG  
CCGACGGCACCGCCTTTTCCAACACGGGCGGGGCTATAACCGCCGCAAAATCATGCGCCGC**GCCATTFC**  
TCACG**GGTAC**ATGCTGGGCTGCGCGAGCCGACGCTCTACCGGTTGGTTCGAACTGGTGGTGGACAGCATG  
GGCGGCGCTTACCCGAGTTGCAGCACAACCAGAGCCGGGTGCAAGCGAGCGTGAGGGCGGAAGAAGAGC  
AGTTCC**TCAAG**ACGCTCGAAGGCGGCAT**TCAAG**CGCCTGGGCGGGCTGCTGAGCGGCATGGAAGGGCAG  
CACGCTGGTGGGCAAAGACGCCTTCGAGCTGTACGACACCTACGGCTTCCCGGTGGACCTGACCAAGGAA  
ATCGCCGAGGAATACGGCGTGAGCGTGGACGAGGCCGGCTACGCCGAGAGCCTGGAACACGCGCAGGAAA  
TCGCCCGCGCGGGCAGCAAGTACGGCAAGTCCGAAGTGTTCGGCG**GCCAT**CAGGAAGCGCTCGACGGC**CT**  
**GCC**CGCCACGCAGTTTGTGCGGTACGACCAGACCTCCGGGGACGGTCAGGTGCTGGCGCTGCTGAGCGGC  
GGCGAACGCCTGAGCCAC**CTGCC**GGCGGGCAGCGAGGCCACCGTCTGCTCTCGCAGACGCCCTTTTACG  
GCGAAGGGCGGGCGAGGTGGGCGACACCGGGCGGCTGGAGTGGGACGGCGGGGCGGTATCGTGC GCGA  
CACCCAGAAGACTAAGCAGGCGTGTTCCTGCACGACGCTGCTGGTGGAGCAGGGTGGAGTGAAGAAGGC  
ACCCGCGTGC GGCGGTAGTCAGCCCCGAGCGGGCG**GCCATTCA**GCGCCACCACACCCGCCACCCACCTGC  
TGATGCGGCACTCCGGGCGGTGCTGGGGGGCGGCGTGCAGCAAAAAGGCTCGCGGGTGGCTGCTGACCA  
ACTCCGCTTCGACTTCTCGCACGGGGCC**GCCAT**GAGCGCCGAGGAAATCGCCAGGTGCAAAACGCTGGTC  
AGCCGCTGGGTGAG**TGCCA**ACTTCCCGGTGTCGTGGCGCGAAATGCCGATTGCCGAGGCGAAAG**CTGCC**  
**TGCCA**CCGCGCTGTTCCGGTGAAGAAGTACGGCGACGTGGTGC GGCTGGTGGGTTGGAAGGCGACGTGGA  
CTTTGGCGGCCACGCGGTCTCCAGCATGGAGCTGTGCGGGGGGCGCACGTCTCGCGTACCGGCGACATC  
GGCGCCTTCGTGATTGTCTCCGACGAGAACGTGGCGGGCGGGCGTGC GGCGTATCGAAGCGCTGGCGGGCG  
AGCAGGCTACGGCGTGGCTGCGCGGGCGCCTGAACGCCTCGGCCAAGGCGGGCGGGCTGCTCAATACCTC  
GCCGAGGGCCTGGAAGAGCGTGTGACGGCCCTGCAGGGGCAAC**TCAAG**GCCCGCGAGAAGGAAACGGCC  
CAGGCGCGCCGCAACTCGCCGAAGCGC**AGAT**GGGCGGGGGTGGCAGCGCGGCCAGACCCGCGAATTGG  
GCGG**TTCAAG**GTCCGACGCTGAAGCTCTCGGGATCGAGGGCAACGA**ACTGC**GCGGGCCCGCCGACAA  
**ACTGC**TCGACAGAGCGCGGACCTCGTGGTATCGCTCCGACAAGGGCCTGGTTCG**TCAAG**GCGGAC  
AAGGACGCCGTGGCGAAGGGCGCGCACGCCGACAACCTGGTGGAGCAAGCTCGCGGGCGGGCGGCGCA  
AGGGCGGGACGCCCTGACATGGCGCAGGCGGGGATTACGGACGCGGGCGGGCGGTGGGAGCGCTGGA  
CACGGCGTTTTGA

46 2669 **1.7 observed, 2.15 ± 0.45\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: c554453-551982 leucine—tRNA ligase

ATGACCCAGGAAGCCACTAAACCCAA**CATTCA**GGAAACCCCGCGCTGAGCGTTATAACCCGCAC**GCCAT**CG  
AACAGAAGTGGCAGGGCAGTGGCAGGAAAGCGGCCTGTACAAATTCGACGAGAACGCCCCCGCGAGAA  
GTTCTACGCCCTGACGATGTTCCCGTACCCCTCGGGCAACCTGCACATCGGCCAC**TGGTAC**GCCAACGTG  
GCCCCGACGCCCGCGCGCTGGCTGCGGATGCGGGTATAACGTGCTGTTTCCGATGGCTTTCGACG  
CCTTCGGG**CTGCC**CGGAGAACGCC**GCCATCA**AGAACACCAACCCGGCGACCTGGACCTACGCCAA  
CATCGAGCGCATGACCGGGCAGTTCTCGCGCATGGGCAC**GATGA**TCGACTGGTCGCGCAAGTTCGCCACC  
TGGACCCCGA**GTA**CTACCGCTGGAACAGTGGTTTTTCATCGAGTTCTGGAAGCGCGGGCTGGCTACA  
AGAAGGGCGGGCTGGTCAACTGGTGCSSAAGGACCAGACGGTGTGGCGAACGAGCAGGTTCGTGAACGG  
CC**ACTGC**GAGCGCTGCGGCACGGCGGTGCGAGCGGCGCAACCTGAGCCAG**TGGTAC**CTGAAAATCACCGAC  
TACGCCGAGGA**ACTGCT**TGGACTTTAGCGCCACCGACATGCCCGAAAAGGTCCGC**GCCAT**GCAGACCAACT  
GGATTGGCAAGTCGGTGGGCGCCGAGGTACCTTCGACACGCCCGCGGCCCGGAGACGGTCTTTACCAC  
CCGCCCCGACACGCTGATGGGCGCGACGTT**ATGGT**GCTGGCGCCCGAGCA**TGCCA**AGGTGAAGGAATG  
ACCACCGACGAGCAGCGCCGAGGTGGAAGCTTACGTCG**CTGCC**GCCGGCG**CAAGA**CCGACGTGGAGC  
GCCAGCAGGAAGCGAAAAGACCGCGTGTTCACCGCAGTT**TGCCA**CCCACCCCATCGACGGTCA**CA**  
GTTGCCCATCTGGGTGGCCGACTACGTGCTGGTCACCTACGGCACCGGCTCCATCATGGCCGTGCCCGC  
CACGACGAGCGCACTTTGCCTT**TGCCAAGA**AGTTTGAC**CTGCC**TATCAGGGAAGTCATCCGCGCCGAGG  
GGTCCGAGGGCATGGGCGACCAGCCGAGCGAGCC**GTA**CTCGGGCGAGGGCCAGATCGTCAATTCGGCGA  
GTTTCGACGGCATGCCCGCGGCAAGGCGAGCATCGCG**GCCAT**CATTGCCCGGCTGGAGGAACCGGGCGTG  
GCGAAGGCCAAAACGACCTACCGCCTGCGCGACTGGCTGTTTCCCGCCAGCGC**TACTG**GGGCACGCCGA  
TTCCCTTCGTCC**ACTGC**GAGAAGTGCGGGATGCAGCCCGTTCGGAGGACCAGTTGCCGGTGAAGTTGCC  
CGAGAACGTGGCCTTACCCCGACCGCCAGAGCCCGCTGAAGCTG**GATGAA**GAGTGGAAAGACGACCACC  
TGCCCTGCTGCGGGCGGCCCGCGAGCGCAACCGACACCATGGACACCTTCGTGGACTCGAGC**TGGT**  
**AC**ATGTACCGCTACCTGTGCGCCAACTACGACGGCGGCCCTTCGACCCAGTAAGGCGGGCCTG**CTGCC**  
GGTGGACCTGTACACGGGCGGCATCGAGCAC**GCCATT**TTGCACCTGCT**GTA**CTCGCGCTTCTGGACAAA  
GTGATGCGGCACATGGCGTACGACGCAAAAGCGAACCCTTTCGGCGCTGCGCAACCGGGC**ATGGT**GC  
TGGGCGAGGACGGCGA**AGATG**TCCAAGTCGCGCGGCAACGTGGTCGACCCCGACGACCTGCGCGA  
GTACGGCGCCGACACGGTGCGGCGTTCCTGATGTTTCATCGCCCCCTGGGA**ACTGGGCGGCCCTGGGAT**  
CCGCAGGGCATCAACGGCCCCGAGTAAGTGGCTCTCGCGCGTCTGGAACGTGTTCTTCGAGGAGAAGGTCA  
GCGGCCCGGAGGAA**AGATG**CAAGGCGCCGACGTGCGCTTCGCGCTTCACTCCGCGCTGAAGAAGGTAAA  
CGACGACTTCGAGCGCATGAGC**TTCAA**ACCATCATTTTCGACCTGATGGAGCTGACGAACCGCTGGTC  
AAGGCCAAGCGCTCGCCGGTCTTCGGCACGCCCGTCTGGGAGGAAGCGCTGGACATC**TTCAA**CCGCATGC  
TGGCCCCCGTTCGTGCCCCACATCGCCGAGGAAATCTGGCATGAGCGTGGGCAAAAGGGCAGCGTGCACAC  
CGCGCAGTGGCCGAGGTGGACGAGGCGGGCGGTGCGCGACACCGTGACCATCGGGCTGCAGGTGTCG  
GGCAAGGTGCGCGGGCAGGTGAGCATCTC**CAAGAC**CGCCAGCCAGGAAGAAGCCCTGAGCGCGGGCGGG  
CCATCGCCGAGGTGCAAAAGCATCTGGAAG**CAAGA**CCGTG**TCAAG**GAGATCTACGTGCCGGGGCGGAT  
TATCAACATCGTGGCGAAGTAA

47 2469 1.9 observed, 2.15 ± 0.5\* expected

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: 628397-630346 threonine--tRNA ligase

ATGCACGTACCCCTTCCCGACGGTAAACAACCTCGAC**TTCAA**ACCCGGCGCCACGGCGCTCGACGCCGCA  
AAG**CCAT**CGGCCCGCGCTGGCG**CAAGAC**CGCCCTGGGCGCCACCGCC**AATGG**CGAACTTACCGACCT**GAT**  
**GA**CGCCGCTGAGTGATGGCGCCAGCATCACCTGATCAC**CAAGA**AGAATCCCCGGCGACCCGCGCCGCTG  
TTCCGGCACTCGCTGGGCCACGTATGAGTCAGGCGGTGGGCGAATATTACAAGGCCAAAAGGCTACGGCC  
CCGAC**BCCATCAAG**CGCGGGGTGGGGCCGTACATCGAAAACGGC**TGGTA**TCAGGATTCGAC**CTGCC**CGA  
GCCGCTGAAAGAAGAGGAC**CTGCC**CGAGATCGAAA**AGAT**CATGCGCGACATCATCGGGCGGGCCTGGAC  
ATCACCCGGCGCAAAATCAGCAAAGACGAGGCGCTGGCGCAGTTTCCCCACGACCCCTACAAGGCTGAGC  
TGATCGAGGGTTTGGCCGACGACGAACCCATCACCTTCTACAGCCAGGGCGACTACACCGACCTGTGCCG  
GGGGCCGCACTTCCCTAGCACTGGCAAG**CTGCC**CGAGAGC**TTCAAG**CT**GATGA**GCACCAGCGGGCGCTAC  
TGGCGCGGCAACGAGAAAAACCCATCCTCCAGCGCATCTACGGCGTGGCGTTCGCCACCCAGAAGGAAC  
TGGACGAGTACCTGTTTTCAGCTCGAAGAAGCAAGCGGGCGGACACCAGCTGGGCAAGGAACCTTGA  
ACTGTTACCATCGACCCGCTGGTGGGCAAGGGC**CTGCC**GCTGTGG**CTGCC**CAACGGCACGGTGTGCGC  
GAGAGCTGACCACTTC**ATGAA**AGAGCAGCAGTTCAGCGCGGTTACCAAGGGCGTGGTACCCCAACA  
TCGGGAACCTCGACTGTACCGCACCTCCGGGCACTACCC**GTA**CTACTCCGAGTCTCAG**TTCAA**CCGAT  
TCAGGTGGACGAGGAAGAGTACATGCT**TCAAG**CC**ATGAACTGCC**CGCATCACGTGCGAATTTACGCCAGC  
AAGCCGCGCAGTACCGCGAC**CTGCC**GGTGC GGCTGGCCGAGTTCGGCACGGTGTACCGCTACGAGCAGT  
CGGGCGAACTCAACGGCCTGACCCGCGTGC GCGGCTTACGCAGGACGACGCCACATCTT**CTGCC**GCCC  
GGACCAGC**TCAAG**CGCGAGTTTCTGGACGTGCTCGACCTGACGGTGTGGTGC**TCAAGA**CCCTTCGGCATG  
ACCGACGTGCGTTTTCCGCGTGGGCGTGC GCGACCCCGAGTCCGACAAGTACGTGGGCGACGAGCAGAACT  
GGGCGCTGGCTGAGCGCCAGATTATCGAGGCGGTGGAGGAAGTCGGC**CTGCC**CTACACCATCGAACCCGG  
CGACGCAGCCTTCTATGGCCCCAAGCTCGACTTCGTGG**TCAAG**GACGTGCTGGGGCGCGAGTGGCAGCTC

GGGACCATTTCAGGTGGACTACAACCTGCCCAGAGCGCTTCGACATCTCCTACACCCGGCGAGGACGGGCAGG  
AGCACCGCCCGATCATGATTACCCGCGCCCTTCGGGTCCATCGAGCGCTTCACCCGGCATCCTGATCGA  
GCACTACGCGGGCGACTTTCCGCTGTGGCTGGCGCCCCGGCAGGTCATGATCATCCCTATTGCCGACCCG  
CACAACGCCTACGCTGAGGAGCTGCGTGAGGA**ACTGC**ACCCGCGCCGGTTGCGCGCCGAGGTGGACGATT  
CCTCCAACCGCATGCAGGCCAAGGTCCGCGACGCCGAGCTGCA**CAAGAT**TCCGGTGATGCTGATCGTGGG  
CGACAAGGAGCAGGAGGCCCGCGAGGTGAGCGTGGCGAGCGCACCCGGCGAAGGCACCAAGGAGCGCAAG  
GGCGTGAAGTTTGACGAGCTGAAGCGGAATTGCTGGAGCGGCGGAAGAACCAGGTTCTTAA

40 1946 **2.1 observed, 2.15 ± 0.5\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: 1028561-1030057 cysteine—tRNA ligase

ATGACCCAGCCCACCATCACCCAGGACACCCGCCCGCAGCCCAACCCCAACATCACCCCTCTACGACACCA  
TGCAGCGCCAGAAGGTGCCTTTTCGTGCCCGGCACGCCCGGCTACGTGGGCATGTACCTGTGCGGCCCGAC  
CGTCTACTCCGACGCGCACCTGGGTACGCG**CAAGA**AGGAAGTCGCTTCGACGTGATTCGCCGCGCCCTG  
ACACATTTCCGGTATCAGGTGCGCTACGTCTCCAACATCACCCGACGTAGGCCACCTGCTGAACGACGCCG  
ACGAGGGCGAGGACAAGTTGCAGGCGCGGCCCGGCTGGAGCAACTGGAGCC**AATGG**AAGTTGCCGACAA  
**GTACT**TCTGGTGCCTTTTCCGCGATATGGACGCCCT**TGAATG**TCTTCGCGCCCCAGCATCAACCCGCGCGCC  
ACCGGGCA**CATTCA**GGGAGCAGAT**CAAG**CTGATCGAGGAACATCGAGAAGGGCCACGCCCTACGAGTCGG  
CGGGCAGCGTCTACTTCGACGTTCCGGAGTTGGCCCGAGTACGGCAAGCTCTCGGGCCGCAAACTCGACGA  
CCAGGAGGAAGGCACCCGCGAGGCCGTGCGCGACGAGAAACGCGACCCCGCGACTTCGCGCTGTGGAAG  
AAGCCGAGCCCGAGCACCTGATGCGCTGGGACTCGCCCTGGAGCGTGGGGTTTCCCGGCTGGCACAT**CG**  
**AATGCTCGGCCAT**GAGCCTGAAGTACCTCGGCGAGGGCTTCGA**CATTCA**CGGCGGTGGCCCTCGACCTGCA  
ATTTCCCCACCATGAGCCGAAATTGCCAGGCGGAGGCGGGCCACCCTTCGCGCGCT**FACTG**ATG  
CATAACAACATGCTGACCATCGGCGCGAG**AAGATG**TC**CAAG**AGCAAGGGCAACTTCACGA**CCATTCA**GG  
ACATCCTGAAAAATACGACCC**ATGGT**GGTGCGCTTCTGCTGGTGAGCAGCCACTACCGCAGCGTGAC  
CG**AGATGAA**CAGGAAGCCTTTGCGAGCGCGGCGAACGGCTACCGCCGCTCAGCGAAACGCTGCACGAA  
ATCGAGCGCCGCTTGAAGGACGCGCCTGCTGGCTCGGAC**ACTGC**GCTGGACAGCAAAATCGCCGCCCGCG  
TGACCGAGTTCGAGGAC**GCCAT**GCGCGACGACT**TTCA**ATACGCCCAAAGCGGTGGCCCTACTGTTCCGGCT  
GACCGGCGAGCTGAACACCGCGCTCAATGCCGGGCCAGTGGGACGCGATACCCCTGGAGCGCGCGCGAC  
GCCTACCGCAGCCTCGGCGGCGACGTGCTGGCCCTCTTTGCCGAGACAGGCAGCGCGGCTTCGGTCGCCC  
AGGACGACGCCTCAGTGATTGACGCCCTGATGGACCTGGTCTCAAAGCCCGCCAGAACTACCGCCTGCA  
AAAGCAGTACGCCGAGGCCGACGAGCTGCGCGAAACGCTGGGCAAGGCGGGCATCACGGTGGAAGACACC  
AAAGACGGCGCGCTGGAAACGCTGA

26 1493 **1.7 observed, 2.15 ± 0.6\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c1288720-1286675 methionine—tRNA ligase

GTGCAAAATCCCCCAGCACCCCTGAGGCCAGTCCCCGAGACCCGGGACCGTGAGTTTTTCATCACCG  
CC**GCCAT**CGACTA**TGCCA**ACGGCACCCCGCACATCGGGCACGTCTACGAA**AAGAT**TCTGGCCGACGCGAT  
TGCCCGTTACCAGCGCCTCGCCGGGCGCGACGTGACGTTTGTGATGGGCACCGACGAGCACGGCGAAAAA  
ATCAGCAAGGCGGCGCCAAAGGGGGCGTGACCCCGCAGGAGCTGGTGGACGACCTCTCGGAGCGGGCCT  
TTCAGGGCCTGTGGA AAAAGCTCGGCATCAGCTACGACTTTTTTATCCGCACCACCTCGGCCAAGCA**CAA**  
**GAA**ATATGTGCAGGACGTGCTTTCAGCGGGTCTACGATGCGGGCGACATCTACTTTGCCGAGTATGAGGGC  
CTCTACTCGGTGGGCGCCGAGCGCTACGTGACCGAGAAGGAAGTGGTGGAAAGGCCCCGACGGCGTCCGGC  
GCTTCCCCGGCGACAAGGACCCGCCCGAGCTGCGGCGCGAGGCCAACTACTTCT**TTCAA**CATGCAGAAGTA  
CCAGCCCTGGCTGCTGGAGACCTTGCAGCAAAACCCGACCTG**ATTCA**GCCCCGCGGCTACCGCAACGAG  
GTGCTGG**AGATG**CTGAAAGAGGACATCGGCCACTGAGCATCAGCCGCCCAAGGCGCGGGTGCCTGGG  
GCATCGAG**CTGCC**CTGGGACACCGACCACGTGACCTACGTGTGGTTCGATGCGCTGCTGAGCTACCTGAC  
GCCGCTGGTGAGCCAGGGGACGAGACGCCAGCATGAGCGGCAAGGCGTGGCACGTCATCGGCAAGGACATC  
**TTCAAG**CCGCACGCGGTGTTCTGGCCGACCATGCTGCGCGCCCGGG**CTGCC**GCTCTACCGCCGCTTGG  
TGGTGCACAGCCACATCCTCGCCGAGGACGGGCG**CAAGATGGGCAAGA**GCCTGGGCAACGCGATTGACCC  
CGAAGA ACTGGTTGCCGCGTGGCCGGTGGAC**GCCATTC**GGTATGCGCTGCTGCGCGAGGCGTGCCTGGGG  
GCCGACAGTCCCTTCGGTGAGGGC**GTACTG**GTTAGCCGCTGAACTCGGACCTCGCCAACGACCTCGGCA  
ACCTGCTTTCGCGCACCGTGAGCATG**ATTCA**GAAGTACAGAGGCGGCGTGATTCCGGCAGCGACCGAGCC  
GACCGACCGCAACGCGAGATTGAGGCGGCGGCGCGCCCTCCAGACGAAGTGTGCGGCTGGTGGAC  
GAACTGAAAATCAATATGGCGATCGACGCCCG**GATGAG**CTTCGTGCGCGACCTCAACCGCTACATCGCCG  
AGTCCACGCCCTGGACCTCGCCAAGTGCCCGAGACGCGAGGGCCGCTCGACACCGTGTCTACACCGC  
CGCCGAGGGGCTGCGCGTAGCGAGTGTGGCGCTCGAAGCCGTGATTCCGACCAAGGCGAAGGA**ACTGC**GC  
GAGCAGCTCGGCCTGGGCCGCAAGGCTACCCCTGCAAGCCGCTTGGGGCTGACCCCGCCGGCACCC  
GCGTGCAGGGCGGC**GCCAT**CCTGTTTCCAAGCCGAGCCGAAGGCGGACGAAACGAAAAACGCCGAGGC  
CAAGCCCCCAAGCCCCAAGC**CAAGAA**AGAGAAAAAGACCGTGACCGAC**ACTGCT**CCCGC**CAAGA**CCACC  
GAACAGAAGCCCGAAG**CTGCC**CCCC**CTGCC**CAGAACGACGGCCTGATTTGATCGACGACTTCGC**CAAGA**

TCGACCTGCGAATCGCCGAAGTCGTGGCCTGTGAAGCGGTGGAAAAGGCCGACAAGCTGC**TCAAG**CTGAC  
CG**TCAAG**CTCGGCGACGAGACGCGCACGGTGGTCAGCGGCATCCGCAAG**TGGTAC**GAACCCGAAGCGCTG  
GTGGGCCGCAAGGTCGTGCTGGTCGCCAATC**TCAAG**CCCGCAAGCTGCGCGGCATCGAGTCGCAGGGCA  
TGATTCTGGCAGCGGAGGACGACGCCGGGAACCTCGACCTGGTGGGAACCGAGCTGGAC**CTGCC**GAGCGG  
GACCAAAGTTCGTTGA

36 2042 **1.8 observed, 2.15 ± 0.5\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: 1365358-1366641 histidine—tRNA ligase

ATGGCG**ATT**CAGCGTCCCAAAGGCACCAACGACCTC**CTGCC**GAGCGGCAGTCC**CAAGA**CCGAAGCGT**TTC**  
**AA**CGGGCCGCGGCCACGAGTGGCTGATCGCGCAGGCCCGTGAAGTGCTCGAACGCGCCGGGGCGCAGCG  
CATCGACACGCCGATGTTTCGAGGAAGCCGAACCTCG**TCAAG**CGCGGGGTGGGCGGCTCGACCGACATCGTG  
CGCAAG**GAGATG**TTTACCCTACTACTTTCGGCGACCACGGCGGCTACGTGCTGCGTCCCAGGGGCACGG  
CGAGCATCGTGGGGCGTATCTGGAGAACGGGC**TCAAGCAACTGCC**CGCGCCG**TCAAG**CTCTGGACCTC  
GGGGCCGATGTTCCGCGCCGAGCGGCACCAGAAGGGCCGCTACCGCCAGTTCACGCAGGTGGATTACGAG  
GTGCTGGGCTCCGCCGACGCGCTCGTCGACGCCGAGGCGATTGCGCTG**ATGGT**GGCGGTGGTGC AAAAGC  
TGGGGCTGCGGGCGTGC**CGCTCAAG**CTCGGGTCCATCGGCGACCCGGAGGACCCGAGCGCTACAACGC  
GTACCTGCGTGATTTGTTCTCGCCGCACTTGGGGCGCCTCTCCGACGATTCCAAAGACCCGCTGGAGCGC  
AACCCGATGCGGATTCTGGATT**CAAGAG**CGCGAGTGACCAGGA**ACTGC**TGGCCGAGCTCAGCGTGCGGC  
CCATGCTCGACTTCCTCGGCGAAGGCGCCGCCGCGACTTCCGGCAGGTGCAGGCCTACCTGAGCGACTG  
GGCGTGGCGTTCGACCTCGACCCGAGCATCGTGCGCGGGCTGGACTATTACCGCCGACCCGCTGGGAA  
CTCCACTACGAGGGCATCGCGCCAAAGTCGGCCCTGGGCGCGCGGGCGCTACGACGGCCTGAGTGAAC  
AACTCGGCGCGGCAGCGTGC**CGGGCATCGGCTGGGCGTTCGGCGTTCGAGCGGCTGTTGCTGGCGCTCGA**  
GCAGGAAGGCATCGAGATTCCGGGCGAGCGGGCGGCCGAGCTCTACGTCGCGGCGCTGGACGAAGAAAAC  
GTGCCGCTGGCGGCGCGGGTGGCGCTCGCGGCCCGCGCTTTCGCCGCGCGAGTTCGCTACCGC**GCCA**  
**TGAA**ACCGGGCAGCGCT**TCAAG**GACGCCGAGCGCGGGGCGCGCTGTGGGTGGGCCTCATCGGGTCGGA  
GGAAGCCGCCCGGGACCCTGAGCCTGAAAAACCTGCACACCGGCGAGCAAAAGACCGTGCCGGTCGGG  
GAGCTGGGACAGGCCCTTGCTTGA

18 1280 **1.4 observed, 2.15 ± 0.65\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: 1367030-1368760 aspartate—tRNA ligase

**ATGAA**ACGCACGTCCCTGATCGGCCAGCTCGGCCAAGCGCAGCAACAACAAACCGTCAACCTGCAAGGCT  
GGGTCAGCCCGCCGCGACCTGGGCGGCCTGATTTTTCTAGAGCTGCGCGACCGCTCCGGCACCGTGCA  
GGTGCAGGTCGAGCCCGATTCTCCCGCTTTTGCCGAGGCCGACCGCCTGCGCGCCGAGTACGTGGCCGAA  
ATCGAGGGCACCT**TCAAC**CCCCGAGTCGAGCGCAAGGGGGCCTGGCCGACTTCGAGGTCAT**FG**  
**CCAG**CCGGT**CAAG**GTGTTGAACCGCGC**CAAGAC**CGCCGCTTTTGAAGTTCGACAAGGGCGAGCGTGGC  
TGAGGATATTTCGACTGAAGTACCGCTACCTCGACCTGCGCCGCGCCG**AGATG**CAGCGCGCCCTGATGCTG  
CGCTCCAAGGCCGTGACCGCCGTACCGAATTTCTCGACGCCGAGGGCTTT**ATTCA**GGTGC AAAACGCCGA  
TGCTACCAAGTCCACGCCGAGGGCGCGCGTGACTTTCTGGTGCCCTCGCGCT**TGAAT**CCCG**TGAAT**T  
CTACGCG**CTGCC**GCAGAGCCCGCAGCTG**TTC**AGCA**ACTGCTGATGAT**CGCGGGCTTCGACCGCTACTAC  
CAGCTCGCCCGCTGCTTCCGCGACGAGGACCTGCGCGCCGACCGCCAGCCGACTTCACGCAGCTCGATA  
TGG**AGATGAG**CTTTGTCGAGCAGGACGACGTGCTCGAAGTGCAGGAGCGGCTGCTCGCGCACGTGTTTCAG  
GCGGGTGCTGGACGTGGAGTTGCCTCTCCCTTTCCGCGCATGAGCTACTTCGACCGGATGGACCGCTAC  
GGCTCCGACAAGCCGGATTTGCGCTTCGACTCGGCCTTTACCGACGTGACCGGGCTGTTCCGGGGCGGGG  
AATTTGCGGCGTTTGCAAGTGCCCCGAGCGTCAAAGTCCCTCGTCGCGCCCGAGCTGACCGCGCAAGCAGAT  
CGACGAACCTCGAGCGCGTCGCCAAGCAAAACGGCGCGGGCGGTCTGGCCTGGCTGCGGCGTGACGGTGAA  
GGTTTACGGGCGGCATCTCAAAGTTCGTGGGCGGGATTGCCCGCAGCTGATCGAGCAGACCGGCGTGC  
CGAGGGCGGCACGCTGCTTTCGCGGGCGGGCGAGTGGAAAAAGCGGTACGGCGCTGGGAGCGGTGCG  
CCTCGCGCTGCGGGACCTGTTTCGACCTGGCGGGCGGGCGGGCCGAGTTTCATGTGTCTTGGTGGTGGAC  
TTTCCGCGAGCTCGAATTCGACGAGGACTCGCAGAGCTGGACCTACATGCACCACCCCTTCACCGCGCCG  
ACCCGGGCGACGTGGCGCTGTTCCGCACCGAGCGGCAGGGCG**AGATG**CGCGCGCAGGCCTACGACCTCGT  
**GATGAA**CGGCTTTGAAATCGGCGGGCGGCTCGGTGCGT**ATTCA**CGACCCGAGGTGCAGGCG**AAGATG**TTC  
CAG**GCCAT**CGGCTTTTCCGAGGAGGCTGCGCGCGAGAAGTTCGGCTTCTTTCTCGACGCGTGGAAACG  
GCACGCCTCCGCACGGCGGCATCGCCTGGGGCTTTGACCGCTGCTGATGCT**GATGA**GCGGCGGGCAG  
CATCCGCGAAGTCATCGCCTTCCCCAAAAACAACCGCGGCGCGGATCTGATGGCGCAGGCCCCCTCCCC  
GTCGAGGATGCGCAGCTCGCCGAAGTGGGCGTGCAGGTTTCGGGGCGAATAA

25 1727 **1.4 observed, 2.15 ± 0.6\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: 1378406-1381642 isoleucine—tRNA ligase

ATGACCAAACGTACCTTCGCCCTGTCCCCACGCAGCCTAACTTCCGCCAATTGGAAGCCGACATCCTCC  
AGAAATGGAAGGATGAGCAGGTCTTCGAGCAGACGCAGACCCGCCCGCTCCGGCAGGCGAGTTCGTGTT  
CTATGAAGGGCCGCCACCGCAACGGCAAGCCCGCCCTCCACCACGTGCTCGCCCGCTCGTTCAAGGAC  
CTCTTTCCCGCTTCAAGGTGATGCAGGGCCACCACGTACCCGCAAGGGCGGTGGGACACCCACGGCC  
TGCCGGTGGAAATCAGCGTGGAAAAAAGCTCGGCTGGCTGGGCGCAACCACGGGGCGAGCCGCGAGGA  
ACTCGAGGAATCAACCCTGTGCCGCACGTCCGTCTGGGAAA**CCATTCA**GGACTGGAACGAACTACC  
GAGCGCATGGGCTACTGGCTCGACCTGGGCGACCCCTACATCACCTACCAGAACAGCTACGTGGAGTCGG  
TGTGGAACTTGCTGCGGCGCCTGCACGAAAAGGGTCTGGTTCGCG**CAAGAT**TACAAAAGTGGTGCCGCTCTC  
GCCGCGCATCTCCTCTACCCTGTGCGGTGCCGAGCTCGGTGAAGTGGACTCTTACC**GCATGGT**GGACGAC  
CCGAGCGTGTATGTTTCGCCTCCCATCGTCTGGGACACGTTGCCCGAGCGGGCACACGCCCACTCAGCA  
GTCTCAGTGGTGAGCAGCGCCAGGGCCTGAGCCTGTGGTGTGGACGACGACCCCTGGACG**CTGCC**GAG  
CAACACCCTGG**CTGCC**GTCAACCCGGATCTCGACTACGTGGTTCGCCGACAGCCCGAGCGGGCGCGTCA  
GTCGCCGAGGGCGCGGTGGAACGGCTCTCGGCGCTGCACAAGGACGCGGCTCCGCTGGAAGTCTCGCGC  
GTT**TTCAAG**GGTTCGCGATCTG**GAATGG**GTGGAATACGAACCGCCCTTCCCCGAAGTAGCGTCTCAACTCGG  
CGTGTGTCGGAGCTCCACGAGCGGCGTGACGGCAAGCCGGTCTTGCACCTTCGTGGTGTGATGGCCGACTTC  
GTGAGCGATGTGGACGGCTCCGGGGTGGCGCATGAGGCGCCCGCTTACGGCGCGGAGGACTTGGAGGTCG  
CCCGCGGTACGGCGTGGCGCTGATGTTTCGGGGTGGACGACACGGCATCCTGCAAGTACGCACGAGCA  
AGGCAAGTTC**TTCAAG**AGCAGCCGACAGGGCCCTCATCGCCGAC**ATGAA**GGCGCGCGGCTGATGTTCTGG  
GCC**GGTACT**TGAAACACCGCTATCCCTTCCACGACCCGACCGGGCGACCCCATCTCTCG**CAAGA**  
AGGGC**TGGTAC**ATCCGCACGTTCGAGGTGGCGGGCG**AGATG**CTGGCGCAGAACGAA**AGAT**CAACTGGGT  
GCCGGGCAAC**TTCAAG**CACGGGCGCTTCCGCAACTGGCTGGAGGGCAACGTGGACTGG**GCCAT**CTCCCGC  
GAACGC**ACTGGG**CACGCCG**CTGCC**CTTCTGGCAGTCCGAGAGCGCCAGCTGCGCGTATCGGCAGCG  
TGGCGAACTCTCCGAGCTGGCGGGGCGGACCTGAGCGACCTCGACCTTACC**GGCTTAC**ATCGACGA  
CATCACCTTTACGCTCGACGGTGGAGGATACCGCCGCGTGGCCGAGGTGCTCGACGTGTGGTTCGACTCG  
GGGTGATGCCCTACGCGCAGTGGGG**ACTGCT**GCTGAACGAACAGGGCGAAGCGGTGCGCGGCGCCGAGC  
AAT**TTGCCA**AGCACTATCCCGCGACTACATCTGCGAGGCGATCGACCAGACGCGCGGGTGGTTC**TACTC**  
GCTGCAC**GCCAT**CTCGACGATGCTCTACGACCAGCCGGCGTA**CAAGA**ACGTCAT**CTGCC**TGGGGCACATC  
GTGGACGAGAAGGGCCTG**AAAGATG**TC**CAAGAG**CAAGGGCAACGTGGTGGCGCCG**CTGCC**GCTCTTCGACC  
AGTACGGCGCGGACTCGGTGCGC**TGGTAC**ATGTTTCATGGCGTCCGACCCCGGCGACCAGAAGCGCTTTTC  
CGAGCGTCTGGTTCGCCGAGGCGCAGCGCAGTACGTGAACACGCTGTGGAACGCTATTCGTTCTTCGTC  
CTGTACCGCAACCTCGACCAACCTGACCTCGCCGCTGCTCCCGCGTGGCCGAGCGGGCG**AGATG**ACC  
GCTGGCTGCTTCCCGCCTGGAAGAAACGGTGCAGCGAGCTGACGGCGCGCTCGAAAGCTATGACGCCCG  
CAGTGGGGGCGCGCCCTGGAAGGGTTCGTGGACGACCTGAGCAAC**TGGTAC**GTGCGCCGAGTCGCAGC  
CGCTTCTGGGGTGGGGCGGCACCGTGGACACCGCCGCTACGCCAGCTGCACGAGGCGCTGCTGGTGG  
TGTGCAACTGACCGCGCCGTTACACCTTTCTGGCCGACGCGCTTACC**GCAAC**CTGTGGGGCGAGGA  
GTCGTGGTGCACCTGACGCCGTGGCCGACCGTCCCGCGCCGAGCGGCTCGACCGCAAGCTGACCGCCGAC  
ATGGCCGCGCT**GATGAA**AGTCTGTGGAAC**TGGG**CGCGCCGCTGCGGGGGCGCACAACCT**CAAGA**CCCGGC  
AGCCGCTCGCCGGGGTGC**AGGTAC**GCGCCGCTCGCCCGAAGCGCTGGACGCGCTGAAGCGCTCGCAGAC  
GCAGATCATGGAAGAACTGAACGTCAAAGCCGTGACCTTCTTAGAAGGCGACACCGACCTCGTGCAGTAC  
AGCCTGCGCCCGAAC**CTGCC**GGTGGTGGGCAAGCAGTACGGCAAGCAGCTTCCGGTGTGAAAAAGGCC  
TGACAGAGGCGGACGCCCGCGCCGTCGCCACGGCGGTGCAGGCCGCTCAGGGCTTACGCTGCAGGCCGA  
CGCGGTGACCTTCGACCTGACGCCGGGAGCGTGTGGTGGACGCCAAGGCGCCCGAAGGTGTGGCGG**CT**  
**GCC**GAGGACCTGGGTATCTGGTGGCTTCGATACCGCCGCTGACCCCTGAAC**TGGTAC**GCGAGGGGCTGG  
CCCGTATCTCGTT**CGCATTCA**GAGGGCCGACGGCCGAGCGGGCTTCGAGGTGCAGGACCGGATTC  
CCTCGCGCTGGAAC**TGGATGG**CGAGCGCTGGAAGCGCGAGGCGTGGCAGGACTT**CATTG**CCGGT**GAG**  
GTGCTGGCGGAGCAGGTGGCCTACGGCTCCGGCGAAGGCTTCCGCGCCGAGGTGGAAGGCGGAGCGGTGA  
CGC**TTCAAGA**AGCTTTGA

65 3234 2 observed, 2.15 ± 0.4\* expected

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: c1620566-1619541 tryptophane—tRNA ligase

ATGACCACCCCGACCC**CTGCC**GCTACAC**CTGCC**CGCCCCCGCTCCTGACCGGGGACCGGCCACCGGGG  
CGCTGCACCTCGGGCACCTCGCCGGGTGCGTGCAAAACCGGGTGCCTTG**CAAGA**CAGAGCCGAGCTGTT  
CGTGTGCTTGCCGACGTGCAGGCGCTCACCGACCACTTCGACCGGCCCGAGCAGGTGCGCGAGAACGTG  
CTGGCGGTGCGACTCGATTACCTCGCCGCGGCTCGACCCGAGAACGACCTGCGTGGTGCAGCTCGG  
CGGTGCCCGGAACCTCGCCGAGCTGACGGTCTATTTCTCAACTTGGTACGGTGTGCGACCTGCACCGAA  
CCCCACCG**TTCAAG**GCCGAAATCGCGCAAAAGGGCTACGGCGAGCGGGTTC**CTGCC**GGGTTTTTCGTGTAT  
CCGGTGTGCGAGGCCCGGACATCGCCGCGTTCGGGGCCACGCTGGTGGCGGTGGGCGACGACAGTTGC  
CGATGCTCGAACAGACCCGCGAAATCGTGGCGGCT**TTCAA**TGCGCTGTATGCCCGGGTGTGCGCGAGCC



GCAGGCGCAGCTCAGCCGGGTGCCGCGT**CTGCC**GGGGCTCGACGGGCAGGC**CAAGATGA**GCAAGTCGCTC  
GGCAACGCGATTGCGCTCGGCGACT**CTGCC**GACGAGGTGGCGCGCAAGGTCATGGGCATGTACACCGACC  
CTGGCCACCTGCGGGCAAGTGATCCCGGACGGGTGGAGGGCAATCCGGTCTTCACCTTCTCGACGCCCTT  
CGACCCCGATCCGGCGCGGGTGCAGGCGTTGAAAAGACCAGTACCGCGCGGGCGGCCCTCGGTGACGTGAAG  
GTCAAAAAGCACCTGATTGACGTGCTCAACGGGGTGGTGGCTCCCATCCGCACGCGGCGCGCCGAGTATG  
AGCGCGACCCGACGCCGTGCTCCGGTTCGTACGGAGGGCACGGCGCGGGGCCGGAGGTGGCGGCGCA  
GACGCTCGGGCAGGTGCGGCGGGCGATGCGGTTGTTTGGGCACTGA

12 1022 **1.2 observed, 2.15 ± 0.75\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c1446067-1444568 proline—tRNA ligase

ATGACGAA**AAGATG**CGCG**CAAGA**AAGACAACCAGGGCCAGGA**CAAGA**AGGCGCAGCAGTACGGCGTCACCC  
CCCAGAGCGTAGAT**TTCAAC**GACT**TGGTAC**AACGAAGTGGTCAAAAAGGCCGACCTCGCCGACAACCTCGCC  
GGTGGCGGGCGCGAT**TGGT**CGTGC GGCCCTACGGCTCGGCGCTGTGGGAAAA**CATTCA**GCGCTGGCTCGAC  
GACCGCT**TTCAAG**GCGAGCGGGCACGAATCGCTGATCTTCCCGACCCT**CATTCCATGAA**CTTCATC**ATGA**  
**AGGAGGCCG**ACCACGTGAGGGCTTCGCGCCCGAGCTGTTACGGTGAA**CAAGAT**CGGCACCGAGGAACT  
GGCCGAGCCCTACGTGATGCGCCCGACCTCCGAGACGATCATCGGGCACATGTGGAGCGGCTGGCTCAAC  
TCCTACCGCGAC**CTGCC**CTTCTGCACTACCAGTGGGGCAGCGTGTTCGCGCCGAGCTGCGGACCAAGG  
CGTTTCTCCGCACCTCCGAGTTTTTCTGGCACGAGGGCCACACCGCCACGCCGACGAGGCCGAGGCCCG  
CGCCGAAGTGCGCCAGCAACTCGACCTGTACCACGAGTT**CTGCC**GCGACGTGCTGGCC**CTGCC**GGTGGTG  
CGCGGCAAAAAACCGCTCCGAGCGCTTTGCCGGGGCGGTGCGCACCTACTCCATCGAGGGCATGATGC  
GCGACGGCAAGGCGCTGCAATCGGGCACCTCGCACTACCTCGGCCAGAACTTCAGCCGGGCTTCGATGT  
GAAGTACCAGACCCGCAACAGAAAGAGGAATTCGCGCACACGACCTCCTGG**GCCAT**CTCCAGCCGCATC  
ATCGGG**GCCAT**CATCATGACCCACGGCGACGACTCCGGGCT**GATGAT**GCCGCCCCGCATCGCGC**CCATTC**  
**AGGTGGTTGTGATTCCGGTGGGCGCAAGGACA**ACTTCGACC**AGATGGT**GCAGGAAGGCGAGAACTGGC  
CGCCGAGCTGCGCGCCAGGGCCTGCGCGT**CAAGG**TGGACGGGCGGACGGCGTGACCAACGGT**TTCAAG**  
TACAACGACTGGGAAC**TCAAG**GGCGTGCCGGTGC GCATCGAACTCGGCCCGCGGATCTGGAAAAGCGGTG  
TGCTGGTCTG**TAAGA**ACCGCCACAGCGAGGACAAAGAAACC**CTGCC**GCGCGCCGAAGCGGTGACGGCAT  
GAGCGCCCGTCTGGACACCATCCACGACTTCT**GATGAA**GCGGGCGACGGATTTCTGCTGGCGAACACC  
GCCGAAGTGGACAGCTACGACGCTTCCAGCGGAAATCGAGGCCGGTCACTGGGTCCGCGCTTACCCT  
GCGGCGAGCCGGCGTGCGAAAAGAGCA**TCAAG**GAAGACACCAAGGCGACCGCCCGCAATGTGCCCTTCGA  
CGACGCCGAGTTCTTTGCCGAGCGCGGCGAGGGTCAGTGCGTGAAGTGCGGCCAGCCGAGCGGTATGGC  
AAGCGGGTGTGTTCCGGGCGTCAGTATTGA

38 1496 **2.5 observed, 2.15 ± 0.6\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c2237914-2236463 glutamate—tRNA ligase

ATGTCTGCTCCAGCTCCCCGCGCGTGACCCGCATTGCCCCAGCCCCACCGGCGACCCGACGTCGGCA  
CGGCGTACATCGGGCTGT**TTCA**ACCACACCCTCGCCCGCAGAGCGGAGCCGCTTCATCCTGCGGGTGGAA  
GGACACCGACCGCAACCGCTACGTCCCCGACAGCGAAAAACGCATTTTCC**AGATGA**TGCAGTGGCTGAAC  
CTACCCCCGACGAAAGCCCGCTGCAAGGCGGTCCCAACGGGCCCTTACCGCCAGTCCGAGCGCTTCGACC  
TCTACGGCGACTACGCCGCTCAGCTCGTGCAGAGCGGGCACGCTACTACGCCTTCGAAACGTCC**GATGA**  
**ACTCGCCGCCCTGCGT**GAGGAAGCGCAGAAGGCCGGGCACGTCATC**GCCATTC**CCAGCCGTGACTTGGGC  
GCGGCGCAGGCGCAGGCGGGTGGACGCGGGCGAGCCGGCGGT**CATT**CGCCTGAAGGTGGACCGCGACG  
GCGAAACGGTGGTCAACGACCTGCTGCGCGACCCATCCACTT**TGCCA**ACAAGGAAATGACGACAAGGT  
GCTGCT**CAAG**GCCGACGGCTTTCCACCTACCACCTCGCCAACGTGGTGGACGACCGCTTAATGCAAGTC  
ACCCACGTGGTCCGCGCCGAGGAGTGGATTACCTCCACGCCT**ATTCA**C**GTA**CTGC**TTT**TACCGCGCCTTCG  
GCTGGCCGAGCCCGTGTTCGCGCACATGCCGCTGCTGCGCAACGCCGACAAATC**CAAGAT**TTCCAAGCG  
**CAAGA**ATCCCACCAGCGTGGAG**TGGTAC**CAGAACCAGGGTTTCTCCCCGAAGCGATGCTCAACTTCTG  
GCGACGATGGGCTGGACGCAACCCGACCGCGAGGAAATCTTCGACCTCGCCAAGCTGCGCGGTTTCGAGCGGGTCTTTC  
GTCTCGAGGACGTGACGCTCGGCGGCCCGGTCTTCGACCTCGCCAAGCTGCGCT**TGGTAC**AACGGCAAGTA  
CCTGCGCGAA**GTA**CT**GAG**CGAGGACGACGTGGCGCGGCG**ACTGC**ACGCCTTCTGATGCAGAACAAAGTC  
ACG**CTGCC**CAGCGTAGACGGCCCGCAGGACCCTTACTTCCGCGCGTCAACCCACT**GATGA**TTCCCCGCC  
TGGAAGTCTTCGCCGACTTCATGGAA**CAAGA**CGCTCTACTTCTGGTCCGAGGATTACCCGGTGAACGAGAA  
GGCGCAGAA**GCCAT**CGACGCGGGCAAAGAGTTG**CTGCC**TGAGCTGGCCGCCCGCC**TCAAGA**AC**CTGCC**G  
ACCTTCGATGCCGCGAGCAT**CAAGGAGATG**TTCCACGCCTACGCCGAGGAAAAGGGCCTG**AAGATG**GGTA  
AAGTGATGCCCCCATCCGCGCCGCGTCCGCGGGCACGATGGAAAGCCCCGAC**CTGCC**GG**AGATG**CTTGA  
AGCGCTGGGCCGCGAGCGCGTGGCGCGGGTGGAAAAAGCCGCCCGCTGA

38 1448 **2.6 observed, 2.15 ± 0.6\* expected**

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: c2354014-2352473 lysine—tRNA ligase

CTGCACGAACAGACCATCGCCCGCCTGAATAACCTGGACGCGCAGGTTGACGCCGGCTTCGAGGCCTATC  
CCTACAGTTACCCCCAGACCCACCACGCCCGCAGCTGTTTCGCCGCTCACCCTGCCAAGAGAAGAGGGCAC  
AGGGGAAGGGGAACGCTGGAACCCGGCCAGAAGTGGGAAGAAGAAAGCTACTCGCTCGCGGGCCGCGTG  
ACCCTGATGCGCCACATGGGCAAGGCCGCTTTGCGGACCTGAGCGACGAGGACGGCAAAATCCAGCTGT  
TTTTCTCCAAGCAGGACACCGCCGGCTTCGACGCCACCAAGAAAATCGACCTCGGTGACATTATCGGCGT  
GAAGGGTCACCCCTTCGTGACCAAGACCCGGCAACTGACCCTGCATGTGACCGAGTGGCAGCCGCTCGTC  
AAGAGCCTGCACCCGCTGCCGAGCAAGTTTCACGGCCTGCAAGACGAGGAAGTGCAGCCCGCCGGCCGCT  
ACGTAGACCTGATGGTGACCGAGGGCGCCCGGAGAAGTTCCAGGCCCGCAGCCGCATGATTCGGTACAT  
CCGCAACGAACTGGACGAACCGGCTTCATGGAAGTGGAGGGGCCACCCTTCAGGTACACGGCGGGCGGC  
GCCGAGGCCCGCCCTTCATGACGCACCACAACCGCTCTCGTACGATTTCAAGTTGCGCATCAGCCTGG  
AACTGTACCTCAAGCGGCTGCTCGTCGGCGGCTTCGAGCGGGTGTACGAGATCGGGCGGGTCTACCGCAA  
CGAGGGCATCGACCGCACCCACAACCCCGAATTCACCATGCTGGAAGTGTACTGGCCCTACCGGACTAC  
AGCGACATCGCAAGTTGGTTCGAAGACCTGCTGAGCGGGCTGGCAAAGGAAGTGCACGGCTCTTATCAGT  
TCGAGTACCAGGGCAAGACGCTGGACTTTACCCCGCCGTTTGCCCGTGTGGATTACCTCGGGCGGCTGCG  
CGAGCACGTGCCGGGCTTCGACTTCGACCCGCTGGACCTAGACCCGGCTGCGCGCTTTCTGTGACGAGCGC  
TTTCCGAGTGAAGGGCGTGGCGAGCTACAAGCTGCTGGACAACTGTTTCGGTGTGATTCGTGGAGCCGC  
TGCTGAGCAATCCGACGTTTCGTGATGGACCACCCCGCGTGATTTCCCGCTCGCAAGACCGGTTTC  
CCGCCCGAGGCCGTCACCGAGCGTTTCGAGGTGTTCTGCTCGGGCTTCGAACTGGCGAATGCGGCTTC  
GAACTCAACGAGCCCTTCGACCGCGCAACGCTTTGAGGCGCAGACTGCCCGCCAGCCCGCGGAGACG  
ACGAGGCGCACCCGAGGACGAGGACTTCTGCTCGCGCTGGAATACGGGATGCCGCCCGGGGGCTT  
GGGCATCGGCATAGACCGGCTGGCGATGCTGCTCACCAGCTCGGATTCCATCCGCGACGTGCTGCTGTT  
CCGCTGCTGCGGCCGAAGGCGGCGAGGCGGAAGAAGCTGACGACACCGCGCAGGAAAACACCGCCGGCTGA

23 1538 1.5 observed, 2.15 ± 0.6\* expected

Deinococcus radiodurans R1 chromosome 1, complete sequence GenBank: CP015081.1: 2574688-2577429 valine—tRNA ligase

ATGACCGACCCCCACACCACCACCTCCCCACGCAGTTCGACCCACCGGCATCGAGCCGAATGGGCCG  
CCCGCTGGCGCAGTGAACCGTTCCGCGCTGACGCCACCAGCGGCAAAGACCCCTTTACCATCGTGATTC  
GCCGCCAACGTGACCGGCAACCTGCACCTCGGACACGCGCTCGACAACACGCTCATCGACACCCGTGAT  
CGCTACAAGCGCATGGCGGGCTTCGAGGCGCTGTACTGCCGGGCATGGACCACCGGGCATCTCCACGC  
AGGTGGTGGTGGAGCGGCAGCTCAAGGACGCGGGACCAGCCGCCACGACCTCGGGCGGAGGGCGTTCT  
TGAAAAGTCTGGGAGTGAAGGGCAAATCCGGCGGCATGATTTCTCGATCAGCTCACCCGCTGGGCGTG  
AGCGCCGACTGGACCCGCGAGCGCTTCACGATGGACGAGGGCCTCTCGCGCGCGTCCGCACCCAGTTTC  
TGAAGCTCTACCAGCAGCCCTCGCTACCGGGCGAGCGCATCGTGAAGTGGGACCCCGCCCTCACAGAC  
CACGCTCTCCGAACTGAAATTTGACCGCGAAGTGCAGGCAAGGGCAAGATGTACACCCTGCTACAAGCTG  
GAAAACCTCCGCCGAGCGCGGGAGCAACGGCGAAACTGGCGAAATCCGCATCGCCACCGTGCGCCCCGAAA  
CCATCTTCGCCGACCAGGCGATTGCTGTGCATCCCGAGGACGAACGCTTCCGGCATCTGGTGGGCCAGAA  
GGCGCGCATTCCGCTGACCGACCGCTGGGTGCCATCATCGCCGACGAAGCGGTGGAGATGAAATTCGGC  
GTCGGGGCACTGAAAATCACGCCCGCGCACGACCCACCGACTTTGAAAGTGGGCGAGCGGCACGGCCTGG  
AGCGCCCCAGCGTGATTGATCTCGACGGCAACCTCACCCGTGACGAGCTGGTGGCCGCGAGTTTCAGGG  
TCTGGAACGCTTCGCCGCGCGCAAAGCGGTGGTGAAGGCGCTGGAAGAATCCGGCGATCTGCTGGAGCAA  
AAGGACCACGACACCGCATCGGCCTCTCCGAGCGCACCAAAGTGCCTGGTGGAGCCATCATTTCCGAGC  
AGTGGTTTCGTGAAGATGAACCCCTTTGCCGACCAGGTCTTGGAAAGGACTGGATAAGGGCGAGATCAAGCT  
GGTGGCCGAGCGCTACACGAAGGTCAACCGTACTGGCTGGAAAACATCCGCGATTGGAACATTTCCCGG  
CAGTTGTGGTGGGGCCACCAGATTCGGGCGTGGTACGACAAAGAGGGCAACATCTACGTGCCCGACCCGG  
AAAACCCCGAATCGACTGCCGACCCAGGACCCCGCTACCGCGACCTCAATCTGCGCCGCGATCCCGACGT  
GTTTCGACACCTGTTTTTCGAGCAACCTGTGGCCCTTTTCCACCCTCGGCTGGCCTGACACCGCAGTGAG  
GACTTCCGCAAGTTCTACCCGACGAGGTGCTGGTGGACCGCTACGACATCCTCTTTTTTCTGGGTGGCGC  
GCATGCAGATGGCGGGCTACGGCCTGACCGGACAAGCACCTTTTCCACCGTGATGCTGCACGGCCTGTA  
TCTGGACGCCAAGGGCCAGAAGATGTCAAGAGCAAGGGCAACGGCATCGACCCGCTGGAAGTGTGGG  
CAGTACGGGGTGGACGCTGCCGCTTCACGTTCTCTCGACCGGGCGGGCAGGACATCAAGCACG  
ACGCCCGCGCTTCGAGCAGGGCCGCAACTTCGCCAACAAGCTCTGGAACGCCACCCGCTTTGCGCTGAT  
GCGTCTGGGCGAAGCCGAATCGAGGGGAGCGACGACCTCTCCGCTACGTCCGCGCCCGCTGACGCTG  
CCCGATGGCGTGCTGCTGCGGAGCAAGGACGTGTTGGCACAGCTCAAGGAGCGCGACGACCTCACCTTCG  
CGGACCGCTGGATTATCAGCCGCTGAATGACGTGACCGCCGAAGCCTCCGCGCAGCTCGATGCCTTCGA  
CATCGGCGCCGCATCCGACGCTGTACTCGTTACCTGGGACGAGTTCTGCGACTGGTACATCGAGGCC  
GCCAAGCCCGAATCGCCAGCGGCAACCTCGGCACCATGGCGACGCTGAAGGTGGTGTGGAGCACGTC  
TCAAGCTGCTGCACCCCTTCATGCCGTTTCATCACCTCCGAGCTGTACGCCCGCTGGGGCACCGCCCA  
GATTGCCGTGCACAGCTGGCCGAGCCCGACCGCTCTGCACGACGCGGAAGCCACCAAAGCTTCGAC

GCGCTGCGCTCGGCGGTGGACAGTGCTCGCAGCC**TCAAGA**GCGAACTGGGCTCTCGCCGACGACCGCC  
TGAACGTGGCGGTGGACGGCGACCTCGCTGACGTGGTGGCGACAGAACGCCCGGTGGTGGAGGGCATCGC  
CAGGGTGAACCTCGTGCCCGCTCTGGAAGGCCGCACGCTCTCGCAGGTGGCCCCGGCGTGACCATCCTC  
GCGCCGCTGGAAGGCACGGTGGACATCGCCGACTGGGTGAAAAAGCAGCAAAAACGCTCGCCGAACTCG  
ACAAGCAGAT**CAAG**CAGGCGCAGGGCAAGCTGAACAACGAGGGCTTCGTGGCCCCGCCCCCGCGAGGT  
GATCGAGGAGGAGAAGCGCCGGGTGGCCGACTTCGGGGCGCAGAAGGAGCGGTTGGAAGGAGTGTGGCT  
CAGTTGGGGTGA

34 2738 **1.2 observed, 2.15 ± 0.45\* expected**

*Deinococcus radiodurans* R1 chromosome 1, complete sequence GenBank: CP015081.1: c1437593-1436316 serine--tRNA ligase

ATGCTGGACCTGAAATTTATCCGTGAGAACCCCGACGCCGTGCGCGAG**GCCATTC**GCG**TCAAGA**ACGTGG  
CCCTTGACCTG**GATGA**CCTGTTGCAGCGTGACCGCGACCTCGTGCGC**TCAAG**CAGCGCGTGGA**GCCAT**  
GCAGACCGAGCGCA**TGCCA**ACGCCAAGCTGGTGCCCAAGGCCAGCCCCGAGGACCGCCCCGGCCT**GATT**  
**CAA**AAGGGCAAGGACCTGAGCGAAGACCT**TCAAG**GCGCTCGAGCCCC**ACTGC**GCGAGCAGGAAGACGCC  
TGAAGCA**ACTGCT**GTGCTGCGCGTGCCCA**CATTC**CG**CTGCC**GGGCGTGCCCGTCGGCAAGGACGAAGACGA  
CAACGTGGA**ACTGC**GCCCGGAAGGCGAG**CTGCC**TGGGTTTCGACTTCACGCCGCTCGACCAGGTTCGAGATT  
CTGAAAAGCAGGGCTGGGCCGACTTCGAGCGTGTGCCCCGCTCTCGGGCAGCCGACGTACCTGC**TCA**  
**AG**GGCGACGGCGCTGCTGGA**AATGG**CGGTGCTGATGTTTCGCGCTCGATTTCTGTGCGAGCGGGCTT  
TACTCCACTTTTCGACCACGGCGCTCGTCCGCCGGAACGCTGGTCAACTCGG**GCCATT**TCGGCGGCGAC  
GAGGAATCGGTCTACAAGCTTGAAGGC**BATGA****ACTGCT**GTGCTCGTGGAACCGCCGAGGTGCCGATCAATA  
GCCTCTACGGGGGAGCAGCTCAGCGCCGACGAGTTGCCGCTGACCTTCGCGGGTTTCAGCGCCGCTT  
CCGCCGAGGGCCGGGAGCGCCGGGCGCGACGTGCGCGGGCTGATTCGCGTCCACGAGTTCCGCAAGGTC  
GAGCAGTACGTCCTGTGCCGCGCCGACCAGGAAGAGGGCCTGA**AATGGT**TCGAGCGCTGCTGAGCAACG  
CCGAGGGGCTGTTGAGGCACTGGAG**CTGCC**TACCGCGT**CATTCA****GAACTGC**ACTGGCGACATGGGCGC  
GGGCAAGGTGCTGATGTACGACATCGAAACCTGGGTGCCGAGCGAGCAGAAGTACCGGAAAACCACTCC  
TGCTCGTACCTGGGCGACTGGCAGGCCCGCCGACCGGCTGCGCTACCGCGACGAGCACGGCAAGCTGC  
TCTACGCCACACCCTCAACAACACCGGCATCGCCAGCCCCGCG**CATTC**TGGTGCCGCTGCTCGAAAACCA  
CCAGCAGGCTGACGGCACCGTGCGCGTGCCCGCGCGCTGCGCCCTTACCTCGGCGGCAGGGAAGTGTG  
GGCAGCCGGTGCCTGA

32 1274 **2.5 observed, 2.15 ± 0.65\* expected**

**Observed mean proximity: 1.8 Expected mean proximity: 2.15 ± 0.55\* (\* 90%-  
confidence interval calculated from the mean sample size equal to 1913).  
Data from <https://www.ncbi.nlm.nih.gov/nuccore/CP021507.1?report=genbank>**

## AL-pentamer proximity for CRISPR Cas proteins

*Mucilaginibacter paludis* DSM 18603 chromosome, whole genome shotgun sequence GenBank: CM001403.1: 2879349-2879678  
CRISPR associated protein Cas2

ATGCGTTTTAA**ATGAAT**ACCGGATTTTGTGGATATTAGTATTTTTTGGAC**CTGCC**GACAGAAACAAAAAAG  
ACCGCAAATCTACAGTAGGTTTCGTAAGGATATT**ATGAA**GGATGGATT**TGGTA**TGTTCCAATCTCCAT  
CTACCTGCGGCATTGCAGCAGTCGTGAAAACGCTGATGTACACATCAAACGTGTGAAGAAAATACTCCA  
CCTAAAGGTCACGTCGGTATCATGA**CCATT**ACCGATAAGCAGTT**CGGAATG**ATGGAATTGTTCT**ATGGTA**  
AGGAAGAGAAGGAGCTTCCAGACACCCACAGCAATTGGAATTGTTTTAG

9 326 **2.8**

*Mucilaginibacter paludis* DSM 18603 chromosome, whole genome shotgun sequence GenBank: CM001403.1: c2880706-2879693  
CRISPR associated protein Cas1

ATGATAAACGAACGCTTCACTTCAGTAATCCGGCTTACCTGAGCTTAAAGGAAAGCCAGT**TGGTA**ATTG  
ATCTTCCTCACCTTAAAATTTTAGGGGAAAAGGAAAGTAAAAAGTC**GGTACC**TATTG**AAGAT**ATAGGAAT  
TGTT**GTA**CTGGATCACCAGCAGATCACCATCACCCATGGCTGCATTGCAGCATTATTATACAACAATTC  
GCGATCATCACCTGCGATCATACACATCACCCACCGGCATGATGCTACCAATTGACGGACATGATACGC  
AAAGCGAGCGCTT**CAGGTAC**CAAATCGATGCTTACAGCCACTTAAAGAAGCAGCTTTGGCAACAAACTAT  
CCAAGCCAAAATATTAACCAAGCGGCGGTTTTGGCAGATCG**TGGTA**TTCCCCACGATAATATGCTTAC  
TGGGCTAAGTCTGTACGTT**CAGG**CGATCCGGATAACTACGAAGGACGTGCCGCCGCTTAT**FACTG****AAGA**  
**ATG**TATTCCTAAAAAAGTGAATTTCTTTCGTGGCCGTGAGGGTGACCCGCCAAACAAT**FACTG****AATTA**  
**TGGT****TATGCCAT**ATTACGGGCTATTGTAGCGCGCGGG**CTGGTA**TGTTCCGGCCTGATACCTACATTAGGT  
ATACATCACAGGAATAAATAACGCCTATTGCTTGGCT**GATGA**CATTATGGAGCCTTACCGCCCTATG  
TTGATTTGATCGTTCTAA**AGAT**CATCGATA**AATGG**CGAGAATTTTTTAGAGTTAGGCAATCTATAAAATC  
TCAGTTATTAGGCAT**TGCC**ACCCTTGACGTACAATTTGAAAGGAATAGGAGTCCGTT**AATGGT**TGGTGTG  
CAAAGCAC**ACTGCT**TTCTTTAGCCAGGTGTTACGAAGGAAAAACAGGAAAATAA**ACTAC**CCCTATC**ATGA**  
**AGA**ATTTTAATCCTAAAAG**GC**CAATATCCCTAAGCCGGACGAAAAT**TTCAATGATGA**GGATACAGAAAA  
TACTAT**TGAAT**ACAATATTGGC**GATGA**GAATAG

31 1010 **3.1**

*Mucilaginibacter paludis* DSM 18603 chromosome, whole genome shotgun sequence GenBank: CM001403.1: 7594207-7594989  
CRISPR associated protein Cas6

GTGCGATTTAAATTAACATTGTTGCCTACGGGCGAGTAAAACGACCGTT**CCATTCAA**ATTATCAGTATGCC  
TTGCGGCGGTAATTTTTTAAAAGGTTGCG**CTGCC**GACGAGCAATACGCCAACTTTTACATCAGCAAGG  
TTATGCCCTGAACAATTACTCCAAACATTTTAAAGTTTTTTACCTTCTTAACCTGGAAGGTAGGTTTACG  
CCACAGCAAAATGCCCTGTTGCTGCAAGGTAATCAAACCGGTTTTGTGCTTTG**CTGCCA**CATGCCCGAAT  
TTGCTGCGCATTTAATAACCGGTGTTTTTACCAGATCAGTTGATATCAATAGGCGGGAGCGGGCCGTGGG  
CCGGTTTTAGGGTTGAGCAAATTGTATCGCTCCCG**CTGCC**TTTAAAG**GATGATGA**CGAAATACACACGATG  
CAGTTTTAAACCGCTTTTCGCCCTTGGTTGTAGGCCGGACCAACGCTAAAAGGCAAT**GATGA**TTATTTATCGC  
CAG**AGATGA**GGATTTTATACCCCTGCTACATATTAACCTGGCCGATAAGCTTGCCGTTGCTTATAATGA  
TAGTTATAGCGGGCGAAT**TGGTATCAAG**GTAAGC**ATGAA**CCCGAAAGCTAAAATCTCGCTGGTGACG  
ATTAAGGAAGGGACAGCGCCGAAACTAA**GGTAC**GTGGCTTTTTTCGGGTTTGCCTGG**AGATGACTGC**GC  
CTGGCAGGGTATCAATTTTGTGTTGGACGTGGGGCTGGGCG**GATGAAT**AGTATGGGGTTTGGTTGTGT  
GGAGGTGAAATAA

24 779 **3.1**

*Streptomyces virginiae* strain IBL14 Cas2 gene, complete cds GenBank: KY176009.1

GTGTATGTAGTCGTCTACGACACCCACGCCAAACGAAACGCACGGATCTGCGTAC**CTGCC**GCAAGT  
ACCTCCACCACGTCCAGCGCAGCGTCTTCGAAGGCCACCTCAGCGACGCCAACTCCGACACT**TTCAAG**C  
TGCAGTCGA**AGATG**CCCTCGACTTGGAGTACGACAACGTCTCTACACGTTCCAGCCGGCACCG**TG**  
**CCACA**ACGCCT**TGAATGG**GGAGCAGTCGAACCCGCCCCAGTGACATCCTGTGA

9 260 **3.5**

*Legionella pneumophila* strain L15-118 CRISPR-associated endoribonuclease Cas2 (cas2) GenBank: MF440347.1: 5115-5414

ATGCAAAT**AAGAT**TATACAAGGAATACATAGTTTCTCTACGACATCAGGGACAACAGCATCCGACGTACAA  
TTTACGAAAGGCTTAAAGACATTGGCTTAAACCCCG**TTCAA**AAATCAGTATTTTGGGGATATTTAAACGA  
ACCTGAAAA**AAGAT**ATTTAACGAACTTTCTAAAAAAGAACT**TGCCAAA****CAAGAT**ACCGCTTTTATTACA  
CCTTGCAAGCTGGC**TCAAGA**ATACGAAGTAAACCTCATTAATCTTGAAGT**GATGAA**CTTAGGAAACCGG  
ATGGACACATCATCGTATAA

10 296 3.4

*Tannerella forsythia* 3313 DNA, complete genome GenBank: AP013044.1: 2465356-2465619 CRISPR-associated protein Cas2

ATGTATGTGATTCTGGTCTATG**ACTGCG**GGCGAGAAACGGGTTGGGAAAATGCT**TCAAG**CTCTGTCGGAAAT  
ATC**TGAAT**TGGATACAAAATTCGGTTTTTGAAGGCGAAATTACAGAAGTTAAATTTAAAGAGTTAACTTT  
GAAGGCTGAGAAAATA**ATGAA**CAGC**GATGAA**GACTCTCTGATTGTTTT**TTCAAG**TGCA**CAAGAAA****AATGG**  
**TTGGATA**AGCAAATCTT**TGGTAA**AGAGCGGAGTAGTGTCGATAACTTCTTGTA

12 260 4.6

*Nostoc* sp. 'Peltigera membranacea cyanobiont' N6 chromosome, complete genome GenBank: CP026681.1: 5862085-5862366  
CRISPR-associated endoribonuclease Cas2

ATGTTTTTATATGTGATTGCTTATGA**CATT**CCTTGTGATAAACGACGTAAAAAAGTAGCTGATTTACTTG  
AAGGCTATGGACAACGGG**TTCA**ATATTCTGTCTTTGAGTGTCACTTACTACCGATAAAATA**TCAAGAT**TT  
ACGTCGTCGGTTAAGAAAAAAGTTGAAGTTAACT**GAAGAT**AGTCTACGATTTTACCCTTTGTCACGACAT  
ACCCTTTC**TCAAG**TAGAATCTTGGGGTGTGGG**GCCAT**CTGTGATTGAACCGCCTAGTTCGATAATCATTTAA

8 278 2.9

*Saccharolobus solfataricus* strain SARC-B chromosome, complete genome GenBank: CP011055.2: 2127278-2128213 type I-A  
CRISPR-associated protein Cas7/Csa2

ATGTGGATATCCTTCTCCGTTAGGTATTT**TGGTAA**ACGTAG**AAGAT**CTAAATAACGTAGAGTCTGCAGGGA  
ACTACGTAAGGCATAGGAGAGCT**CCATT**AGTTTTTAAGGATAAAGACTCCTACACGGTAACATATG**TGCC**  
**AGCAGT**TAGTGGAGAAATGATAGCACATGGGTAT**CAATGAAT**CTAGTCGAAGTACCGCT**TTCAA**AGAAAT  
CTACCAGTAGATAGTTTAGCTAAGCAGGGAATATTGATAAAGAGGGGTTCT**GATGA**TAAAGGTT**ATGAA**  
GCACTAAATGCACT**GATGAAA**AGGGTCCGATTACGAACTCTGCGTCATTAATGAGGATATAGTCGAGGA  
TGTTGCTGG**ATTCATGAAT**CCAAACAAGCTAGTTAAGAGA**ACTTCAA**ATGTAGCCTTTAGTTAC**ATGGT**T  
CCTGCAAT**AGATGC**AGTTAAAGCT**TTCAA**CAATATCTTCCCAATTTACGTTAGGTATGCTAATAAAGAAC  
**TAATGG**ACAAATA**CAAGA**ACGAGAA**CATTCAA**TCCCTATATAATATCGAAACAGCGAGCGCATCGTAT**GT**  
**ACTT**ACGGGTTACTTGAACGTTAACAGTGTAGGAAAGACACAAAACCTATCCGGTAAAAGAAGTTGATAAG  
AAAAAGGATAGGGAAAAGGCTGCATT**AGATGC**CCTTAATGTTAACTTTAACGCAATTCCTATTTGGCGCAA  
AGTTGACTAGATTTAAGCCCATAGTGGAGATAGAGGCCTTATTTGTTAGCGCTTCAGAGAAAGCCT**TTCAA**  
TCTACCACCCGTA**ACTGG**AGACATTAAGAAATACATAGACCTTGTTAACTCTACAACGGATTCTTTCGCT  
AAAATTT**TGAAT**ATTTAAACGGCCCGTAGTAAA**GTA**CTATCTAAAGGAGGAGAAAGGTAACGTC**CAAT**ACGC  
CAAT**AGATGC**GTTTTACGGTAATGTAA

29 932 3.1

*Saccharolobus solfataricus* strain SARC-B chromosome, complete genome GenBank: CP011055.2: 2120631-2121506 type I-A  
CRISPR-associated protein Cas4/Csa1

GTGTTCTTTACT**CATT**CAGATATGCTGTTGCT**TTCAA**AGAGGATAAAGAAGTTACCTAAGAACGT**GATG**  
**AGGAG**CTGAGGGGATGGAATTGGAGCGAACCACAGTTTACACTAGGTCCCTTTACAGGTATCAATTTTC  
CGAA**ATGGT****TACTGTAGTACT**CTGAGGAACGTTTACCTTAAGGTAAAGGGCTTTAGGGGAGAAATAGGT  
AGGCAAATAC**TTCAAG**GTTCACTTATCCACACGATTTACGCTATTGGCATTGAAGCAATCAAACGTTTTTA  
TTT**ATTC**AGGGAAAGCATTGATGGG**GTA**CTCTGAGAACGCT**AATGG**AGATG**AG**TTTTACTCCCTACT  
TAAGGACTTAAGGGAGGAAGAGGGGATTTACGCAAAGGTCTTATGGGATCACATTACGAACATTT**ATFCA**  
GCTGAACTGGATAGGGTTAGGAGTAAGTTCACTAACTTAACTAGGGATTCCCTGGTCTCCAAGTTGTGC  
CGTTTTACTGTTAGTTTCCAGTT**GATGGT**TCCCTTCTGGGCTTGACCAATTTAAGGGTCGACGCCTTTAT  
CCCTCACCTTCCCCTAATAGCGGAA**ATGAAG**ACTGGGAAATACAGATACACTCAGAGTTGTCCCTTGCA  
GGATACGCTTTGGCAATTGAGAGCCAGTACGAGATAC**CCATT**GATTTTGGTTACTTGTGTATGT**TACTG**  
TGACTGAAAAGGAGGTTAAGAATA**ACTGC**AAGCTGATTC**CCATT**TAGATTCCCTTGAGGAGTGAGTTCTT  
GGACATGAGGGATAAGGC**TCAAGA**CATA**ATGG**ATAAGGGTGTGGATCCTGGGATAGCTAAGGATTTGTGAG  
AGTGATTGCATGTTTTATAAGGTTTGTACCCTTAA

27 872 3.1

Methanoseta concilii GP-6, complete genome GenBank: CP002565.1: c2758294-2758007 CRISPR associated protein Cas2

ATGCTGGTCTGGGTGATCTACGACATATCCGAAAACAGATACCGCAGCCGGGTGGCT**AAGATCTGCAAGA**  
ACTACGGCCTGTTTTCGGGTGCAGAAGAGCGCATTTCTGGGAGACTTGAACCG**GAATG**AGTCCGACTCCCT  
GACTCTGGAGTGCAGGGCATTATTGACGAATCGG**ATTCA**GTCTACGTCTTCCCCATGTGCGAGG**ACTGC**  
TTTGAT**AAGATCAAG**CTGATCGGAGCAGGGTTCGATCGAGAGCTGGTGAGCGG**GATGA**CCATAACCAAGG  
TGTTTTGA

8 284 **2.8**

Desulfobacula toluolica Tol2 complete genome GenBank: FO203503.1: 1119557-1120546 CRISPR-associated protein, Cas1 family

**ATGAA**ACAATCCTTATATCTTTTT**TTCAA**ATAGTACCATCAAAAAGAAAAGACAATTTCTTTGATGTTTGAAA  
TTGACGGAGAAAAGAAATACCGTCCTGT**TGAAT**CAATTTGTGACCTTTTTCTTT**TTCAA**TA**CGAA**CATACTCT  
**GAAT**ACAAAATTGCT**TGAAT**TTTTCTGGGGCAACAAAAAATTCCTGTTTATATTTTTAATTATTACGGATTT  
TATTCGGTTCATTTGTCC**CAAGAG**GAGCAATATCTTTCAGGGCATGTCACCATCAGTCAGGCAAAAGCATT  
ATCTTGACGATAAAAAA**GAATGGT**TTTTGGCCT**ATGAAT**TAT**TGAATG**CCGCAGCT**TTCAA**ATATTATCAC  
CAATTTAACCTATTATAAAAAACAGGGAAAGAGATATCGGTGAAAAAATTTTA**AAGAT**AAAAGAGTTAAAA  
AAATCATTAAAAAATGCG**ATTCA**ATTTTCATATTT**AATGG**CATAGAAGGAAATATCAGGGATTTGTATT  
ATTCTGCTTTTT**ATGAA**ATTG**TTCAA**ATAAGGT**TGAAT**TTAAAAAACGTATTAAAAAGACCGCCGATAA  
TTTTATAAATTCCTGATTTTCATTT**TGGTA**ATTTCCCTTGTTTATACAACGGTTCGTCTGAAATTTATAGA  
ACCCAGCTTGATCCTACCGTGAGTTTTCTTCATGAGCCGGGTACAGGC**GGTACT**CCCTTGCCCTTGATA  
TATCGGAAATTTTTAAGCCCATCCTCCTTGATA**GAATG**ATCTTTACCATGATCAACAAAAAGAATTAAC  
GGAAA**AAGAT**ACGGATAAAGACCTTAACCTATTGTTATCTGAA**AAGAT**CAGGGCCG**GATGAA**ATTTTAAGA  
AAATAT**GATGAA**AGGCTGAAAAAGACTATTAACATAAGAAACTCAATCGCAATGTTTCTTACAAACGGC  
TTATACGGCTGGAGTGTATAAAATTTGTTAAACATATTTCTGGGTGAAGAAAAATATACCGGG**TTCAAGAT**  
**GTGGTGGTAA**

39 986 **4**

Desulfobacula toluolica Tol2 complete genome GenBank: FO203503.1: 1118078-1118929 CRISPR-associated protein, Cas6 family

ATGATTGCAGGCTTCACTAACCCCTGAAGCCTTCCGGAAAGGTATGATTTTGGGTATAAAAAATAACAGCAA  
TAACGAATCAACCGCCTTTTTTCCCTATAATTATCAATATGCAGTT**CATTCA**GCTTT**GFACT**CTTT**AT**  
**TCAAGA**ATCATCAGCAA**ACTATTCAAG**TTTTTTACATGATAAAGGGTATGTAA**AAGATG**GCATCAATAAG  
CTTTTTAACTATTTACCTTT**TTCAA**AGCTCAACTTTAGTCCTAAAAAT**AATGGGAATGG**CGGTTTTTATA  
AGGT**TTCAA**AAATACAATTTGTGTTTTCCACCATCATGGAAAAAAGCATGAGACATTTAATCCTCGGCAT  
ATTT**TTCAA**ATCAGAAAATGATGTTAACTTGAGCGGGCAAAGACATTTTTTTTGATATTGTAATGTTGAT  
**ATTCAGGAAGAT**TGTCAGTTTACCAA**TACTG**AAAAATTTATTTGTTTGTGCGCCATAACCGT**TTCAAG**CA  
TAATGATCAAAGAAAAAGTAAAAGGGTGCAACATTTTTTAACTAT**ATGGT**GCCC GCGGAAAGAGATCA  
TTTGTGTTGAAATTTAAAGAAAAATCTTATTAACAAAT**ATGAA**ACCATCAATAATATCCCAT**ATGAA**AT  
**ATGAA**ATTTCTTTTTCAGTTTTCATTTTGAAAACAACTATATCGTAAAAAGA**AATGG**CAGAATATCAAAAT  
TGAT**TGAAT**TTAAAAATAATATTAAGTCAAA**GCCAT**GGAAGCCCTTTTTTATATAGAGGCAGATCCTGA  
GTTAATTAATCGGGTATGATTGCGGATGGGGTGAAAAAAACAGCGCCGGTTTTTGGTTGTGTTGAAAAA  
GCAGCAAAATAA

30 848 **3.5**

Desulfobacula toluolica Tol2 complete genome GenBank: FO203503.1: 1118938-1119459 CRISPR-associated protein, Cas4 family

TTGAACCGGTTTTT**ATGGG**GAGAAAATATGGATAATATTAATTTTACCGGCACACAGATTAATTAATTATTT  
TTGTCTGTAAGCGAAA**ACTCTG**GTGTTTA**CAAGAG**CTATTTCGGTTT**GAAAGATGAA****ATGAAT**ATGTGCA  
GTTGGGAA**ACTCATTGATGAA**AGCACTTATAAACGCAATAAAAAAGCAGATTGAAATCGGCAATATCAAA  
ATTGATTT**CATTGATA**ATACGGGGTT**ATTCATGAA**AT**CAAGAA**ATCAAA**CAAGAT**TGAAAAAGCCATA  
TTTATCAGCTGA**AGTACT**TATATATATTTTTTAAATAAA**AATGG**GGTTAAAAACATAATCGGTGAAATAGA  
TTATCCAA**ACTGAAG**CAAAGACAAAAAGTGATATTGGATGCTG**AAGATGAA****ATGGA**ATTTAAGACTATT  
TTTTCTGATATCC**ATGAA**ATCCTTGACAGGAAAA**AACTCC**GGCAATAATAAA**CAACCG**TATTGTAAAA  
AATGCGCTTATTT**ATGAAT**TTTTGTTTTGTCTAA

26 518 **5**

*Desulfobacula toluolica* Tol2 complete genome GenBank: FO203503.1: 1120549-1120830 CRISPR-associated protein, Cas2 family

ATGTATATTATTGCAATGTATGATATAAA**ACTG**AAACAAAGCGGGTCGGAA**AAGAT**TAAGACATATGT  
TTAAATTATTGAAAAATACCTTATCCGT**ATTC**AGAATTCTGTCTTTGAGGGGGAGCTTACCAAAGCTAA  
ATTTGAGAAA**ATGAAGTATGAAGT**CAACGATATTATTGATAGCTCAA**AAGAT**TCGGTTATTTTT**TTCAA**  
AGCCGTGATATCA**AATGG**ATGGACA**AAGAT**ATCTGCGGGTTTGAAA**AAGATGA**TACGGATAATTTTTTATGA

12 278 4.3

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: 137724-137753  
CRISPR-associated endoribonuclease Cas2

TCAAAGAATTCCACCGATAGGTGCCTTT**TTCA**ACCC**CAAGA**ACCTTCCTTTTAAGTGCTTCTCGCTCCTC  
ATAGTATAAAGTATCACAGAATCCTCATCTTTGTTTATAAATACCTTTTCAGACCATCCTTTATCTC**TTCAA**  
ACTCACTTCTTGTAACCTCTCCCTCAA**ACTG**AGTT**TGAAT**CCAGTGC**AGATGA**GTTCTGAGGAA**AGA**  
**TTTG**ACCTTGTTAACCTGTCAACGTTGATGTCATAGACAATTATAACGTACAT

8 260 3.1

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: c1379814-1378849  
CRISPR-associated protein Cas1

**ATG**AAAAGA**ACTACT**ACCTTATGTCCGATGGATTACTTAAAAGAAAGGAAAAACACTGTATTTTGCA  
ACA**AAGATGGT**AAAAAGCCCATACCCATAAA**CAAGATCTATTCAA**TATACGCCTACGGATCCCTGACCTT  
TTCATCCCAGTCTGTGCACCTCCTGGCAAAGGAAGGCATCCCAATACACTTCT**TTCAA**CTACTACGGATTC  
TACGACGGCAGTTTCTACCCCCGGGAAACCCTGCTTTCAGGGGATCTTCTGGTGAAACAGGCAGAACACT  
ACCTCAACCCTGGAAAAAGGCTCTGGATAGCTAAAAAGTTCATTGAGGGTGCCGCAGGTAACATGGAAAA  
GGTCTGGGCTATTACCACGTTGAA**AATGG**CATTGGAGAAACAGTGAAGGATCTAACCTTTTTCATCCAAC  
GTGACCGAGGCA**ATGA**ACGTTGAGGCACGTATGCGTTCCGA**GTA**CTACTCAAA**AATG**GATGAAATATTGC  
CTGATGGATTTAA**AATGG**AGGGTAGGAGCAGGAGGCCTCCAGAGAAC**ATGGT**GAACGCCCTCATAAGTTT  
TGAAACT**CAATGGTCTATTCAA**CCGTTATCTCGGAGATATACAACACTCAGCTCAACCCACAATCTCT  
TACCTCCACGAACCCTTTGAGCGGCGCTACTCCCTAGCACTGGATCTGAGCG**AGATGTTCAAG**CCGATAC  
TGTTGACAGGCTCATATTTTAT**ATGGT**AAA**CAAGA**AAATGCTGACCGAGGACGACTTCGAACAGAAGGT  
GGAC**ACTG**TATGCTCAACGACAGCGGGCGGAAGAAGTTCA**TCAAG**GAGTAC**GATGA**GAGACTTAAAAAG  
ACCATAAAGCACAGGGA**ACTCA**ACCGTAAGGTTTCTTACAGGCGTTTGATCCGGCTTGAAGCTTACAAGT  
TAA**TCAAGCA****CATTC**TGGGCACACAGGAGTACAAA**CCATT**TGTTATGTGGTGGTGA

33 962 3.4

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: c1381325-1380807  
CRISPR-associated exonuclease Cas4

ATGCTCCCAA**ACTCAG**AAAAACACCTCCAAGTCAGAG**GGTACTG**AAATAAACTACTACTTTCATCTGCAGGA  
CCAAGCTCTGGCTGTTTTCCACAACATCC**AGATGG**AGCAGGAAT**TGAAT**CAGT**TTCAA**TGGTAAACA  
GCTCCA**CAAGA**AGAGCTATGGCCGTGAGAAGAAGGACTTCA**CCATT**GACAACCTCATAAGCGTGACTTC  
ATAAGGAGGGGTGAAGCCTTGGAGCTCCATGAGGTTAAAAAAGCAGTAA**AATGG**AGGACGCCACCCT  
ACCAGCTCCTTTACTACAT**GTA**CTACCTGAAGAAGGAGAAGGGAATCCAGAAATATTGTAG**ATTCA**TAGA  
CTATCCAAAGCTGAGGAAAAAAGAGAAGGTTGAGCTCACAGAATCTGGAGAAAAATATTTAGAGGGTATC  
ATTGAAGGTATTGGTCAAATA**ATTCA**TGGAA**CCATTC**CAAAACCTGAAAAAAGAAAAATATGCCGTAAAT  
GTGCCTACTACGAGTTCTGTTGGGTTTGA

15 515 2.9

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: c1385116-1384403  
CRISPR-associated protein Cas5

ATGGAAAATAAGCAT**TGGT**ATTTGATATATGGGGGGATTATGCACATTTTCGTAAAAATTGAAACCACCA  
CTTCTCCTCTCACAT**ATTCAA**TTCCAACAGGG**ACTGCT**GTTTGCTGGATTAATCGCAGCAATGTTAGGTTG  
GCAAAGAGATTCTTATTATGATACTCTTTCTAGGGATAATTT**TGAAT**TTTCTATTAGGATATTAGAACCT  
ATAAAAAAAGTTCG**GATGAAT**ATTAATTTGATAAAAA**AGATGA**AGGATTTTACCTATGGGATATTAAGG  
GGACTC**CAAGAT**CCCCTACTCCCTATGAGTTTATAAAAAATCCATCTTATAGAATTTACAT**CCATT**TAAA  
AGAAGCGAACACATATAAAACGTTGAAATATTATCTTGAAATCATAAAACAATCTTTACTCCCTATTTA  
GGTATTAGTGA**ACTCATTG**CTA**ACTC**AGTTTTGTTGG**TGAAT**ATGATATAAAATCCTAAAGAGTTAGAAA  
GAGCTGAAGTACAT**ACTG**TATTAAAAAAG**GATGA**TTACAAAATTTATGT**TGGTA**ATTTAGAAAGAAGGAAT

GAGGTGGTCACGAGAGAACAT**TGCCATT**TATAT**ATGAATTC**AGATCGTTGCGTAACT**TGAATATTC**GAAGTT  
CTTCTTGAACAGACTGGAAAACCTCTTATGATTGAAA**ATGG**ATTATTCT**ATGAA**ATTGGTG**AAGATA**ATG  
TTGTTTTCTGTGA

29 710 4.1

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: c1386034-1385126  
CRISPR-associated protein Cas7

ATGAGTGATGTTGTGAAAAATAGATCAGAAATAATTTTTTTATACGATATTGA**GAATG**CAAATCCA**AAATG**  
**GAGATCCAAATGGATG**AAAACAAGCCTAGGAT**AGATGAA**GAAACTGAAAAGAATATTGT**FACTG**ATGTAAG  
GCTCAAAAGAACCATCCGCGATTTTTTTATACAAATTTTT**AGATGAA**GAAATTTTCATAGTAGAAGAAAT  
AATA**AGATG**GAACCCAAAAACCAGAGAAAATAGACTTGAAGAACTTAAAAAAGTTCTAAAG**AGATG**  
GAGTTAAATTACTTGAAAATTATGTCGATTTAAGGTTATTTGGTGCAACAATTGCAGTAAAAATA**TGCC**  
**ATTGACATGGATAGGTCCTGTTCAATTTAAATTTGGTA**AGTCACTCCATAAAGTAGAACCAAAAA**CCATT**  
AAAGGAAC TAGTGAATGCC**TTCAA**AGGA**GAATG**TAAC TAGAGGAAC TTTTATTGAAACTCAAATCCCTCC  
CAT**ATTCAT**TTAATTTCAATTTTATGGAATAATAA**ATGAA**AATGCAGCCAAAAAAACCGGTTTAAACAGAACT  
GGATGTTGATTTAAT**GATGAATGGG**CTTT**GGAATGG**AACAAAGAATATCATCACACGATCTAAAAATTGGC  
CAAATGCCTAGACTGTTATTAAGGGTGGTTTATAAAG**AGAT**AATTTCCATATTGGAGACTTGGATAAGG  
GAATAAACTCAC**ATTC**AATGAGAATATTGAAGAAAAAGCTTTAAGAAATATATCTGAGGTTAAATTAGA  
TATCTCGGGTTTAAATAGATAAAAATTGAGTTAAATAAGGATAAAAATTGAAAGGATTGACTTTAAAAATCAAT  
GATTTAGAAACAACCTCCTACATTA**AAAGAT**TCGTTTGAAGGAATCAAACCTGAAAGAATTAATGTTTAA

34 905 3.8

*Methanobacterium congolense* isolate Buetzberg genome assembly, chromosome: I GenBank: LT607756.1: c1388623-1387892  
CRISPR-associated endoribonuclease Cas6

ATGCGCTTAAAAATAAGTCTGGCATCCCAGAACAG**TGAAT**ACAAAATTCCTTACAATTACAATTATCAAC  
TTTCTTCGTTAATTTACAGGA**AGAT**AGCAGACCTTGATCTAGCCAGCGAGCTCCATATTTCTCTGATTT  
TAAATTTTTTACTTTTTCTCAGATTTACGTTCCAGAAAGGATAAC**CAAGA**ACGGCATCATC**TC**AAGG  
**GATGGTA**ATCTGCATTTTTACATTTCTCTCCGAACGACTACTT**GATTC**AGAGT**ATGGT**TGAAGCTTATC  
TCGAGGATCTTGAGTTAACTTCCGGGGCGATAGGCTGCAGGTT**CAGGAGGTTGAGCTTTTAAAGAAACC**  
TAA**ATTC**AAGAAAAGTGCAGAG**TTCAAG**ACCATATCTCCTGTGATGG**CAAGA**ATAAAAAGG**GATGA**TAAA  
ATATGGGATCTGAACCCTGGAGACCTTCAGTTCTATACAGCTCTCCAGAAGAACCTTCTGAACAAGTATA  
GGGCATTTACGGTAGCTAC**AAATGGT****GATG**AGTATTT**GAAAGAT**TGCTCCAGACATGGGATCTGTGAAACG  
CAAACGAATAACCA**TC**AAGAAGGACGGCGTGGAACTTACCACAGGGCTTATC**AGATG**CGCTTACGGCT  
GAGGGGGATCCGAGACTTTTAGAATTTATGTATGATTGTGGATTTGGGGAGAAAAATAGTATGGGTTTTG  
**GAATGGT**TAATTTAATTAATTATAGTTATTAG

27 728 3.7

*Metallosphaera hakonensis* JCM 8857 = DSM 7519 strain HO1-1 chromosome, complete genome GenBank: CP029287.1: c384573-  
384313 CRISPR-associated endonuclease Cas2

ATGTA**TGCCATT**TTAGTTTATGACAT**AGATGA**GAGCAGAGTTTCCAAGGTCTTAAAGACCTGCAGAAAAGT  
ATCTAACGTGG**ATTC**AGAGATCCGTCTTCGAAGGTGAGAT**ACTG**AGGGGAACCTAGAACTACTCTCAGA  
AGAATTGAAAAGAATTA**TC**AAGGATCAGGACAACGTCTTGATCTATCTAACCCCAACGAAAAAGAACTTC  
AGGGTCGTCT**CCATT**GAAAGGAGAGGGATAGAAGCGACAACCTTCTTATAA

9 187 4.8

*Metallosphaera hakonensis* JCM 8857 = DSM 7519 strain HO1-1 chromosome, complete genome GenBank: CP029287.1: 392756-  
393250 CRISPR-associated protein Cas4

ATGTCAATTAAGTTAAGAGAGGACGTAG**TGGTA**ACT**TGGTAC**CTTGATC**TGGTAC**CTT**AAGAT**TTGCCCA  
GGGAAGTT**TGGTAC**ATGGCGAGGCAGATCACACCCTATGAGGGTCATCCGAAACTGGAGAAGGGACGAGC  
**GATTCAT**CAAATATACTCAGATCGA**ATTC**AGATCTCTCTGGAGGG**AATGAA****GAT**TGACGGGTTTAGGAGA  
GAGGAAAGAACCCTGATTGAGATAAAAATCGTCT**TTCAAG**ACATCT**TGAAT**CTGCAAGGGCGCAGACAACT  
ACTACTTATATAGATTGAAGGAGGTTGGTT**TGAAT**TGCGAAGGTGAGATTTACGTTCC**CAAGA**AAGGAGA  
GAGGCATAGGGTTACGCTCAACGAAGACCAGGTTAGGAGGGATCTTGAGGAGGTTAAGAGTATCGTAAAT  
TCGGAATCCACCTC**CAAG**AGTCAGGATTAAGTATTGCAAAAAGTGCGCATACAAGGACTTCTGTGGG  
GGTAA

20 491 4.1



Metallosphaera hakonensis JCM 8857 = DSM 7519 strain HO1-1 chromosome, complete genome GenBank: CP029287.1: 397622-398572 type I-B CRISPR-associated endonuclease Cas1

**ATGAAA**AAGGCTCTTTATACT**TGCCAACGGTAC**CGCTGAAGAGAAAAGGGGAACACACTCTCTCTGATCACCA  
AGGAGGGAAACAAAGTACGTACCCATCACCACCATCAG**TTCAA**TTTACGTGTTTCGGTGAGGTTAGG**TTCAA**  
**CAAG**AGGCT**CATTCA**GTTCCCTGAGCCGTAATCACATACCCGTTCACTTTTTTCGGGAAAACGTTACCT**GGT**  
**AC**CTTCTACCC**CAAG**ACT**ATTCAA**CGCCTCAGGTT**TCATGGTACT**TAAACAGGTT**CAGACTTACCTGGACC**  
C**CAAGATG**AGGGCTGGGCTTGCTAGAAAGTTTCGTGGAGGGG**GCCATT**AAGAACATGGCCACC**GFACT**CAG  
GGAGAACAATTTTCAGGAGGACGCCGTGAAACTTAGGGAGTACCT**CAAGTACTGCC**CGAAGACGT**AGAT**  
**GGGATCATGGGAGTTGAGGGGAACCGGAGGAGCTTCTACTTTCGAGTCCCTGGAT**AGAT**CCTTGATCAGG**  
ACTTCAGGCT**TGGTA**AGGAGGAGCAGGAGACCTCCCGAGAACATGGGGAACTCGCTCATGAGTT**TAATGAA**  
CACG**GFACT**TTACGCTGCAGTTCT**GATGAG**GGTTTATCAAACCCACCTGGACCC**AGGGTTGGGTTTCTC**  
CACTCAACCAATTTCCGAAGGTTCTCCCTGAACTTGACGTTGGCCGAAGTCTTCAGGCCGGTTATCGTFCG  
ATAGACT**GATGAT**TTCCATGATAAATAGGAGGGAGATAACCAAGGACGACTTCGAGG**AGATGA**CGGCTGG  
ACTTAAGTTAACCAAGGGGGGCTAAGGAAGGTGTTAACGGCTTTAGATAAGAGGCTCTCCTCAACGGTA  
AAGGTTGGGAACAGGAAAATGAC**GFACT**CCAGGATTATTAGGGCTGAACTGTATAAAAT**TGAGAGACAT**  
TGTCCGGGGACGTGGAG**GFACT**CCCCGTTCTTAATCAGGTTAG

38 947 4

Metallosphaera hakonensis JCM 8857 = DSM 7519 strain HO1-1 chromosome, complete genome GenBank: CP029287.1: 401050-401937 type I-B CRISPR-associated protein Cas7/Csh2

ATGACT**AGAT**TAAACAACAACACTCCGAGATCCTCTTCTTATATGAGGCCAAAGCTCACCAACCCCAACGGGG  
ATCCCGACGACGAGAACAGGCCCA**GAATGG**ACCC**CAAGACCAAG**AGGAACCTGGTTAGCGATGTAAGGCT**T**  
**CAAG**AGGTATTTTCAGGGATTACATCATTTCCAAACCTCGGAGAGGACAGCGTATGGGTTACT**AAAGAT**TGAG  
GGTAAGAACGTCGAT**TGCC**ACCGAGAGACT**CAAG**AGGATAGGTAGTCCCAAAGACGCTTTGGCCAAGTGTA  
TAGACGCCAGGCTCTTCGG**TGCC**ACCATACTGAGAAGGGGAACAGAAGGGGAAGTCCACGTCCTTCAC  
GGGACCGGTCCAGTTCACCTGGGGATATTCCCTCCACAAGGTTGAC**ATGGT**CGACTCCAGGT**CCATT**ACT  
TCCCTATTCTCAGGAAGGGACACCG**GGTAC**GGTAACATAGGCAAGGACTACAGGGTATACTACTCCCTGA  
TAGCGTTCCATGGGGCCGTGAGCGC**CAAG**AGGGCCGAAATCACGGG**TGCC**AGCGAGGAGGACCTAAAGAC  
GCTGGATAACTCCCTATGGGACTCCCTCGT**TACTG**AGACGGTAACCAGGAGCAAGCTGGGACAGAGGCCA  
CTCCTCTACCTTAGAGTTGA**GFACT**CCTCCCAGAGA**AGATGAA**GGGGGACCTGAGGAGGTTTCGTGAAGG  
TTAAGGAGAAAAGTGAGAGCGTGAGGGACTTGACAGACTTGGAAGTGAACCTCGACGCTTTGG**TGAAT**CT  
CGTGAGCTCCAACGACGTGAT**CAAG**GTTGTAGCGAGGTGTTCCGACGACT**TTCAAG**GGATTCTGCGAACAA  
GTTAAGAGTGCCTGC**GATGAG**GGGAAGAAGA**ACTG**CATCGTAGGGTAG

27 884 3.1

Methanobacterium sp. BRmeth2 chromosome, complete genome GenBank: CP022705.1: 468549-469262 CRISPR-associated endoribonuclease Cas6

ATGAGATTGAAAATCAATCTAACATCACAAATGAGTAATTATTTAATTCCTTATAATTACAACCACATCC  
TTTCTGCTATTATTTATCGT**AGAT**ATCTGACCTTGATTTAGCAGCAAAACTTCATTT**TTCAA**AGGATTA  
TAAATTTTTTACTTTTTTACAGATTTACGTTCCACAATTTAAACGTACAAAACAGGGCATTATTTCCAGA  
**AATGG**AAAATTAGAATTCTATGTTTCTTCTCCTAAT**GATGA**CTTAATAAAAAAGCCTCGTAGAAGGGCACC  
TTGAGAACACCGAAGTAAAT**TTCAA**AGGTGATAAACT**TACTG**GTGGAGCAGATTGAACTTCTAAAAAAACC  
TATTTTTAAAGAAAAT**ATGAAAATGAAA**ACCATGTCACCAATTATGGCCAGAATAAAGCGAGAAGTTGGG  
GGAAAAC**TTAAGAT**ATGGGATCTGGATCCTGG**AGATGA**CGGGTTCT**ATGAA**AGTGTTCAGAAAAATTTGA  
TTAATAAATATGTTAAATTTTACGGGGATTTT**ATGGT**GAAAGGTGGGTTAGGATAAAGCCTGAT**ATGAA**  
ATCTGCTAAAAGGCGCAGGATTGATATTAAAGGAACCTTATCAC**AGATG**TTTCAT**GATGAAT**CTTGAGATT  
GAAGCTGATTTGAGGCTTTTTGGAATTTGTTTATGATTGTGGGCTTGGTGAGAAGAATAGTATGGGATTTG  
GG**ATGGTTGAAT**AA

20 710 2.8

Methanobacterium sp. BRmeth2 chromosome, complete genome GenBank: CP022705.1: 471663-472280 CRISPR-associated endoribonuclease Cas6

**ATGAA**AACGCTAAGAATATTGATTGAAGGTTGGGTAACCTTCTTTTCAGATACCCCTGCTTTTATCAGCGGAT  
TTCAGCCAACTTTACCAGTACCT**CCATT**AAGTACAGTTTATGGCCTTATATCTGCAGCTAAAGGAGAAT  
AGTTACTCCAGAAGACTTATCTGTTGGTTATATA**TTCAA**TTACCAGGGAAAAGCAGTAGATTTAGAAACA  
ATAT**ATGAAT**TATCTGTTTTAAAGGTAAATCAAACGTTATTTAAAGAGAATTTCTCGTAAATCCTGAAA  
TTTTATTTATATATG**GATGAT**TTGGATTATCGAGATTACTTTAAAAAACCGCACTATCCTCTCCTTTTAGG

GCG**TTCAAC**CAGATCTAGTAACTATCAGTGAAATTAAGAAGTAGAGCTTGAAAAAAAAATCAAACATAAAA  
CT**TGGT**GAAACCATACTC**CCATT**AGGCACAAACGGTGCTTT**TGGTACTATTCAA**TCATTACCAACACATT  
TTACTAACACAATTCACGTAAAGCAGTAGGCACTAAA**CCATT**TATATTA**ATGAA**CCAGTTCTTTGATTA  
**TCAAGATGAATG**TTATTTTGATAAAGAAAAGGAGTGGGGAGTTTGGTTCACAAATGA

24 614 3.9

*Methanobacterium* sp. BRmeth2 chromosome, complete genome GenBank: CP022705.1: 474545-475060 CRISPR-associated protein Cas4

ATGCT**TTCAA**ACTCTGAAAAACAACCTCCGGATAATAGGAACCCAAATTAATTACTATTTTCATCTG**CAAGA**  
CTAAACTCTGGCTTTTTTCTCATCACATCC**AGATG**GAACAGGAATCTGATCTAGTTTCTCTTGAAAAATT  
ACTG**CATTCA**CAAAGTTACA**AGATG**AAAAAGAGTTTA**CCATT**GACAATCTAATCAGTATTGATTTTATT  
AAAAAG**CAAGAT**TATTTAGAGCTTCATGAGGTCAAAAAGAGTAAT**AAGATGAA**AGAATCCCA**TGAAT**ACC  
AGCTTCTTTACTACATGTATTATTTAAAAAATGAGAAAGGTATTGAAAATATCAAAGGCGTAATTAATTA  
TCCTACAATAAGAAAAAAGCTACGATTATTT**AGATGAG**TCTAAAGAGAAGGAATTAATGAGTATAATT  
**GAAGATATGGT**GAAATTATTAATAATGACACTCCAAAACCTGAAAA**ATGAA**AATTTGCAGGAAATGTG  
CATATTTTGAGTTTTGTTGGGTTTAA

22 512 4.3

*Nitrospira moscoviensis* strain NSP M-1, complete genome GenBank: CP011801.1: 1417060-1417353 CRISPR-associated endonuclease Cas2

ATGGCAG**AAGAT**CGGGCC**TGGTACT**TGATCTGCTATGACATCCGCGACCCTAAGCGGTGGCGGCGCGTGT  
TTAAGCTGCTGAAAGGCTATGGGCATAGCCTTCA**GTA**CTCCATCTTCCGCTGTGCGTTGACTAAGCGCCA  
GATCGAACAGTTGCGTTGGGAATTGGCGACTCTGCTCGATCCCGTCGATAGCCTGATGGTTGCAAGTCTG  
TGTCAGGATGTGCTTCTCGGATGGAGG**TTCAAG**GAGGGAAGGGTGACTGGGATTCCGAGCATGAGAAG**T**  
**TCAAG**GTTCTTTGA

9 290 3.1

*Aeropyrum pernix* K1 DNA, complete genome GenBank: BA000002.3: c784515-784237 CRISPR-associated protein Cas2

TTGGTGTACGTTTTAATAGCCTACGATATAAGCAATGATTCTAAACGCC**TCAAG**GCCGCACAGAAACTTC  
TGC**AGAT**GGGATTTCGCTAG**GGTAC**AGAAGAGCGTGTACATCGCGAAGGGAGGACGGAGTCT**TGCCA**AGGA  
GGCGTACCGAGCTCTCCAGAGGCTAGCGGACAGTGGGAAAGA**CAAGATCATGGT**GATGGT**TAT**TCCGGGC  
GACAGTGTAGGGATGCCTACGGTCTTGGGGGTAGTCTTGAGGATGGGAAGAGAGTTGTGGTGGTGTAG

8 275 2.9

*Thermoanaerobacter kivui* strain DSM 2030, complete genome GenBank: CP009170.1: c1756606-1756046 CRISPR-associated protein Cas4

ATGATTCTTAATGTAATTTTTTTATTTTTTGTGGTGTATTTACTT**ATTCAAG**CTATTATTGAAAAAAGAC  
CACCTTATAAG**AAGATG**AGGAGAAGTATTGGGTTTCGCGGTGGCAAAATAATATACATAGACAAGCAA**CA**  
**AGA**AGTTAAAGAAAAAGGCGTTACATATGGGAAACTATTTAAATCGGATAAAT**ATGGT**TTATCTGGAAAA  
CCTGATTATATTTATCAGTTGGGAGAAGAAGTGTTCCAATAGAGTTAAAAAGCAGTGAAGCT**GATGA**TT  
CTCCTTATTATA**AAGATG**TTATGCAGCTGGTGGCGTATTTCTTGATAATAG**AAGATG**TGTATA**CAAGA**A  
AGTAAGAAGAGGGAGAATAGTTTATAGAAACACTATGTTTGAGGTATACAATAAAAAATATTTGAAAAA  
GAACTTTTAAATCTTGTAGATAAA**ATGAAAAGATGAA**ACAAGGAAATTAATTTCCAGAGATAAAACCTT  
CATTTGCCTTATGCAGATTTTGCCCTGCAGAG**GGTACTG**TCTGTGAAATATACAATGACGCTTCCAAATAA

22 557 3.9

*Thermoanaerobacter kivui* strain DSM 2030, complete genome GenBank: CP009170.1: c1666773-1666372 CRISPR-associated endonuclease Cas2

A**TGCCA**AATGTTAATATCAGAGCTCTAAAAAATGAGATATTTAAAGAAGTTAAG**ACTGCCAT**CAATCCTG  
ATAAGGACAA**AAGATG**GCAAGGTATTTCTGAGGTTGCAAAGAAAACAGGGCATCTTTTAAAGAATCCAA  
TGTAACGGTTACACAGCTGCGCAAGGTTTTACAGAGGTCAAAAGACTTTCCCAGAG**GATGA**GAACTAC  
AAATACAAACTAAAGCTTTTAAAGCAA**AAGATGG**CATATACATCAGG**AAGAT**TTCCAAAGCTAA**AAGAT**T  
**TTCAAGAT**ATAGTC**GATGA**GGCAT**TGCCA**ATTGCAGAACAAAATGAGAAAAACATTGGAGAGATTTAAAGA  
CTTTTTTGAGGCAGTTGTTGCATACCACAAATTTCTTTGGCGGAAGAGAATAA

19 398 4.8

Observed mean proximity: 3.6 Expected mean proximity:  $2.15 \pm 0.9^*$  (\* 90%-confidence interval calculated from the smallest sample size equal to 187).  
Data from <https://www.ncbi.nlm.nih.gov/nuccore/CP021507.1?report=genbank>

## AL-pentamer proximity of the capsid proteins of 15 virophages

Zamilon virophage complete genome NCBI Reference Sequence: NC\_022990.1: 3374-5203 capsid protein

ATGTCAAATAGTGCTGTACCTTTAAATGTTGTAGCA**ATTCAAGA**ACCCCGTCTTGAACCTTAATA**ATGAAA**  
GAACTTGGGTAGTGGTTAAAGGTGGTCAACAAGTCACCTATTATCCTTTCC**TTCAA**CTTCTTTCAGCTC  
TAATCAATTTAATTTTCATTTGTAATCCTCCTTCAGCACAACTGTATTGGATCGATTAG**CATTCA****TTCAA**  
**GTTCC**TTATGATATCACT**TTCAAT**GTTAATCCTGCTCATGCTGGTGTAAGTACTGACAATCTTTTACAACCTG  
GTCGTGATGCTTTCCGTGCTTTCCCTATTAGTTCTATTACTAATACTTTA**ACTGCA****CCATT****AATGGT**TT  
CCCT**GTGAAT**AT**TGAAT**TAGGACAAAT**ATTCA**CGCTCTTAGCAGATATCATACTCCTCTTAAACTTAAA  
**AATGG**ATGGATGTCTATGCAACCTTCTTT**TGAAGATA**ATTATCAATCATATCGTGATGCTGAT**GGCA**CTA  
ATAACAATCCTTTAGGTGATTTTACAAGCGCATCTGGTCTTTCTGAGCTTC**CAAGA**GGATCATAT**TCCAT**  
**GAATG**TTGTGTCAAATACTCCTACT**TACTGCT**TAGAAT**TACTG**GT**GTACT**TT**ATGAA**CAAGTAT**TCC**TTCT  
**CCATT**TTTGG**GATGAACATCAAG**CTGGTGGTCTTGCAAATCTAACCAGTTTAA**CC****TTCAA**TT**GGGTAC**  
**TAAATA**AATAATCTTGCTAGAAT**ATGGTCT****CATTCA**GATATTACTAATGATGTTTCTGGAAACA**GTACT**AT  
TGG**TTCAATGAATG**TCAGCTTCCAACAACCTCAATGTATCTTGGATTTGTTACTCCTCGATTAAATATT  
CCTATTCCTCCAGAATAACTTATCCTTAT**TTCAA**ACT**TTCAAGAT**ACACTACACAAT**TCCAAA**ATACTC  
TTGCTCCCAATGCAACC**GTACT**TACAAATCTAATATT**GTTCAA**CTTGATTCTATTCCAAGGAACTTTA  
TGT**ATTCATGAA**ACAATCTG**ATTCA**GTGATTTATCAAATCTCAATAATCAAATTACCCTCCTGATGTA  
TTCCTACAAATCAATTTCTCTCAATTTAACTTGGAAACAATCAACAGGGTGTCTTTCTGGAGCATCTGCAC  
AAAATTTATATGACTTCAGT**GTTCAAAATGGT**TACAA**CAAGA**CTTGGAC**TGAAT**TT**AATGGT**CTTACACA  
ACAATTATCAGGAGTCTCTGGATCACCTACCAAAGTAATTGGTCTTGAAGGAGGAATAGTTTGTCTCGAA  
TTAGGTA**AAGATG**TAGGATTACGTGAC**GATGAA**GCTGAAGGTGTGCT**TGGTA**AT**TTCAA**TTTACAAGTAC  
AAATGACTTGCACATACTAATACTAATCAATATCTCACAATCGTACCTGATATGTATATTATTCAGTATAT**GA**  
**TGGTACT**CTTGTATTCTAATACTAGTGTATGGCATCAATTTGGTGTTCCTCTAAAGAAGAAGTATTA  
AATGCT**CCATT**AATCATGATGTCATATCACGAATTACAAACTGTATATGGAGGAGATTCTTTAGCA  
CATTTAAGAATTTCTTGAAAAGCCGCTAATGTAGCCGAAAGGTAAATAATTTTCTCA**AGAT****TTCAA**  
AGTAGCTAGTTCTGTGCTAGGAGCTATTCCATCCTTATGCT**TCAAG**TTCTGGACAAATTTTAAAGGAAT  
ATTGGTT**ATGGT**GAAGGTGGAGTATCTGCTGGTGGTGTGTTGCTGGATCTGGAAGAAGAAAAGGAGTG  
GAGTATTAGTAGGAGGCAG**TGAATG**TGAAGGTGG**AAGAT**ATATGTCCAAAGC**TGAAT**TAAAGAAAGCTTT  
AC**GAATG**TAA

83 1826 4.6

Sputnik virophage isolate Rio Negro virus capsid protein gene, complete cds GenBank: KJ183141.1

ATGTCTA**ATTCA**GCTATTCCCTTAAATGTTGTAGCAG**TTCAAGA**ACCTAGACTTGAACCTTAATA**ATGAAA**  
GAACTTGGGTTGTTGTTAAAGGAGGTCAACAAGTCACCTATTATCCTTTCC**TTCAA**CTT**CATTCA**GTTC  
TAATCA**ATTCAA**TTTTATTTGTAATCCTCCTTCTGCACAAACA**GTA**CTTGATAGACTTGTTTTT**ATTCAA**  
**GTTCC**CTATGATATCACTTT**TACTGCCA**ATCCATCTCATGCAGGTATAACTGAAAATTTAC**TTCAA**CTTG  
GTCGTGAT**GATTCC**CGTG**CATTCC**CTATTAGTTCTATTACTAATACTCTCAAT**TGCCA**CTATA**AATGGT**TT  
CCCT**TGAAT**ATTGAATTTGCACAAAT**ATTCA**TGCCTTAGTGCATATCATACTCCTCAAAGTTAAA  
**AATGG**ATGGATGTCTATGCAACCTTCAATTTGAAGACAATTACCAATCTTATAG**AGATG**CTGATGGAGCCA  
ATAACAATCCTCTTGGAGTATTTACTTTCAGCAGCTGGTCTTT**TGAAT**TACCTAGAGGATCATACACA**AT**  
**GAATG**TTGTAACACTACTACT**TACTGCT**TAGAAT**TACTG**GTGTTCTTT**ATGAA**CAAGTTTTCTTCTTCT  
CCTTTCTTGTGGGATGGAGAACAAGCTGGTGGTTAGCTAATCTTACTAGTTTAA**CATTCAA**TTGGGTAT  
**TGAAT**AATAATTTAGCTAGAATCTGGTCA**CATTCA**GATATTACTAATGATGTTTCTGGAAATA**GTACT**AT  
TGGATCA**ATGAAT**ATTAGTTTCCAACAACCAAGTATGTATCTTGGATTTGTTACTCCTAGACTTAATATT  
CCTATTCCTCCTAGAATTACTTATCCTTAT**TTCAA**ACTTTCCAGATACACTACTCAAT**TCCAAA**ATACTC  
TTGCTCCTAATGCTTTCTTCTA**CATTCAA**ATCAAATGTTGTACAACCTT**ATTCAA**TTCTTAGAAAATCTTA  
TTTGTGTTGTAACAATCTGATAATGTAATTTATCAAATCTCAATAATCAGATTACTACTCCTGACGTA  
TTCC**TTCAA**ATTAATAACT**TGAAT**TTAACATGGAACAATCAACAGGGTATTCTTTCTGGAGCATCTTAC  
AAAATCTTTATGATTTTCAGT**TTCAAATGGT**TACAATAAGACTTGGT**TGAAT**TT**AATGGT**GTAACCCA  
ACA**ATTCAATGGT**GTTTTCAGGACAACCTACTAAAGTCAATTTGGACTAGAAGGTGGAATTTGTTGTAGAA  
TTAGGTA**AAGATG**TTGGTCTTAG**AGATGATGAA**GCTGAAGGTGTTATTGGAAAC**TTCAA**TTTACAAG**TT**  
**AAATG**ACTGTAACATACTAATACTAATCAATATGTAAGTGTACTCCTGATATGTATATTGTAGCAGTTTAT**GA**  
**TGGTACT**CTTGTATTCTAATACTAGTGTATGGCTTCTATTGGAGTAGCTTCTAAAGAAGAAGTAT**TG**  
**AATG**CTCGTATTACTCATGGAGTTTACATA**ATGAAT**TACAAAGAATTTATGGAGGTGATTTCTTTAGTT  
**CATTCAAGA**ATTTCTTAGGAAAAGTAGGTAATGTTGCTGGAAAAGTCAATAATTTCTCA**AAGAT**TTCTAA  
AATAGCAAGTTTCAGTTCTTGGTGAATTCCTCATCCTTATGCT**TCAAG**TTCTGGTCAAATCCTTAAAAAT  
GTAGGT**TATGGT**GAGAGTCATGTTGGTGGAGGTAAGAAAAAGGAGGAGTTCTTATTGGCGGAAGACAAC  
TTACAAAAGC**TGAAT**TGAGAAAAGAACTTAAATGTGA

89 1784 5

Megavirus vitis isolate vigne Megavirus vitis transposiron mobile element, complete sequence NCBI Reference Sequence: NC\_040848.1

**ATGGT**TAGAAGAAAAAATGTATAAAATTTAGAGGTAAAAATGTATATTATGCAAAATATA**CAAGAT**TTGT  
CTAAAAAATTAATATTTCTTATTACGCAAGCTAGGAAAAATAATAAATGAT**AATGGTA**GAAGAATTGTTAT  
TG**ATTC**AAATAATAACACATTATTAATTAATAT**CAAGA**AAGAAAAATTT**TACTG**GACTTTTGAAAACAAAT  
TT**TGGTAT**TCAGTAGAATAGATAATAAAAAAATTTAATGATACTCA**CATTC**TTAATAAGGGTATTGATA  
TTGTTAAGGAATTAGGACCAAATACAGAAGGAAAAATTTAATAAAAAATTATTCTTTACATAACATTTGT  
GAGTCCTGATCCAGTTA**ATGAAGATGAA**ATATTAACATAATGATAAAAGTGGATTATGTTAAAAATACAAAAAT  
TTAATT**GATGAA**CATCGTCTCGAGGAACGAGAAATTACTACACCATAT**AATGGT**TACAATGACGAAATTC  
CTAATTTTATAAGAAGGAAAAATTAAGGAATAT**ATGAAA**ACATTAAAAAATCTGGTGGTCAATTAATACA  
TT**ATGAAT**ATCGTATTGGTTCATATTATAAAAAATAG**AAGAT**TAAAATTTAAAAAGGGATATGTTAGGGAA  
TTAAATAATGTTT**ATGAAT**TAACAG**AATGGG**TAAA**CATTC**AAATACGAAAATAGTGATAATACAAATGATA  
CATGCGTGGTTAGAACAATTTCCAAA**AAGAT**ATCCTAAATTT**GACTG**GAAAAATAAAAAAATTTGAAACTAT  
CCACGGCGTTAAAGTT**CAAGAT**TTTTTTAGGATTTTGTCAAAAAATATGATATCGGTTACAATATTTTT**AAAT**  
**GAACTGGT**AAAAAATTATATAATTATCAGGGTTCGGACGGATTACTTAACTGTATTATTTACAATAATC  
ATATTTATCCAATA**AATGG**CGGTAAACCAAAAAAATATTCTACAAAAGAATA**CAAGAT**AAATTTTGT**AGA**  
**TGAT**TCGTTAAAAAACTTAAAAATATATTAACACTGAAAAAAATTTACCTTCTAATATAAAAATGAT  
TCAGTTAGAAGAAAAAAT**AAGAT**AGATAATATTA**AAGATG**TTAATATTACATCATTTACAGTAGAAGATA  
AAAA**ATTC**ATTTGTAAT**GATGAATATCAAGAATG**TTTTAAAAATATTAACATA**AATGG**GATATAG**TGAATA**  
TATTTTTGATAATATACGAATAACAAATATTCTTAATTTATTAGAAAAAATATTAAGTAGACGATGTA  
TCATCATTTATTTCTGAAAAAGAATTATTTAAAAACATCACCGCTATTATGGAAATGTAATAATGTAATAA  
AGCCGAATCGTGTATAACAAAAAATAATAATGATGTAATA**GAATGGT**TAACAACCATCGACAAAAATAA  
ATGTT**ATTC**ATATTGTCTTTATTCTTTACCTTATCTTATAAAATTCGATTATAGAAAACATAAACTTAAT  
GATAAACCAACAAAAATTATTGATAATAATTTATATTTGGCAAAACCAAAATTTTGACTATTTAATGC  
CTAAACT**AAGAT**ATATCCTGGATATTTTTTATCA**GAATG**TAGAAATAT**TGGTATGAAT**TTGATTTAT  
AGAAGAATTAGAAACAGAAATAGTTCCTAATTTATTAGAAAAAATAATTAATTAATGTACGATAATATG  
TCAGC**TGAAT**CATTTAAAAGTGCATTAATATATTTATA**GGTAC**ATTTGAAAAGAAATGATATGCAAAACAT  
ATAATTACGAATATAGCGGTATATTTACAAAAGAAGCAGTAAAGGCA**CAAGA**AGGTTTTTATAAAAA**AAT**  
**GG**CGCATATTATTTATTT**ATTC**AAAGAAACCGAACAATATTTTCATGTTTCGTGATAGATTACCAAT**TGCT**  
TCACAAATAAAGGATATGTCTA**GAATG**ATTTTGTATAAAAAAATAAAAGAATTAGGATTA**AAGATG**ATC  
AGATTGTACAAATAAA**TACTGATTCAA**TATCATATTT**ATGGTGA**ACT**GCC**AAAGAATTAGATCCAAGTAT  
TTTTATCTGGTTGGAAACGATCGGATTTTTAAAGAATTAGGTGATGTTAACAATTTTATC**GATGAA**GAAGTC  
AGTGTAAAAATATTACTAACAGTAATAAAAAAA**CAAGA**ATT**ATTC**ATCAACAATATGCAGGATCGGGTA  
AAACGACATATATTATTGATAATCTTGTACCCAGATTATTAAAAAAGGGTATATCATTTATTGTATTAAC  
GCCTACACATAAGACATT**AGATGAAT**ATAGAAAAAATAATATAAACTGTGAGATCATGCAGAAATATGTT  
TTCACAAATA**CATTC**CAGAAGAAGACTATATTATTGT**AGATGAA**ATTGGTTTTGTTGATCGCGGATGTC  
ATGATTTACTTTATAAAAAATAATAAAGCCAATAAAAAATTT**TGAATG**TTTTGGGGATTTAATCAACTACA  
ACCAGTGGGAGAAAATCAACCATAAATCAGCCACATTATTTGAAATATATG**TTCAA**ATATATATACACT  
GATTTTATTAATTATCGCAATAACTTCACGAAAAGAATACTAT**GATGAT**TTAATAAACGAAAAATAT**AGATG**  
TGATATGTGAAGTAAATA**AATGGT**CGCATAAAAAATAT**GATGAA**AGCGGAATATATTCTATGTTTTCGACA  
TAAAACAAAAAATAAATAATAACATCATGATAAAAAAATTAGGTTTT**ATTC**ATGGCGTGATATTGGT  
GTGAAAATTTGTTGTGATAAAATAAATTAATTAATGAACCATATTTATAATAATAAA**ATTC**ACCATAT  
CAGATTTAATAAAAAATGATAAAAAATTAATAATTA**AAGATGATAATGGT**GAAGAATTTGTTGTGACCGA  
AAG**AAGAT**TATTATCTAATTTCGAACCTGCTTA**TGCC**ATTAAATATACA**TCAAG**CACAAGGTATGACACTA  
**AATTC**ATTTTTTGGGCTTCT**GATGAT**GATAATTTTATT**AATGG**GAATGTTGCTTATACAATTTATCAGCA  
GGTAAAAACAAAAAATTAATTTTTCATTGGAAAGAGAATTT**ATGAA**AATATTATTGACTGAAAAATAATTA  
TTTATAA

123 3083 4

Maverick-related virus strain Spezl (virophage at the origin of large DNA transposons), complete genome NCBI Reference Sequence: NC\_015230.1: 14876-16696 putative major capsid protein

**ATGAAT**ACACCCCGAAGACTCGA**TACTG**TTTTACAAGCTCCATATGCTTATAATTGGCCCACTTCTAAAA  
ATGTTAAAAATGGCTTCTAGAATAGG**CATTC**CATACTCAACAT**TTCAA**ACA**ATTC**AAACCGTATC**AGATG**C  
ACCTAAT**AATGGT**ATAGGTCAAATAACTTTTAATCAACCTCTTGAAATTTAACAGGAGGTGCTCCACGT  
TTAAGAGTAAGTTT**TACTG**CTGAAATTAATAATTTTAGCTG**ATTC**ATCTTTAA**AAGAT**CAAAATAGGCC  
TGAAATCTTTTCCAGTTAATAGAAAGTATACCTGTTGCAGTAATTAAT**ATGAATGGTA**AAACTTTTACCAG  
TTACCCCGCACAAATTAATTAACCTTCATCAATATAATGCGGATCCTCTAGAACTTGCTTTGCTTTCTCCC  
TGCTCTGATGT**AGATGAAT**ATAATAAAAAATAAAAGCCGTTTCT**ATGAAT**AACCCTTATAGACAA**GGTAC**AG  
AATCTACCGATAGCC**GAATG**TCTCGCGGTTTAGGTTGTAATTATGCCTATTATATCCACCACGGGCAGC  
AGGTAGCACATCCGTAATAAATTTGATTTTGTAGT**AGATGAA**GCCTTGTAGCCAATCCACACAATATAAA

AATATTAAGACCCAGTACCTTTTAGAAACCTGAACACATTTAAAGTCATATTAGATGGACAATTTAAAC  
CTGAAAATATGATGGTATTGCGGATGTTAACTTGTGGCGGGTAAAGCAGATTTTGAAGTGGATAT  
AACCGGATTTAAATTAATATGCTTGTTCAAAACCTGGGTGCGCCCTCTTGAAATGGTGAATTCTAAA  
ACTATAATATATAACACCCCGTTAATTTTCATTAGAAAGGTAATATAAGTTTCGATGTGCCCAATACAAAAG  
ATCCTTACGGTATCCCTGGTGAACGAAATAAACACATATTAACAACCATAGTATGGCAATGAATAATGT  
CCCTTCAATGTTTGTCTGTTATGGTTTCACAAGAACACCAACAAAAAGTTTGCCTGATCAACTTGTCT  
GGAATTTATTGGATTAGAAATTAAGGTGGATTCTGATGTTGGTATTTTGTAGAGAACTGAAACAACAAC  
TGTACGAATTTATCTTCATCTAATGGATATAATAAAGATTTTCATGCTTTAGTGGTGTCTGGCTAATGG  
CTTGACTGTGGCAGATCCCGCTGTGGCTGCTGGTAACAAATTTAAAGAAGCTATATTTGGTGTGGTCT  
GTAATTTTCTTCCGTCCTTCCGATCTGGGATTAATAAGATTTATAATGTTATGGCTAATGCTAATAAGTCCA  
TTAATATGCAGGTACAAGCAACATTTGTACCCCTGAAGCTGCGGGTACTGTGCTCATTATAAATTAGA  
AGTGTTTTCAATTAGAGATAATCTCACCTACAGCTTTGAGATGGTACTTTTATGATGATTTAACCTT  
TACACTCCCGACCAACTATTACGAAGCCCTTAAACTTACTGATGACAATAACAAATTAATGCGGGTTA  
TGGGCGGGTCTTTTATGGGTGATGTTATGACTAATTTTAAACCATATGGCAGCTCATCTGTGACTAAAC  
GGTACTAAACTTTTAAAGAAATGCGGGAATTAAAGATTTATGCGGGAGATGGTACAATGATGGGAAAC  
ATTGCTCCGTTTATGGTTATGGTAAAAAAAAAACAACCTAGAAAAAAGAAAGGTGGAGAAATTTGTTT  
TACTAGGATCAGAAAAAAGGCGGTAAAAAATTATCAGATAAACAACTTACGACTTAAGAAACCTCTAA

71 1817 3.9

Moumouvirus australiensis isolate 10A, complete genome GenBank: MG807320.1: 471905-475239 capsid protein 1

ATGGCAGGAGGATTATTACAACCTCGTAGCTTATGGTGCTCAAGATGTTTACCTAACTAGTTGGGTAGAA  
AAGTAGGACGCCGTGATGAATATGGATAAGTCACGGATAAACCGGTTTGTGGTCCATTAAATCGTAATAA  
ACTTACGAAACCACGCCTGCTAGTGGGAATAATTTTATTCTGCGACACCTTCAAATTCAGATGAAACCC  
TAAAGCCGTATAAAGTAGATAAAAAAATGAATTTCTCAAAAATTTGTTTTATTACAAATTCATGTTATTAT  
TAATATGGAACTAAAATATGCACACAATGTAAAGATGAAAAATCTTTAACAGAATTTGGTGTAAAGAAAT  
AAAAAGAAAAATATTTTACATCAATGGTGTAGAGTATGTCTTAACGAATAAGCAAAAAATACCGCGAAG  
AGAATAAAAAATTTGCTTAAAAAAGACAAAAAAGTGGTATAAACTAAAGGTAGAAAATGGAAAAAAT  
ATATGATAAAAAATAATCTCGAAAGAAACAAGACTAAGAGATAATGAAAGATATCATTCTGATCCAAGTTTT  
AGAACCAAAAAAATACTAAGATCAAGATTAGCAAAAAGTACTTAAAGGTGAGAAAAAATCTAAAAAACTT  
TAGAATATGTAGGATCTGATATATTTTATCAGACAATGGTTGAATTTTCAAGTTTCAAGTTTCAAGTT  
TTGGGAAAAATCAAGTACTTATTGGGACGTAGAACATGTTATACCTTGTCTGCTTTTGTCTTCTCTT  
GAGGAAAAATTTTATAAGTGTTTTAAATTTGGAGAAATTTAAGACCTTATGAAAGAAAGCAACCGTAAA  
AAGGCAATAAAAAATAATTATTGAAGACATATTTAATCAACAAGAAAAAATACAGAGAATTTGAATTTATTTA  
TCTATTCGGTACCAAGTAACATGGGAACTATGTTATGGCCTGGAGAAAGTAGCCAGGGTATGGTCAACAAG  
CCGACGGATGATGAATTTGGTTAAAAAATTTAATCAAAATTTTGAATGGGCAAAATGCTGAGCCAAGTCT  
AAACCTCGTTATGCAAGAGGCAAGGATGCAGTCCAGAGACTAGAAGGAGGTGCGCGTGAGAGAATTAGCA  
GTTCTCAGTATCGTGTGAGGTATAGTCCGTCCCATCTCGAAAAGATGGGATCCGACAGACAGGGAATC  
CGCAATAACTTTTTTAAAGTTTTGAATACAACATTTCTGGCCTTAGTAGTTATTAAGGTAATGCTAG  
TCATTTTATCAATATTGGTGTAGTAAGCAACCTGCGCTGAAAAGATCCGATATTGAAAATGGCGACAC  
TGTCAAATTTGCGGGAACGTCCTAAAGACGTCGTGTACCAAGTCAATGGAGAAATCTATTGATGGCCGAGA  
TTGGAACCTCGGATGGTAAAAATCACGCGTATGATTAATAGAGTCATTCTATTGAAATGGATAAT  
CCGCAGCCAAGTCTAAATACCGTAAAAATGATGAGTATTTCAAAAAGTAAAAATTTTGGAAATCACGGTA  
CATGGATGCAGTTCCAGACATAAATGTCAAGTGGCAGTGAAGAACTAATCATTCTTTATGATTCGTTA  
AGTTATAGTCCGGCCACTAGGAACTATGTGGACTAACCGGTATATCGTCTGCTACTACTAACTTTGCGA  
CTGAATCAATTGAGCAATATTTTAAATGGTAAATACTAATTTTCGGAAGAAAAATCCACCGCAGAAATATCTCG  
TAATGGAGATCTTATTACCAAGCTTTTCTTAAAGTAGTATTACCCGAAAGTAAAGATTTACTGTCTTTTT  
GATAAATTCGGACATGTTAAATTCGCTTGGGTCAACATATTGGTCATGCTATTGTTGATGAAACTGAAC  
TCGAAATCGGTGTTTCTACTATTGATAAACAATATGGTATTGGCTTACCTCTGGCAAGAGTTTCTCA  
AAGTAAAGATCATGAAGCCGGTGGCAATAAAGATGCTGGTAAATGTTCCGAATTAACATCTATCACTACT  
CTTGATTGGGACAGTTATGATAACAATCTTCTTAAAAATTCATACACTCTTTACGTTCCCTTCAAATCT  
ATTTTTGTGCGCAATTAATGGTCTCGCTCTCCCTCTTATTGCTCTCCAATATCACGTAGTCAGAATCTACGT  
TAAATTCAGACCCGCTGAACAATGTTACATGGCAAGTATGCTTTCAAGAGTGGTGTGATAACTTTGAA  
CTTGACGATGTTCAATTTATGTCAACTATGTTTATCTTGACACTGAAGAAAGACGAAGATTCGCTCAAG  
TTTCTCATGAATATCTTATTGAACAACCTCAATTCCTGGTGAAGAATCAATGGTACTAGCAACTCTGC  
TAAATATAAACTCAATTTACAATCACCCCGTCAAAAGCTCTTTATTGGATCACCAAATTAGGTAATTAACA  
GGAGGATAAACTCAATTTATGATCACGAAGTGGAACTGCTCGTGAAGAAATGCGCTAAACTCTTAA  
TCCTTGCTCAATATGATCTTTGACGAATTTGGTTATTTCAATGATGTTGCTCCTGAAACAAAACGATAACAG  
TTATCATGGTATTGTGGTCTTCAATATGTTGGTGTAGATCTTCAAGTACTTCTGAAAGAACCCATTAT  
ACCTTCAATGATTCACACCCGCTGATGCCTTTGATGGTACTGTTCTTATGGTAAATTAGCTCCTTTGTG  
TTCCTCTCCTTAAAGAGAAACAAGATGTTGATCTCCGTGATAAAGTTGATGGTGTATCCGCATTTGTAC  
TGATTTGAAAATGATAACTTAAAGATATCTGAAAGTTGAAAAAGTTACCAGAAACGATCTTCTATTTCT

GATCT**TTCAA**TTCTGTTGATAAATTC**GATGAA**GACAACCGTGTGATTACATTAAGAAACACGATATTA  
CCGTATGGCAACACGACAACACTACGGTCTTCTCATTGACG**GTA**CTAAAAATCCTGTAAGTGATGCTGAACCT  
CCAAC**TAATGGT**CAATCACGTCAAAG**CAAGA**GATCTGGCTTCTGGCATGA**TACTG**TTGAACCTTACATG  
CATCACACTAAAACACCCA**AAGATGGT**CTCAATGTCTTCTCATTCGCTCTTAAACCTGAAGAACATCAAC  
CATCAT**GTA**CTTGCAATTTCT**CAAGA**ATTGA**TACTG**CTCAACTCAATCTCTGG**TTCAA**TCACFTTGCTAA  
CAGCAAATATGCTGATGTTTTTCGCTGACAATGATAATAAAGTTCTTATCTTCGCTATCAATTATAATGTC  
CTCAGAATTATGAGCG**GAATGG**GAGGCTTAGCCTACTCAAACCTGA

150 3331 4.5

Pleurochrysis sp. Polinton-like virus isolate PleuroPLV, partial genome GenBank: KY203338.1: 11562-12602 minor capsid protein

**ATGAA**CGCGCCTGGATACCATATTGTTGACCAGACCAATGTGTATATCAATACCGGCTCTGGACTGAACC  
CAG**ATTCAA**CGGGAGATGAT**TA**CAAGCTGACGTTTAAACGATGTGAACGTGAAATGCGGAGACAACCAAGTT  
CT**TTCAAG**CTCACGTTAAAACAGTTTAAACATGTACAAGTCATGGACGAACG**TGAATG**CGAATAACT**CCATT**  
TTCGTTCTAAACATTGGCGGAT**CATT**CTTACCAGTTCGCTTACACAGCGGAACCTATTCGAGTTTTCGCTG  
ATCTGGTGACT**TGAAT**TCTCCTCAATTGTTGATT**TGGTACTG**ATTAGAGAGGGAGCCTACACTACATCGTC  
AGT**TACTGTGAAT**AACCTAACACAGCAAGCGATATATCCGGTAACGTGGAATTCGATTTTTCGATATCG  
ACCACAGAGGCTCACGG**ATTCA**GTGACGGAGAAA**TAAG**CTGCATGC**GAATG**TGTCATTGGGTGACTTCT  
ACAG**ACTG**CTTGGGG**TAAG**CGTAGCAAGGTGGACAACGATGTGGACGGTTGGC**AGATGA**CCGCGCCTGA  
CGCGAACACTCTGCGCTTCAGAGGATATTACAACCTGTCAGCTGA**GTA**CTGAGAGCTACGTCTACCTACAT  
TGCTCGACATCGAGCTGCAGTATCGCTACGAGGAACCTACACTCTCGGCTCCAAGGACAGCCCTGGACCCC  
AGCTCGATCATAGCACGATCCTGGCT**AAGAT**CCCC**ATTCAA**AATAACGTCTTGGCATATGACTCCAGTC  
AGACGATTCGTAACGCTCTCTGAAAAACAT**CAAG**CGAGATCTCGTATATGGAACCTGACCTTAGGGACAGT  
AAGGGCCGT**AAGAT**CCCACTCATCG**CTGCC**GGTCAGAACAC**AGATG**GGAACCGCT**CATT**C**GAATG**CCTAC  
TCGAAATCCAAACG**ATTCA**GCGGACCGGGTCTCACCAAGGTCACFTTAAAGTCTCCGG**AAGATGA****CAAGAC**  
ATATGCGCATCAACACTCGACCCCTCGTCTTCCA**CAAGA**ACGGACAACCTGG**TA**CTGA

44 1037 4.2

Cafeteria virus-dependent mavirus, complete genome GenBank: KU052222.1: 24414-25325 minor capsid protein

**ATGAA**ACAGTATATTTGGCTT**AA**T**GAA**ACTAT**CAAGA**GTAATAAACAATTAGCTGGACCGGGGCTCTT  
ATAAAAGACCCGTTAGTGTGATATTTTTTCGTTCTTCTACAATTTTAGACCCCTGATAAAAAATTATCTCTT  
AATTGTTGAAGAGTTCCATCTCCATAAAATAAGACTT**CCATT**ATTTAAACCCGCGGGACATGACTA**TCAA**  
**G**TAGGAATTTTTAAACCGTTCCACC**GATGAA**ATTATGGGAGTACGTGAGGTTGATTTTAGTACCTTTGT**AG**  
**ATGAAGATGGGTAC**ATGTACGATTATGTTGATGTAGGTACC**GCCATT****ATGAA**ACTTTAGCTGGATTATG  
TGACGGTATCATAGGAGAAGAAGACATACCCGTTTTTAGTTTTAATAAACATAGTAAAAAATTTGAAATC  
ACCACACTGAAAATTTTAGAAACGGTCATTTTATTATGTTTAAAT**GATGA**CATGAGAGTAGACTTTAACT  
CTTT**TGAAT**TTGACGACATAGACGAAGAT**ATT**CATTAGTTATATTAAT**GAAGATG**ATAGAAACA**CAAGA**  
**TG**CTTCCACTTTAGAAATTTTTAACCCCTATATCGCATATTGTTATAGAATCTAACGACCTCCCGTATCT  
**TATGAAT**TACTCCCT**TTCAA**TATCTAAAAACACCACTATATCAGATAACACGGGGTTTTTCTTACTAATTT  
ATAAATATTTACAACAAAACA**TAAGAT**TATAACAGCATTTTATTTAGGGTTGAAAATTCCTCTAATAA  
ATACCATAATATTTTACAACTAATTTTAAACCGTTTTAATTTATCTTTACCATATACGATTATGATA**AT**  
**GAAA**AACACCCCTTTAACTCTTCTTCTCAAACAGTAATAACA**CTTAA**GCTTTTATT**TGAAT**CAATAGATTAA

33 908 3.6

Cafeteria virus-dependent virophage mavirus, complete genome GenBank: KU052222.1: 25440-27260 putative major capsid protein

**ATGAAT**ACACCCCGAAGACTCG**TA**CTGTTTTTACAAGCTCCATATGCTTATAATTGGCCCACTTCTAAAA  
ATGTTAAAATTGCTTCTAGAATAGG**CATT**CCATACTCAACAT**TTCAA**ACA**ATTCAA**CCCGTATC**AGATG**C  
ACCTAAT**AA**TGGTATAGGTCAAATAACTTTTTAATCAACCTCTTGGAAATTTAAACAGGAGGTGCTCCACGT  
TTAAGAGTAAGTTTT**TACTG**CTGAAATTTAAAAATATTTTAGCT**GATTCA**TCTTTAA**AAGAT**CAAAATAGGCC  
TGAATCTTTTTCCAGTTAATAGAAGTATACCTGTTGCAGTAATTAAT**ATGAATGGTA**AAAACTTTTACCAG  
TTACCCCGACATAATTAATAA**ACTT**CATCAATATAATGCGGATCCTCTAGA**ACTT**GCTTTGCTTCTTCCC  
TGCTCTGATGT**AGATGAAT**ATAATAAAAAATAAAGCCGTTTCT**ATGAAT**AACCCCTTATAGACA**AGTAC**AG  
AATCTACCGATAGCC**GAATG**TCTCGCGGTTTAGGTTGTAATTATGCCTATTATATCCACCCACGGGCAGC  
AGGTAGCACATCCGTA**AAAA**TTGATTTTTGTAGT**AGATGAA**GCACTTGTAGCCAATCCACACAATATAAAA  
AATATTAAGACCCAGTACCTTTTTAGAAACCTGAACACATTTAAAGTCATATT**AGATG**GACAATTTAAAC  
CTGAAAATATGAT**TGGT**ATTGCG**GATGAT**GTAAACTTGTGGCGGGTAAAGCAGATTTTGAAGTGGATAT  
AACCGGATTTAAATTAATATGCTT**TTCAA**AACTGGGTGCCCCCTCTTGAATTTGGTGA**CATT**CTAAA

ACTATAATATATAACACCCCGTTAATTTTCATTAGAAAGGTAATATAAGTTCGATGTGCCTCAATACAAAGG  
ATCCTTACGGTATCCCTGGTGAACGAAATAAACACATATTAACAACCTCATAGTATGGC**AATGAAT**AATGT  
CCCT**TTCAAT**TGTTTGCTGTT**ATGGT**TTCA**CAAGA**AACACCAACAAAAAGTTTTGCCCTGATCAACTTGCT  
GGAATTATTGGATTAGAAATTAAGGTGGATTCTGATGT**TGGTA**TTTTTAGAGAACTTGAACAACAACAAC  
TGTACGAATTATCTTCATCT**AATGG**ATATAATA**AAAGAT**TTTCATGCTTTAGTGGTGCTCTGGCT**AATGG**  
CTTGACTGTGGCAGATCCCGCTGTGGCTGCT**TGGTA**ACAAATTTAAAGAAGCTATATTTGGTGCTGGTCT  
GTAATTTCTCCGCTCCTCCGATCTGGGATTA**AAAGAT**TATAATGTTATGGCTAATGCTAATAAGT**CCA**  
**TTA**ATATGC**AGGTAC**AGCAACATTTGTCAACCCTGAAGCTGCG**GGTACTG**GTGCTCATTATAAAATAGA  
AGTGT**TTCA**ATTAGAGATAATCTCACCTACAGCTTT**GAAGATGGTAC**CTTTATG**GATGA**TTTAACCCCT  
TACACTCCCGACCAACTATTACGAAGCCCCCTAAA**ACTTACTGATGA**CAATAACAATTAATGCGGGTTA  
TGGGCGGGTCTTTATGGGTGATGTTATGACTAATTTTAAACCATATGGCAGCTCATCTGTGACTAAAAC  
GGTACTAAACTTTTAAAGAAATGCGGG**CAATTAAAGAT**TATGCGGG**AGATGGTAC**AATGATGGGAAAC  
ATTGCCTCCGTT**ATGGTATGGTA**AAAAAAAAACAACACTAGAAAAAGAAAGGTGGAGAAATGTTT  
TACTAGGATCAGAAAAAAGCGGTAAAAAATTATCAGATAAAACAACCTTACGACTTAAGAAACCTCTAA

73 1817 4

Qinghai Lake virophage, complete genome GenBank: KJ854379.1: c15342-13609 major capsid protein (MCP)

ATGGCATCCGCAGACTTTAA**ACTACTCTCA****TC****CAAG**CAATCGACCC**TTGCTGATATTACCGACCAATTGG**  
ACTACGCAGTTCTAAGTGGTGCTTCCAGCA**ACTTACCAGCAATTTAACGCTGTATCAACATCCGCATC**  
**AAG**TATTACCTTTCTTTAC**AGGTAC**CTTCTGAGTCTATAGTCGTA**TC****CAAGA**GAAATATTGATAAGAAC  
GACATTAATTT**TACTG**TTGATGTAACAAATGTTCTATCGGTGCAAAGGCATTTGACTACGGAAAAACCG  
ATGCT**TTCAAGCATT**CCCTCTTAATAGTTTATTTAGACTT**CATCCGCACAAATCAACA**ACTAATGT  
CTCTGTTAATACC**CAAGATG**TATTGCCTCAATTGTTAAGACTTAACAATAGCAGAGA**ACTTTATA**AAATAT  
AACTC**ATTTTGTCTTCTCTACCCGATCAAG**CATA**CCATTCA**TATGCC**ATGGT**GTTGATGCTTTTAATA  
ACC**CA**TTAGGGGCTTACTCTAATGCATCTTACGATGTGCACCTTATTCTAGGGGAGCATT**TAA**ACTTAA  
AACATTTTCATATGTATATACTCCTGTGCTGGTGGCTCTCCTGACACTCTTGTCT**CTGCCA**ATGTTGGT  
GATGTTTGG**AGAT**ATCTATCACAGTTGAAGTATGTGAACCTATTATTGGTCTCTCA**CCATT**TGTATTG  
GAG**ACTGC**GATTACAATAAGCAAGGTTTAGT**TGGTA**TTAACGCCTTGCTCCTTTGCT**TTCAAT**ATTGATAG  
TAGTTGTAAGAGATTTTTCAGTTGCTCCTTCCACCA**TC****CAAG**AGACATATCTGTTAGTTTAGGA**ACTTCT**  
**GCCA**ATCCTCA**CCATT**CGAGAAATGTTCCCTGTTGGTTAATTTCTT**GTACT**CAACCATCCGACTTGA  
TTCCTGCTCGTAATGTAGTACCATAACCAGATTTT**CAAGAT**TTTTATCC**TTCAAT**CTGCTACAAATCC  
ACTTGGTGCTGGTGCTTACCACAGTTAATCTCAAAA**CATTCAA**ATTAATCAATTGCCGATTACTTT  
ATCGTCTGTGTAGAAAACCAATGTCTCAACAAA**CA**TTAAGGATAGTTCTTCTCGCTGTTATTAATA  
GTATAAGTGTAACTTAA**CAATGCATCGGGTCTTCTT**CATCTGCTACC**GCCCAAGA**CCTTTGAGAAAT  
TTCTGAGCGT**AA****TGGT**TCTACTCAAA**ACTGGCTAGA**ATTTAGCGGTAAGGCATCAAACGAAGGAAAC  
GATATCATCGCCACTGATATTGCTACA**ACTGGTTCGCTACTTATTTT**GTCTCCT**CCATT**TGACTTGTCTC  
TCCCTGACTACTTAA**ATCAGGAAGCATCGGTCA**ATA**CAACTTCCA**ACT**TTCAA**ATGAATATC**ACCA**ACCA  
ATCTGTGATGCTTACTCCTGAAATTTGTATAATCTGTGTTAATAGTGGTGTATTTACTACGGTTGCT  
GGTTC**CA****CAAG**CATCTT**TACTG**ATCCTTACTAA**CAATGGT**TCTTGATAG**CAAGA**CAACTGAGGCAA  
TTGACCCTGTATCATCCGTTCTATATAAG**GAATG**ATTGGTGGAAATCTTGGAAACAGAA**TGCTACTGC**  
**CG**TTAAGAAA**TGCATT**CTAAGAGATTTAGCAGGAAATTTAGAGAG**AA****TGGTA**GGTATGGGTGTCCA  
GTTGGAGGAGGTATAATGTCCGCTGGAGGAAAATTTGGATGCTCTCTGTGCTTAA

64 1730 3.7

Miers Valley soil virophage partial MCP gene for coat protein, segment MCP, isolate Miers Valley soil GenBank: LT221025.1

ATAAAATACCCACTGTATAGTCCAAATAGGATTACCTCAACATCAAATCTTGCACCTGTTGCTGCTCTAG  
CAAGTATTTCTCTGACTTTACA**CCATT**GACTCTTACTCAAATTCCTTATTGTGTTTATTATGGGTGG  
TCGATC**AGATG**CCGAAACATTAGCAAATGTTAATAATATAGTTACATCAACCGATACTA**ACTTTAGA**AT  
**GATTCA**TTGAATATGACTTGGGG**TACTGT**ACAAGTTATGAGTTTCGCTGGT**CCATT**AGATCATTATAATG  
TTTCTATAAA**GAATGGT**TTAAAGAAGCTCTCATTACAGAATATTATGGAGAGACCAACTCGTTT**CACA**  
ATCTATCGCCGCTCCTTCTACTAAAATTTGGATTA**ACTGGA**AGTGTAAT**AA****AGAT**TAATACC**TGGTA**TTGAT  
TTTGCAATTAGCGGATACCTCTCTAGCACCTGGAG**TGAATG**CACAACTACTTTCCAATTAGTGTAGTTG  
CAACAAATCTTAATAATAAATATGCTGTAACACCTCA**ACTGAATG**TTTTATTTGT**AGATGAA**GGATACGT  
TGATATCCAAGGTGATTTAGCAACTAAGCGTATCGGAGTTATTCCTCGTGAAGCTGTTATCCACA**TGCCA**  
GCAAATGATAGTTATAAGGCT**GAAGATTTAAGAT**ATTTTGGTGGAGCACCATCTTCTCTAGTCGCTTFA  
ATGTGTTTTCTAAAAATTTAGCAAATCGAGTT**AAAGAT**TTTGATATTCC**TGGTA**AA**CAAGA**CAAGCCAA  
TCGTTTTTT**CAAG**CAAAAAAGTAATCAGTAGAGGAT**AGATG**CAGCTTCGTGGTTACCAA**CAAGAT**TT  
AG**TGGTA**ATTTACA**AGAT**TATCAGGTGTAGCTCGTGATTTTGGTTATGGAGGATGTGG**AGATGA**TATGG



GTGCTTATTTACCTGAAAATATAC**ATGAAG**ATAGAGGAGGTTATTTGGCTGTTAATAA**TCAAGAT**AATAT  
GGG**TGGT**AGATCTATTCTCTTGATAGATTACGCAATGCTATGTGTAAACGATAA

42 1031 4.1

Organic Lake virophage, complete genome GenBank: HQ704801.1: 7791-9518 major capsid protein

ATGTCTAGCGAACTTTCTACCTTTCTAGTC**CAAGAT**CCACGTATCAGTCGTATCACCAGCGATG**TTCAAG**  
TTGCGGTGA**AAGATGGT**CCGGCTTCGAGCGTGGTCCAAGGCTACCCTACTAACTCGAACAGCTCGTCCAC  
TACCCTTT**TTCAAT**GTCAATGTCCCAAGCGAAAACACTCTTGATAGACCGCAACCTTCGTGTCCAA**GGTACG**  
ATCCAATGTGTTAAGTGAATTAACGTAGGCACAACACTGAGATTAATGTGATACTTTACAAATGTACCAT  
CTGCCTTTCCCTTAAAT**TCAAGC**ACT**TTCAAT**CGGCATCTCTTACCCTTAAACAATGTCAAAGTCTCGG**TTCA**  
**A**TCGGCGGATATTCTCAACGTTATAACGAAAACAATACCATCAACGATTTTTTAAAGCAAACACATCCAAACT  
ACTCCATCT**ATGGT**AGACAAATATTACGCCAAAGCGGTGATGCTATGTGCGGAGAACAAAACCTCGGGCGT  
GGGGTGCTGGTGT**TGAATCTGCC**GAAAAAGACAGCGACACCGCTGGTGTGCGGACTCTAGCATCAGTTA  
CGTTGTCTACACAAGTGTGGTGTGGCGTTGCCCTCAGGAGCTAACTTGACTGTGCTACAACCTATTAT  
GTAGAAATCTCTCTTGATG**TGAATGA**CCCATATTAG**GAATGCCA**ACTCT**TGAAT**TAAAGAAGAAGAGT  
CGTCTTATTTAGGCATTAACAACCTCGAACTTGTGCT**TTCAAT**TACAATGACTTTAA**GAATG**TATTTAATGT  
AAACAAAGAACTCGTATTGTCTTTCTCGAGTGGAGTTAAATTTGGAACTCAGGCGTCCACTCTTTCTTTG  
**AAAGATGA**CGCCGACTTATGGCTCGGTATATTAGTCTTCATCCATCCCAATATGCAAAAACCAATGCTA  
AAAATATCCTTCCATATGACGAATTTGTAGCCTACAAAACAACCTTTCTCTTCCGTCGG**ATGGT**GTAGG  
TGTA**AATGGGG**CTATGACGAACGTCAATTTCTATGCG**TCAAGT**GCCTGA**CAAGAT**TTACCTTG**TGGTAC**GC  
CCTCAATATCGGTACAAAAGCATATTGGTCAAATAACCTCAGCTACCCAATCTC**TCAAGT**TAAATATCA  
CT**TTCAACA**CAAAAGCGGGACTTCTATCTG**AAATGG**ATGCTTATTTCGCTTTACCAAATGTCCCGCGT**AA**  
**TGGT**TCTCAACAAACTTG**GAATG**AGTTCCGTGGCGTGGTTCGATCTGGGAACGGAGGGAA**ATTC**ACTTCT  
TTAGGAAGTATTATCGTCATCGATCCGGTCCGTGATTTAGGATTGAGCGATATGCTTTTCGAGTTCTAGTC  
TTGGACAATTTGGTTTCCAAGCGATTGTGACTACAGAACC**AATGGA**AAGGTATTGCTAACTCGGCCTTAGC  
CAATATTGAACTTTGTGTTCTTGCGAAC**ATGGT**GGTGTAAATGAT**TACTG**AAAGGGATCGAGTGCAGC**G**  
**ATG**AGTGGATTACTCACCAAATCAGAAGTATTAGAAGCGAAAGAAAAAGGAACTTCCAAAATTGAT**ATG**  
**AAGATG**TCGAAGCTATGTCGGGCGGTAATCTT**ATGA**AGAAAGGCGTGACCGCTCTCGGCGATGTGCTAAA  
ACGCAACAAAGGAGCAATCGGTAAAGGAGTTGCCCTCAGCGGTAGAGTCGGCTATTGGTGGAGCAAAAACG  
TCGTCGT**ATTC**AAACGTCTGGCGGTTCTCGGCTCTCCAAATATATGTAA

52 1724 3

Yellowstone Lake virophage 6, complete genome NCBI Reference Sequence: NC\_028270.1: c6005-4254 putative major capsid protein

ATGAGCGACTTAAAAACAGTTTTTAATTAATG**ATTC**AAAGAAATTG**AAGAT**ATAAGCGGTGAAATTGGGTTTT  
CCGTGCTAGGCGGTCTCAACAATCCACGT**TTCA**ACCACAACAGGCGAATACATCATCTAATACTTCC**CA**  
**TTCA**TGGCAAATTAATGTACCTAGTGAAAATATTGTCGTTGATAGAAGAATTTTACTTGA**TACTG**ATGTA  
GCATTTACAATTACAATTACAAATGTTCCAGTAGGTGAGGTAGCGTTTAAATTGGGGTAAAACTG**ATGGT**T  
TTGGACCTTAT**CCAT**TAAATAACTGTTATCTTACTCAAAATGTTTTAATTAATAATTGTGGTTTTATCTGT  
AAATACC**CAAGAT**ATAATCGCCCTTTACTTTCGTT**TGAATG**ACCAAAAAATACACACTTATTATGATGGA  
TGGACTCCATC**AATGGT**TGATGATACTTATTATAATTATTCTGATGCTGT**AGATG**CTAACAACAACCC**CA**  
**TT**GGAACGTGTTGTTAATAATAGTATGGATAATAAA**ATGAA**ACCACGAGGATGTCATACTATTAATCTAT  
GTTAATTAACATTATTTAAGC**ACTGC**ACCGCTTGTATTGCTGATGCTTCTGTTATTTCTACTAATGTA  
GCGGATACTTGGACTATTGATTTAACTACTCAATTT**TACTG**AACCTTTACTTTTTCTATCTCCTTTTATAT  
CTAATCCGTGCGTATCTAATGAGGCTGGATTTTTAGGAATTAATAATTTTT**ATTCA**ATTTACTATTGA  
TTCTAATTTACGTCGTATTTAGAACTTCTAATGCTTGGACAAATACAT**ATTC**AT**TC**CAA**ATTCA**CT  
**GC**ATTTACTAAATTACAATTTCTT**ATGAAT**TATCTAAGTGTCCAACCTTCCCAATATGCTAAAAATATCCG  
CTCGTAATGTTTTACCCTTTTACAATATGATAGATATATCGCTAACTCTAGTAATACTACCGCTAATAT  
TCCACCTGCTGTTGTTGGTGCTAATAATAGTATTACTTTAACCTCAACCCCTATACCTCATCTAATATT  
GCAT**TGAAT**CAAATCCAGACAC**CCAT**TTAATTTTTGGTGTGATTTCCAGTATCTAGTCAAACCGCTAAAA  
ATAGTGATG**CATTC**TATAAAATTGATTCTATTTCTATTAAC**TTCA**TAACGTAGCAGGTATTCTTAGTAG  
TGCTACACCTCACGAATTATGGGTATTATCTCAAAGGCGGGTTCATCTCAATCTTGG**GATGAAT**TTAGT  
GGACAAGTTAATAC**TTCAAG**TGCATCAGGTGCAAG**TTCA**ATTATTCCAACAACCTGG**TTCA**ATCTCGTTC  
TATCACCATC**AATGGT**AT**TC**AAATTTACCT**TGAAT**ACCTCAGTAGTTCCAGTTTAGGAAGTTATAATTTACA  
ATTTACTATTAATTTATCAAATCAAAGTTTAACTGCTTCTGCTATTGAAATCGTTGTTGTTACTATTACA  
TCAGGATACATCACAAACAC**TCAAG**GAACTACTCAATCTTTTACAGGTGAT**TGAAT**AAAAG**AGATG**GTAA  
TGAGTGCGAAGAAGGGTCCGGCTGTTCTCGTCTCGCACAACTGATT**ATGAA**CGCCTTGTGGTGGTGT

AAGAGATAATAGAGGAGTAATGAATATGATGAAACAATTTAAAAATAGACGTATGGGCTCAACTTCTGGC  
GGTGTAAATGTCTGGTGGTGTAAATGTCTGGTGGTATGCAAACGGTATGTCGTCTAGGAAAATATATTCGTTAA

61 1748 3.5

Yellowstone Lake virophage 7, complete genome NCBI Reference Sequence: NC\_028257.1: c17093-15336 putative major capsid protein

ATGGCAAATAATTCTATGAAAGTTATGAAGGTAATTGATGAGCGTATTAATGTTGCACAGGCTCCAGTGC  
TTATTGTAAATGAATCAACTCCAAGTGTGCAATATTCAACTATTAATGTTGCTGGTAATGCTTCACTTAC  
GCCTACTTTTTACAATTCCGTGTCCTCCAAAATCAAGGGTTAGATAGAACCGTGACGATGGCATTCACTGTT  
CAATTTTTTAATTACTIONGAACTGATTTGGAACAATTTCAAGCAAATCCAGCAATTGCATTGAGAGCGTGGC  
CATTACATAATGCATGCACATCTATGAATATTAATTTAGGTAATGCCGGTATGGTATTAATCCTAGTCA  
ATATGCTGCGGCATTATTAACACAATGGAATTGTGATAGTCATACACAAGCAACTGATTTATCATCATTT  
CCATCGGCTCCAGATAGATATTCAGTTATGCGGCCGCGGCATTATCAACTACATCAACATTTGATTTCTG  
ATTTTGCAGCTGTAAATTTCTGATTATGCTAATACATCAAGACTGGACAAATTACATCTATGAATATTA  
TGATGGCACATATGTATTGGGTGGAACCTGGAGCTACTCCAGTTGCATGGGGTGTCTCTGCAACTCAATG  
GTAATTACTIONGCAACGGTTTTATGAACCTATTGCCGCTTCTCCTTTTGTATAACTIONGAAATTAGAAATCCTA  
AAAAAGCATTTTTTGGATTAGGAAACGTACGGTGTGAGTTTGTAGCTTTAGTAATTTACAAAATGTTAAG  
TTATTATATTCCTGGAACTIONGCTACAATTACATCAATTCAGGACAATTTGCATCTCAATCTTTGTTAGTT  
TCTTACATATTCAGCATTTGAAGATAGTGTATCAAAATTAATGTATCTCCTATGGCATATAATTAATGCAACTC  
TTAGACTIONGCAACATCTGTTTTTACGGTTGGGGCTTCAACTGGTGGGGTAAATGGCATCTGCAACTIONGCTAA  
TTCTGCTACAATGCAATTAAGTGTGTTCCATCACACTTTTTAATATGGGCAACGCCTCAATGAATTTT  
GTTGAATCGCAAACACAATCATTACCTGATTTTTGTTTTACTCTTAGTAATGCAACTATTAATTTTTGCTG  
GTAAGATAACATTTTTAGGAACTAACTGTACTIONGATATCAACTATATAATATTAGTAAAAAAATGGTTC  
AAATACTTCATTTTTCGCAACTIONGACTGGAGCACAAATTTTATCTTCTTCAACACAAAATAATCCAGCAAAT  
CCACCACAATATTACGGTGGGGGAGTTTTAATTTTTAAAAACGAGTGAAGACCTTCGCTCCCTGCAACGA  
GTTGCGCTGGTATGAATCGTGCATTAACCTTCAAAATTCAGTGCATGCACAAATTAACTGATATCGC  
ATTTGGTGCAATTTGTCAATTAATGTGTTGGCAATACTIONGATGGTGTGTTGCTTACTCAACATGGTTC  
CGTGTCAACTTATTGAAGGTGGTATACTIONGAAGATATTTATAATAATGCTCCAGTTGTTTCTGGATTAG  
AAGAACTGGCCGTAAAAAGTTATACAAATAAAAATGGATATTCTGGAGGCAGTTGGGCATCATTTAAAAA  
TGATATGTCAAAATTTGCTGATTATATATCTCCGGTCACGAAACCTATTATTGGAGCATTGACAAAATAAA  
GCGGTTGCTGGAATGAATCCGGCGCTGGTCTGATGAATCGTGGTAAATTCGTGGATTATTGAAAGACA  
TGTATTAA

79 1754 4.5

Dishui Lake virophage isolate 5 putative major capsid protein gene, complete cds GenBank: KU245928.1

ATGGCTATGCTTCTGCTCTCGGACCATTCCATTCCATTCCTCTGTAATACCGTCCAAGCAAACATTAA  
TAACAATACCGTTTTCCAGAATATGCGAGATGTTATGTATCAGTTTCTCCGCTTTAATGACCCGAGAGAA  
CTTGACGATAACAATAACGCTTGTCCACCATGTATGACTCCTATTTTAATTAACAAGATGTCFCGGCA  
CTAATAACAACCCTCTCGAATAATGTAAGCAACGACCAAGATTTCAACAAGAGGTTGCTTTTCGTTCTTG  
ACAGCATACTIONGTAATGAACCCAAAGGTTCTATTGCTTCCAACACTCGTAATGTGATCTTGACCTTTAC  
TACAACCTGAACCTCTTATGCTTTTACCTTTTTATTTGGTGTGATCCTAAGAGCAACAACTAAGGTATGTAT  
GGCGTTCAACACTCAACTTTGTCTTCAACTCTTGGTCTTCCAATCGTATTGTTGCTTTGCTAATAAAG  
ATTTATTTGACGGCGGTGTTGCTACTATTTCTCTTCCCAACTIONGCAATTAACAACGCACAATFACTCAT  
GGAATTTTATAACAAGCAGCCAAGCGATTTGGTTTCCAGCCGTAATGTTGTCCCCTTCGAGAAATATCCC  
CGTTATTTAACCCCTATGACTTCTCTTGGTGTCTTTGCTGCTATAACCCTCAAACCTCAATTAGTTATTC  
CCTCTGTATCTGTAAATTTTTAGTCAACTIONGATTTGAATTTCTGTTCTGATAAACTTATCATTTGTTGCCG  
TAAAGTTCTCGCACAGCAAGCGCCTTTGATAGTGATAGTTTCTTACCTATTACCAATATAAGTATCAAT  
TTCAACAACAAGCAGGTCTGTTGTAAGTGTCTCGTCAAGTATGATCTATGGCGTATGAGTGTGGAATCAG  
GAAGCAACAACCAATTTACGGGTTTACCGGTTTCCGCAATATTGGAATAATGTTCTCTGTCAAAATGG  
TCTTCGTCAAGTCTAAGAGAGTTTCAACTGTGAGATCTGTTCTTTGTTTAGAGATGGGAGACATGTA  
GAATTAGACGATGTTTATGCCCCAGGAAGTATCGGGGCATTTCAACTTTCAATTTCCAAACTACATTTCAAA  
ACAATACCCTTGAAGCAATTGACGACCAAAAACAATCAGTATGAGATCGTGCTTATTACCATGAATAGCGG  
TGTCTTACAATTGAGCGTGGAACTTACAGACTTAACTIONGCTATTCTGTCCCGTGCCGATGTGCTTTCT  
GTTTCTTCTCAACCTTCTCAATTCAGTCTTCTGTTGCTCGTCTTGTGGTGGATCATGGGAAGATTCAT  
TCAAATCCCTTTGTTCTCAATTGCTCCTTATGCTGGTAAAGCAGAAAAAGTTAAAGATCTAATAAATGGG  
TGAAGGTGCGTCAGGTGGCGGTGAGTCAGGCGGTGATGAGAAAACATCTCGCCATGTAA

51 1457 3.5

Dishui Lake virophage isolate 6 putative major capsid protein gene, complete cds GenBank: KU245929.1

ATGTCTTCCAGCGATTTTGGAGAAAGTTTGCCTC**CAAGAC**GATGTGCTTAAACAC**TACTG**ACAAAGTGCGAT  
 ACGCAGTATTTAAGGGAGCACAGAATATTACTCCCGCACAAATACAAC**GCCATT**TCTACAAGCA**ATTC**AAAG  
 TATTACTTGGAA**ATTC**AGT**TGCC**AGT**GAA**T**TACTG**TATTCTCTCGTAGGATC**ATGGTA**AGGGCAA**CC**  
**ATTT**CAGTTAAAATTACGGGTGATTTGCGCGCCACCGCTCCCAACGGGTCTGTGC**TGGTA**AA**TATGGT**T  
 ATGCTTCCGCATTAGGTCCTTTCC**TTCA**ATCTCTTTGTAATACT**ATTC**AGGCAACTATTAACAATAA  
 CACAGTATCACAAAATCAAC**AGATG**TCATGTTCCAGTTGCTACGATTTAATGACCGCAGGGA**ACTTGC**C  
**AA**TATAACA**ACTCA**ACTCCCAACATGTATGATTCTTATTTAGATT**ATCA**AGATGCTCTTGGAACTAATA  
 ACAACCCACTTGGTGGTTGGAATAATGTAAGCAACGAC**CAAGATTTCA**AC**CAAG**AGGTGCTTTCGTTCT  
 TGACAGCAT**TACTG**TAAT**GAA**CCCAAAGGTTCTATTGCTTCCAACACTCGTAATGTGATCTTGACCTTT  
 ACTACA**ACTGA**ACCTCTTATGCTTTCACCTTTTATTTGGTGTGATCCTAAGAGCAACA**ATCA**AGGTATGT  
 ATGGCG**TTCA**AA**CM**CTCAACTTTGTCT**TTCA**ATCTTGGTCTTCCAAMTCGTATTGTTGCTCTTGCTAATA**A**  
**AGAT**TTATTTGACGGCGGTGTTGCTACTATTTCTCTTCCAA**TCA**AGCAATTAACAACGCACAAT**TACT**C  
 ATGGAATTTTATA**CAAG**ACAGCCAAGCGATTTGGT**TTC**AGCCGTAATGTTGTC**CCG**TTCGCAGAA**TATA**  
 CCCGTTATTTAACCCCTATGACTTCTTGGTGTCTT**GTCC**TATAACCCCTCAA**ACTCC**ATTAGTTAT  
 TCCCTCTGTATCTGTAAATTTT**CAGTCA**AT**TCA**GT**TGA**AT**TCT**GTTCCTGATAAACTTATCATTTGTTGC  
 CGTAAAGTTCTCGCACAG**CAAG**ACGCCTTTGATAGTGATAGTTTCTTACCTATTACCAATATAAGTATCA  
 AT**TTCA**ACAACAAGCAGGTCTGTT**TCA**AGTGCCTCGTCAGTATGATCTATGGCGTATGAGTGTGGAA**TC**  
 AGGAAGCAACCAACATGGCGGAATTTACGGGTTTCGCAAATATTGGAAATAATGTTCTCTGTCA**AA**T  
**GGT**CCTTCGT**CAGG**TGCTAAGAGAGT**TCCA**ACTTGTGGATCTGTTCTTTGTTTAG**AGATG**GGGAGACATG  
 TAGAATTAGACGATGTTTATGCCCCAGGAAGTATCGGGGCAT**TTCA**ACT**TTCA**ATTCCAAACTACAT**TTCA**  
**AA**ACAATACCCTTGAAGCAATTGACGACCAAAACAATCAGTATGAGATCGTGCTTATTACC**ATGA**ATAGC  
 GGTGTCTTACAATTGAGCGTGGA**ACTT**CACAGACTT**TACTG**CTATTCTGTCCCGTGCCGATGTGCTTT  
 CTGTTTCTTCTCAACCTTCT**CATT**C**CA**AGTCTTCTGTTGCTCGTCTTGTGGTGGATCATGG**GAAGAT**TC  
**ATTCA**AATCCCTTTGTTCTCAATTGCTCCTTATGCT**TGGTA**AGCAGAAAAAGTT**AA**AGATCTAAT**AA**TG  
**GGTGA**AGGTGCGTCAGGTGGCGGT**CAGT**CAGGCGGT**CGT**ATGAGAAAACATCTC**GCCAT**GTAA

65 1739 3.7

Observed mean proximity: 59.8/15≈4 Expected mean proximity: 2.15±0.8\* (\* 90%-confidence interval calculated from the minimal size of ligases, 908).  
 Data from <https://www.ncbi.nlm.nih.gov/nuccore/CP021507.1?report=genbank>

Standard deviation s for the minimal ligase: s=0.48. Then, the number X of standard deviations between observed and expected mean proximities equals:

$$X = (4 - 2.15) / 0.48 = 3.9$$

and, if the observed mean proximity O is a Gaussian random variable, following the Table of the distribution function below, we have :

$$\text{Proba}(O \geq 0.04) = \text{Proba}(X \geq 3.9) \approx 5 \cdot 10^{-5}$$

Table of the distribution  $\Pi(t) = P(X \geq t)$  of the  $N(0,1)$  random variable X

t	3	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4	4,5
$\Pi(t)$	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999928	0,999968	0,999997