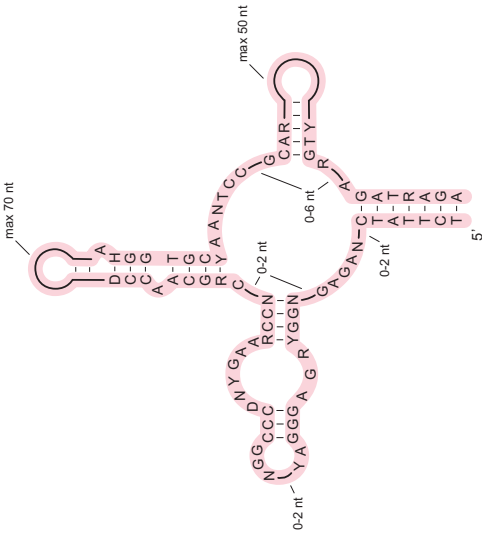


A



Accession	Metagenome	Start position	End position	Regulated function (COG)
CH095766	Sargasso	205	372	COG0646
CH110459	Sargasso	6821	6708	COG0646
CH158823	Sargasso	648	517	COG1897
AAF01048923	Soil	827	708	COG0460
AAF01049186	Soil	413	548	COG0498
AAF01054820	Soil	1103	966	COG0182
AAF01071693	Soil	991	868	COG0460
AAF01081827	Soil	78	236	COG0498
AAF01008477	Whale1	794	652	COG0646
AAF01018340	Whale2	185	307	No similar proteins
AAGA01004112	Whale3	792	939	COG2873
AAGA01012172	Whale3	589	469	COG2021

C

Structure	Accession	Metagenome	Start position	End position	Regulated function (COG)
CTTTGGCTTGGTTATC - AAGAA - AGACCGAGGGATT - GGGCCCGCTGACGCTT - TAGCAACCTTCTGGTAGGATTCGGAGCAAAAGC	CH095766	Sargasso	205	372	COG0646
TTAGTTTACTTATC - AAGAA - AGGTGAGGGAC - AGCCCTGTGAAGCCT - TGGCAACCTTT -	CH110459	Sargasso	6821	6708	COG0646
TACGTTATTTGTATC - GAGTGAATGAGGGATC - TGGCCCTTGAAGTGTGAGCAGCCCTTTC -	CH158823	Sargasso	648	517	COG1897
CGTGCAGACGCTTATC - CAGAG - TGGCTGAGGGAA - AGCCCTGTGAAGCCA - CAGCAACAGGGCC -	AAF01048923	Soil	827	708	COG0460
ATATCAACGGTCTCAT - CAGAG - TGGCAGAGGGAC - TGGCCCGATGAAGCCA - CGGCAACCGGCTCGATAA -	AAF01049186	Soil	413	548	COG0498
GGGGGACGCTTATC - GAGAG - TGGTGGAGGGAC - TGGCCCGAGCAAGCC - CGCAACCCCGGAACCGTC -	AAF01054820	Soil	1103	966	COG0182
GCGGGCAGCACTTATCCAG - TGGCTGAGGGAAAGCCGCTGAGCCA - CAGCAACAGGGCCGA -	AAF01071693	Soil	991	868	COG0460
ACTGAATACCGTCAATC - CAGAG - GGTGGAGGGAC - GGCCCTGGGAAGCC - CGCAACCCCGGCTGTGGCGTTCGGATCC	AAF01081827	Soil	78	236	COG0498
CGGATCCGTGCTTATC - AAGAA - AGGTGAGGGAA - AGCCCTGGGAACCT - TGGCAACCACTATAAGTCCCGAA -	AAF01008477	Whale1	794	652	COG0646
AATTGAAATCGATATC - GAGAGCGACGTGAGGGTTC - TGGCCAGTGAACGTC - CGCAACCCGCGCA -	AAF01018340	Whale2	185	307	No similar proteins
CATCAIGCATTCTCATC - AAGAA - AGGTGAGGGATC - AGTCCCTTGAAGCCT - TGGCAACCACTAGTACTTAA -	AAGA01004112	Whale3	792	939	COG2873
TTATTGAAACTGTTATA - AAGAA - AGGCAGAGGGATT - AGCCCGATGAAGCCT - TAGCAACCCCTTACTA -	AAGA01012172	Whale3	589	469	COG2021

Structure	Accession	Metagenome	Start position	End position	Regulated function (COG)
TGGCAACCTCCCGGTTTCGGTGCTACCTCCC - ACCAGATGAGCTAATCTCTCTTCTCTGGTAAGCGATAGTAAAAACCAGAG	CH095766	Sargasso	205	372	COG0646
-----GACTGGTCCGAATTCCTGATTTCCAAAT-----GTTGGAATAAGATAAGTTAAAAAGAGAT	CH110459	Sargasso	6821	6708	COG0646
-----GCATAAAGGTGCTAATGCCAG-----CCCAACCCGGCAA-----ATGTCGGTGGAGTCGATGGACTTATTGAA	CH158823	Sargasso	648	517	COG1897
-----AAGCCAGGTGTACTTCTC-----GCCGGATA-----AACCGGGA-----GAGATAAGAGCCTCGGAG	AAF01048923	Soil	827	708	COG0460
-----TTCGAGGTGACGAGTCCCAATCCA-----GCTCGAGCGA-----AAGTTCGGCAAAAAAATGAGAGAGCAGTTT	AAF01049186	Soil	413	548	COG0498
-----AACGGTTCGGCGGGTGCAGTCCGA-----CCTCATTTCCGG-----AGCGGATGGGAGAGATGAGAGTTCGAAAAG	AAF01054820	Soil	1103	966	COG0182
-----GAGCCCGGGTGTACTTCTC-----GCCGATC-----AACGGGA-----GAGATAAGAGCCTCGGTAG	AAF01071693	Soil	991	868	COG0460
TGGAGCCCGCCGGGAGATGGTCCCAACTCCG-----GCCITGGACCAAGG-----TGGCCGAGGG-----AAGATGGGAGAAAGGCCCT	AAF01081827	Soil	78	236	COG0498
-----GTTCCGATTTGAAAATCGGTGCTAAATCCTA-----TTCGCTTTA-----AGCGGA-----AAGATAAGTAGAAAAAT	AAF01008477	Whale1	794	652	COG0646
-----TGGGGCAAGGTCCCAATGCCA-----GCCCGGGA-----AACCGGGA-----ACGATGGTCCACATGGC	AAF01018340	Whale2	185	307	No similar proteins
-----TTAAGTGTAGAAAGGTGTACATCCT-----CCTTAGTCTTAAAA-----ACTTATGTGCTGAGA-----AAGATGGTCCGAGAGCCT	AAGA01004112	Whale3	792	939	COG2873
-----TTAAGAAAGGTGCTAAATTCCT-----ACCAATT-----TTTTGGA-----TAGATAACAAAAATGAAAT	AAGA01012172	Whale3	589	469	COG2021

Additional file 9: (A) SAM-riboswitch (S-box) pattern. (B) List of identified SAM-riboswitches. (C) Alignment of SAM-riboswitch sequences.