Biomphalaria glabrata GENOME INITIATIVE

rDNA gene region (18S-ITS1-5.8S-ITS2-28S) of *Biomphalaria glabrata*

Caveat

Use of the sequence data presented here is at your own discretion

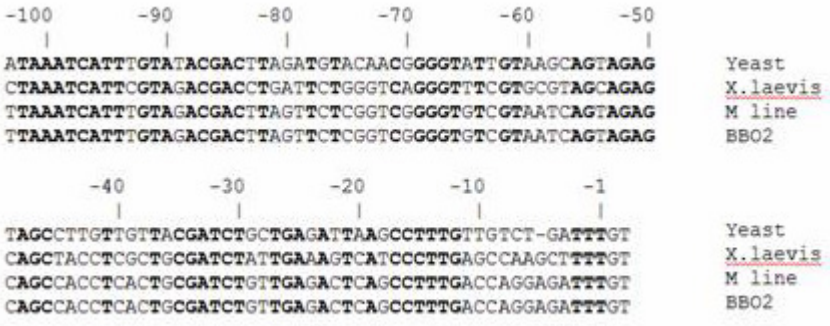
Ongoing gene discovery efforts (byproducts of EST sequencing at UNM; Adema et al.) and the WGS trace files generated during the feasibility studies phase of the genome sequencing at WUGSC (Clifton et al.) have each yielded muliple sequences representing the rDNA gene region of ***Biomphalaria glabrata***. Contig analysis of these independent data sets has yielded mutually confirmatory consensus sequences for the rDNA cassette of both M line and BB02 strain ***Biomphalaria glabrata***. Both sequences represent a majority consensus, and may differ somewhat from specific copies of rDNA gene cassettes that are tandemly arranged in the genome. Similarity to eukaryote signal sequences for termination of transcription of LSU rDNA is interpreted to indicate that the sequences putatively contain the complete, full-length 28S gene. This has **NOT** been confirmed experimentally, however.
Currently, GenBank does not offer a database format that can accept consensus sequences generated from contig alignments. Consequently, GenBank could not accept the sequence submissions shown below. As an alternative, the information is provided here for the benefit of the ***Biomphalaria*** research community, as it may be helpful for phylogenetic studies and for validation of EST data. Full effort has been given to assure that the data are correct, but realize that GenBank has not curated these sequences.

Therefore, **use of the sequence data presented here is at your own discretion**

(Note added in 2019) The below sequence data provide mutual confirmation for Zhang SM, Bu L, Laidemitt MR, Lu L, Mutuku MW, Mkoji GM, Loker ES (2018) Complete mitochondrial and rDNA complex sequences of important vector species of *Biomphalaria*, obligatory hosts of the human-infecting blood fluke, *Schistosoma mansoni*. Sci Rep. 2018 May 9;8(1):7341. doi: 10.1038/s41598-018-25463-z. PMID: 29743617).

Rationale for the assumption that the 28S gene of ***Biomphalaria glabrata*** is complete

The 28S sequences from ***Biomphalaria glabrata*** share similarity with eukaryote transcription termination sites at the 3' termini of the 26S LSU from *Saccharomyces carlsbergensis* (Yeast) and the 28S of *Xenopus laevis*. The alignment is based on data from **Veldman GM, Klootwijk J, de Jonge P, Leer RJ and Planta RJ. 1980 The transcription termination site of the ribosomal RNA operon in yeast. Nucleic Acids Research 8 (22), 5179-5192**. In the sequences below, the residue at position -1 is designated as the **putative** end of the 28S gene.. The pre-RNA may extend beyond the terminus of 28S. Alternative downstream termination sites may exist. Additionally, sequence length and sequence similarities with 28S sequences in GenBank support the assumption that the ***Biomphalaria glabrata*** 28S is complete.



Identical residues are in bold, the negative numbering identifies residue positions upstream of the **putative** end of transcription.

Biomphalaria glabrata M line strain rDNA

Use the below sequence at your own discretion.
This is a majority consensus of a contig-alignment of 1285 cDNA inserts (single pass sequences) representing rRNA sequences that resulted as a by-product of gene discovery efforts using RNA of M line ***Biomphalaria glabrata***.

18S-ITS1-5.8S-ITS2-28S (6674 nt: 1514 A 1683 C 2028 G 1549 T)
putative location of "hidden break splice region" in 28S determined from genomic DNA indicated in orange

CCTGCCA**GTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTTCACACTGTCTCACGGTG**
AAACCGCGAATGGCTCATTAAATCAGTCGATGTTTATTAGATGAAAACTATCCTACTTGGATAACTGTGG
CAATTCTAGAGCTAATACATGCTTACCAAGCTCCGACCCTCGTGGAAAGAGCGCTTTTATTAGTTCAAAA
CCAATCGCCGTTGCCCTTCGCGGGGCGCGGCGTCCCGATTGGTGACTCTGGATAACTTTGTGCTGATCGC
ATGGCCCTAGTGCTGGCGACGCATCTTTCAAATGTCTGCCCTATCAAATGTCGATGGTACGTGATATGCC
TACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACC
ACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAAC
AATACGGGACTCTTTCGAGGCCCCGTAATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCAATTG
GAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCAGTT
AAAAAGCTCGTAGTTGGATCTCAGGCGCAGGCGGGCGGTCCGGCTCGCGCCGGCTCACTGCCCGTTGTCT
CCTGCCCTACCCGTCTGCCGGCTCTCTCCCGTGGGTGCTCTTCGCTGAGCGTCCCGGGTGGCCGGCGCGT
TTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCCGAGCTGCCTGAATAATGGTGCATGGAATAATGGAA
TAGGACCTCGGTTCTATTTTGTTGGTTTTCGGAACTGGAGGTAATGATTAACAGGGACAAACGGGGGCAT
TCGTATTGCGGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGAACTACTGCGAAAGCATTTGTCAA
GAATGTTTTCATTAGTCAAGAACGAAAGTCAGAGGCGCGAAGACGATCAGATACCGTCGTAGTTCTGACC
ATAAACTATGCCACCTGGCCATCCGCAGGAGTTGCTTCGATGACTCTGCGGGCGGCCTCCGGGAAACCAA
AGGTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACC
AGGAGTGGAGCCTGCGGCTTAATTTGACCCAACACGGGAAAACTCACCCGGTCCGGACACTGTAAGGATT
GACAGATTGATAGCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATT
TGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTATTAAATAGTTCGCCGGTTCCTCGATGCGCCG
GCGCAACTTCTTAGAGGGACGAGTGGCGTTTAGCCACACGAAATTGAGCAATAACAGGTCTGTGATGCCC
TTAGATGTCCGGGGCCGCACGCGCGCTACACTGAAGGAATCAGCGTGGATGCCTCCCTGGCCCGAAAGGG
TTGGGAAACCCGTTGAATCTCCTTCGTGCTAGGGATTGGGGCTTGTAACTGTTCCCCATGAACGAGGAAT
TCCCAGTAAGCGCGAGTCATAAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTA
CTATCGATTGAGCGGTTCAGTGAGGGCCTCGGATTGGTCTCGGTCTGGCGTGCAAGCGCCGGCACCGTTG
GCCGAGAAGACGCTCGAACTCGATCGCTTGGAGAAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACC
TGCGGAAGGATCATTAAAGGTTTCTGTTGCTATCGACGATAACGAGAGCAGCAAAAAAATCAGCACAAAA
CAATACGTTTGTATGGCCCGGCCGGGGCGTCGAAGCGCACGAAGCGCCGCACTGGCTGGCACGACCGCTC
CTTTGTCGGGGTACCTACTTGTCCTCGATGCGACCCACGGTGGCGGCTTAGAGCCTTAGGGCTCGCCGGG

TCGCGAGGTTCAAAGAGTGGCCAGCTCTGTCTTGTACTCGCTCTCTATAGTACAACTGCTCTGAGACGGA
GTTGGCGACCGCCCCAATTGTTTTTTTTTTTTTTCTGTCCGTTTGAGAAAGTTTTCAGGGTACCTATGCCC
CTTTACGGCCGCTGGCACCGAACGCTCGGTAGGTGTAAGTGGAAGTCTTGAGGGTAGAGAGCTCCCACGC
TCGTCTTGTGGGCCGCGAGGTTTAAAGAGCCGGTCGGCTGCTGCTCTGTGGCAGGCCGGCCGCCCTGGCT
TTCTTACATGTTACCTTTACACAAACTCATAACACATTGTGATAAATTTTACTTTATGGTGGCAACGCTA
TACACAAAAAGTAACAACTTTGAGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAGCGCAGCCAGCTG
CGTGAATTAATGTGAATTGCAGAACACATTGAACATCGACATCTTGAACGCACATGGCGGCCTCGGGTCA
ATCCCGGGGCCACGTCCGTCTGAGGGTCGGTTTGCTAAAAGCGATCGCTTACTCGAAAGGGCACGCGGAT
TGTCCGTCTCTCGCTCTCCCGGAGAGTTTGACGAACTGCGAGCCTGCTGGTGTTATGGGTTTCCCCCGTA
TTGCGAGTGGCGCATTGGACCGTCGCGGATAGCCAGTCCCATCCATCTAGGAAGGGCTGGCATGTCTCCG
TGGTCTTAAGTACAGGATGCGCCGTCGTCCTCCTGTCCCTTTCGATGCTGCTAGCCTCGTCGGTGATCTC
TCACCGCAGGGCAGGACCCGGCTCGTATGTTAATCTTTACGTATGGGCTTTGCAGCTCTGTCCTGTACCC
GCTCTCTTAAGGTGTGAGATGGGCCACATGGCGGTTTTCGATTTGGGGCATGCGGCGCCCGCCGTGTTGT
ATTAGAACCGGCGTATTTTTTTATTTTAATCTCCGACCTCAGATCGGGCGAGATTACCCGCTGAATTTAA
GCATATAACTAAGCGGAGGAAAAGAAACTAACAAGGATTTCCTTAGTAACGGCGAGTGAAGCGGAAATAG
CCCAGCACCGAATCCCCCAGTGTCATGCTGGCGGGAACTGTGGTGTATGGGACGCCACCAGTCGTATGCG
CGGGGCATCGAAGTCCTCCTGATCGGGGCCATCACCCAGAGTGGGTGTAAGGCCTTTGCAGGTGCCTTGC
CGTGCGGCCGCGAGCGTCTCAGGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCGGGTGGTAAACTCCAT
CTAAGGCTAAATACATGCACGAGTCCGATAGCGAACAAGTACCGTGAGGGAAAGTTGAAAAGAACTTTGA
AGAGAGAGTTCAAGAGTACGTGAAACCGCTCAGAGGTAAACGGGTGGTTCCGCAAAAGTCAGCCCGTGGA
ATTCAGCGCGGCGCTCGGCCTGGGGCTGCTGCGGGTCGGGATCCCTCGGACCCACCTGCGGCGTCGCACT
GGGCCTCGCCGCGTGCACTTTCCATGGGCAGAGCGCTACGATCGGTTTCTCGGCGGCGATTAGCCCCGCT
GCTTCGCGTGGCACGGCCTTCGGGTCGTGCCTCGGGCGGCGGTGGTGAGCTGCTGGGGGACCGAGGAACG
GCCGTTCGCTTCGAGGCTGCTCGGCCTCGCGGGGGAGTTCGACTGGCGGGGACTGGGCAACCGTGCCCGC
CGACCGCTCTTCTGTGTTGGATCGGGCTCGCCGGGCGAGCATGGGGTCAGTGGCGAGTCTGTCGGCCTAC
CACCCGACCCGTCTTGAAACACGGACCAAGGAGTCTAACATGCGCGCGAGTCATTGGGCTGTACGAAACC
CAAAGGCGCAGTGAAAGCGAGGGTCGCCCCGGGCTGACCCAGGTGGGATCCTTTTGCCCCGTGCAGGAGG
CGCACCACCGGCCCGTCCCGTCTGCGTCGTCAGTGGGGCGGAGCTAGAGCGCGCACGCTGGGACCCGAAA
GATGGTGAACTATGCCTGAGTAGAACGAAGCCAGAGGAAACTCTGGTGGAAGTCCGTAGCGATTCTGACG
TGCAAATCGATCGTCAAACTTTGGTATAGGGGCGAAAGACTAATCGAACCATTTAGTAGCTGGTTCCCTC
CGAAGTTTCCCTCAGGATAGCTGGCGCTCGATCGCAGTTTTATCCGGTAAAGCGAATGATTAGAGGCCTT
GGGGACGAAACGACCTCAACCTATTCTCAAACTTTAAATGGGTAAGAAGTCCGGCTCGCTCGATTGGAGC
CGGGCCCTATTCGAATGCGTGTGCCAAGTGGGCCACTTTTGGTAAGCAGAACTGGCGCTGTGGGATGAAC
CAAACGCTCGGTTAAGGTGCCAAACGCAGACGCTCATCAGACACCATAAAAGGTGTTGGTTGATTTAGAC
AGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACTCACCTGCCGAATCAACCA
GCCCTGAAAATGGATGGCGCTAGAGCGTCGGACCTATACCGGGCCGTCTCGGCAATCGAGTCTGGAAGAC
AGGCTGAGACGAGTAGGAGGGCCGCTGGGGTGAGCGTGGAAGCCTGGGAAGCGATTCTGGGTGGAGCCGC
CCCGGGTGCAGATCTTGGTGGTAGTAGCAAATATTCAAACGAGAACTTTGAAGACTGAAGTGGAGAAGGG
TTCCATGTGAACAGCAGTTGAACATGGGTCAGTCGGTCCTAAGAGATAGGAAAACTCCGTTCTGACTCGG
GGGCAAAGCTAATATTATATTACTATTACGAGTAGTATTAATTGGTTGCTCGACTCTTATCGAAAGGGAA
TCGGGTTAATATTCCCGAACCTGGACACGGAGATTGGTCCTTTGGGGCCACGTGCGGCAACGCAAACGAA
GTGGGGGACGTCGGCGGGAGCCCCGGGAAGAGTTCTCTTTTCTTTGTAAGGGGCTTCATCCCTGGAATCG
GCTTGCCCGGAGATAGGGACAATGGCTCCGTAAAGCACCGCGGCTCTTGCGGTGTCCGGTGCGCTCCTGT
CGGCCCTTGAAAACCCCACGGAGACGGTGTGAATTTCGTGCCAGGCCGTACCCATATCCGCAGCAGGTCT
CCGAGGTGAACAGCCTCTAGTCGATAGAACAATGTAGGTAAGGGAAGTCGGCAAATTGGATCTGTAACTT
CGGGAAAAAGATTGGCTCTGAGGGCTGGGCCAGTCGGGCCGGGGTACGAAGCGGGACTGGGATGAGCCCG
GGCTGGGCGAGGCCGCCCGTCGCTAGTCGACGGGGCCGGCCGAGCTCGGAACGCGGCTTCAACCTTCCCG
TGGACTGCCTCAGCTATGCGGCGGCTCCTCGCGGGTCGTCCGCGTCGGCTGGCATTCAACAGCCAACTCA
GAACTGGTACGGACCAGGGGAATCCGACTGTCTAATTAAAACAAAGCATTGCGACGGCCGTCACCCGGTG
TTGACGCAATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAACCAAGCGCGGGTAAA
CGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAAT
GGATTAACGAGATTCCCACTGTCCCTATCTACTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGGTA
GAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCCGACTTTGTGAAGACACATGAAGGGTG
TAGCATAGGTGGGAGCGCAAGCGCCATTGAAATACCACTACTTTTATCGTTTCTTTACTTATTCAGTAAA

GCGGAGAGCGGGGTGCAAGCCCCTCGATTCTGGAGTTAAGCGGCCGCATGGCTCCGTTCGTTAGCCCCTC
TTCACCGGGGGGCGCGGGCGGGGTCGGTCGCGATCCGCTCTGAAGACAGTGTCAGGCGGGGAGTTTGACT
GGGGCGGTACATCTGTCAAAAGGTAACGCAGGTGTCCTAAGGCGAGCTCAGCGAGGACGGAAACCTCGCG
TAGAGCAAAAGGGCAAAAGCTCGCTTGATTTTGATTTTCAGTACGAATACAGACCGTGAAAGCGTGGCCT
ATCGATCCTTTTGACTTTAAGAGTTTTAAGCAAGAGGTGTCAGAAAAGTTACCACAGGGATAACTGGCTT
GTGGCAGCCAAGCGTTCATAGCGACGTTGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGCGAA
GCAGAATTCGCCAAGCGTTGGATTGTTCACCCACTAATAGGGAACGTGAGCTGGGTTTAGACCGTCGTGA
GACAGGTTAGTTTTACCCTACTGATGACAAGTCGTTGCTACGGTAATCCTGCTCAGTACGAGAGGAACCG
CAGGTTCAGACATTTGGTTCACGTGCTTGGCTGATAAGCCAATGGTGCGAGGCTACCATCTGAGGGATTA
TGGCTGAACGCCTCTAAGTCAGAATCCCGCCCGGATATGCGACGATATCTTTCGCACCCCTGCGTCGGGA
GGCAACGATATGGCTGGGTTGTTTGGGGCCGCTACTCCTCGGGGTGGTGGCACCGGCCCGGCCTGGTCGA
GCCGCAGTAAGAGGTCTTTGCCGTTCCGCCGATCAGTGCGGGCGGGGACCTGGACGATATTCACCCCATG
CGGCGTGGAGGTGTTAAATCATTTGTAGACGACTTAGTTCTCGGTCGGGGTGTCGTAATCAGTAGAGCAG
CCACCTCACTGCGATCTGTTGAGACTCAGCCTTTGACCAGGAGATTTGTTCGCT

Biomphalaria glabrata BB02 strain rDNA

Use the below sequence at your own discretion

The above M line sequence was BLASTed against the ***Biomphalaria glabrata*** WGS data in the NCBI trace-file archive. Reiterative BLASTing with the resulting contiguous WGS sequences was performed to expand the termini of the sequence contig. Below is the minimal set of (DDBJ/EMBL/GenBank) WGS entries used for this majority consensus of genomic DNA:

GNL|TI|591561635, BGAA-AAB06A11.G1, GNL|TI|1021177758, BGAC-AAB70F07.B1,
GNL|TI|813160879, BGAC-AAA62F03.B, GNL|TI|591556337, BGAA-AAA77E12.B1,
GNL|TI|591561780, BGAA-AAA60E11.B1, GNL|TI|813155926, BGAC-AAA19F01.B1,
GNL|TI|591561165, BGAA-AAA60E11.G1, GNL|TI|591553978, BGAA-AAA19F12.B1,
GNL|TI|591564370, BGAA-AAA77E12.G1, GNL|TI|813155027, BGAC-AAA05F03.B1,
GNL|TI|591562324, BGAA-AAA60C12.B1, GNL|TI|813161043, BGAC-AAA70B12.B1,
GNL|TI|591557051, BGAA-AAA37B10.B1, GNL|TI|591550672, BGAA-AAA22G05.B1,
GNL|TI|591549995, BGAA-AAA09B12.B1, GNL|TI|591557001, BGAA-AAA78A05.G1,
GNL|TI|591550170, BGAA-AAA28B09.B1, GNL|TI|813160194, BGAC-AAA73A08.G1,
GNL|TI|591559985, BGAA-AAA37B10.G1, GNL|TI|591563063, BGAA-AAA60C12.G1,
GNL|TI|591552794, BGAA-AAA22G05.G1, GNL|TI|591550438, BGAA-AAA09B12.G1,
GNL|TI|813154902, BGAC-AAA04C05.B1, GNL|TI|591564687, BGAA-AAA78A05.B1,
GNL|TI|591558090, BGAA-AAA43D11.B1, GNL|TI|591564320, BGAA-AAA44G09.G1,
GNL|TI|813159116, BGAC-AAA19F01.G1, GNL|TI|591556199, BGAA-AAA21E12.G1,
GNL|TI|813159524, BGAC-AAA62F03.G1, GNL|TI|591556604, BGAA-AAA16D09.G1,
GNL|TI|591549714, BGAA-AAA16D09.B1, GNL|TI|591564756, BGAA-AAA41A10.B1,
GNL|TI|813155271, BGAC-AAA16B08.G1, GNL|TI|591554242, BGAA-AAB16F02.B1,
GNL|TI|1021182168, BGAC-AAC72C04.G1, GNL|TI|591563786, BGAA-AAA68E12.G1,
GNL|TI|1021177759, BGAC-AAB70F08.B1, GNL|TI|1021182171, BGAC-AAC72C07.G1,
GNL|TI|591560672, BGAA-AAA68E12.B1

NTS(incomplete)-18S-ITS1-5.8S-ITS2-28S-IGS(incomplete),
(8007 nt: 1846 A 1984 C 2305 G 1870 T 2 others)
putative location of "hidden break splice region" in 28S indicated in orange

CATCGAGTAAGCGTAGCAAACCAGCACTTAAAATGCTAACCAGCTCAGGATGCTCATGCATATAGGGACC
AGCGTGACCTGAGCGGATTCAGTACTTGATCGCATGAGTCCGTACCACGTAGGGAATGAGATCCGCGAGC
TACAGAGAGGTCAGAGCGTAGGCCTGGTAGAACCCGGTCCCTCGCAACGGGGGTCTGGGTTGTCAAAACG
AAATGGGTCCCTTAGCACGAACGAGGGTCGATTCTCCAGGCGCAGATCTGGCTCCCATCTGTGTTTAGTT

TGCCGAATAAAACATGAGGGGACGCAAAAGGAACCTGTTCTTGAAAGGACCGCATTCTCCCGCCCTGAGC
AGGCGCTGGCTTATGAAGGCTAGCGCTGTGACTGTTTTTAGTAAGTCGGGAATCTTTTAGGAGTAGCGGC
CCGGCGACGATGATAGCTATCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATG
CATGTCTAAGTTCACACTGTCTCACGGTGAAACCGCGAATGGCTCATTAAATCAGTCGATGTTTATTAGA
TGAAAACTATCCTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCTTACCAAGCTCCGACCCTC
GTGGAAAGAGCGCTTTTATTAGTTCAAAACCAATCGCCGTTGCCCTTCGCGGGGCGCGGCGTCCCGATTG
GTGACTCTGGATAACTTTGTGCTGATCGCATGGCCCTAGTGCTGGCGACGCATCTTTCAAATGTCTGCCC
TATCAAATGTCGATGGTACGTGATATGCCTACCATGTTTGTAACGGGTAACGGGGAATCAGGGTTCGATT
CCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAACTTACCCACTCCC
GGCACGGGGAGGTAGTGACGAAAAATAACAATACGGGACTCTTTCGAGGCCCCGTAATTGGAATGAGTAC
ACTTTAAACCCTTTAACGAGGATCAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTC
CAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTCAGGCGCAGGCGGGCGGTCC
GGCTCGCGCCGGCTCACTGCCCGTTGTCTCCTGCCCTACCCGTCTGCCGGCTCTCTCCCGTGGGTGCTCT
TCGCTGAGCGTCCCGGGTGGCCGGCGCGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCCGAGCTG
CCTGAATAATGGTGCATGGAATAATGGAATAGGACCTCGGTTCTATTTTGTTGGTTTTCGGAACTGGAGG
TAATGATTAACAGGGACAAACGGGGGCATTCGTATTGCGGCGTTAGAGGTGAAATTCTTGGATCGCCGCA
AGACGAACTACTGCGAAAGCATTTGTCAAGAATGTTTTCATTAGTCAAGAACGAAAGTCAGAGGCGCGAA
GACGATCAGATACCGTCGTAGTTCTGACCATAAACTATGCCACCTGGCCATCCGCAGGAGTTGCTTCGAT
GACTCTGCGGGCGGCCTCCGGGAAACCAAAGGTTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAA
CTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAA
ACTCACCCGGTCCGGACACTGTAAGGATTGACAGATTGATAGCTCTTTCTTGATTCGGTGGGTGGTGGTG
CATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTATT
AAATAGTTCGCCGGTTCCTCGATGCGCCGGCGCAACTTCTTAGAGGGACGAGTGGCGTTTAGCCACACGA
AATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCGCACGCGCGCTACACTGAAGGAATC
AGCGTGGATGCCTCCCTGGCCCGAAAGGGTTGGGAAACCCGTTGAATCTCCTTCGTGCTAGGGATTGGGG
CTTGTAACTGTTCCCCATGAACGAGGAATTCCCAGTAAGCGCGAGTCATAAGCTCGCGTTGATTACGTCC
CTGCCCTTTGTACACACCGCCCGTCGCTACTATCGATTGAGCGGTTCAGTGAGGGCCTCGGATTGGTCTC
GGTCTGGCGTGCAAGCGCCGGCACCGTTGGCCGAGAAGACGCTCGAACTCGATCGCTTGGAGAAAGTAAA
AGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAAAGGTTTCTGTTGCTATCGACGATA
ACGAGAGCAGCAAAAAAATCAGCACAAAACAAACGTTTGTATGTCCCGGCCGGGGCGTCGAAGCGCACGA
AGCGCCGCACTGGCTGGTACGACCGCTCCTTTGTCGGGGTACCTACTTGTCCTCGATGCGACCCACGGTG
GCGGCTTAGAGCCTTAGGGCTCGCCGGGTCGCGAGGTTCAAAGAGTGGCCAGCTCTGTCTTGTACTCGCT
CTCTATAGTACAACTGCTCTGAGGACGGAGTTGGCGACCGCCCCAATTTTTTTTTTTTTTTCTGTCCGT
TTGAGAAAGTTTTCAGGGTACCTATGCCCCTTTACGGCCGCTGGCACCGAACGCTCGGTAGGTGTAAGTG
GAAGTCTTGAGGGTAGAGAGCTCCCACGCTCGTCTTGTGGGCCGCGAGGTTTAAAGAGCCGGTCGGCTGC
TGCTCTGTGGCAGGCCGGCCGCCCTGGCTTTCTTACATGTTACCTTTACACAAACTCATAACACATTGTG
ATAAATTTTTACATTTCTGGTGGCAACGCTATACACAAAAAGTAACAACTTTGAGCGGTGGATCACTCGG
CTCGTGCGTCGATGAAGAGCGCAGCCAGCTGCGTGAATTAATGTGAATTGCAGAACACATTGAACATCGA
CATCTTGAACGCACATGGCGGCCTCGGGTCAATCCCGGGGCCACGTCCGTCTGAGGGTCGGTTTGCTAAA
AGCGATCGCTTACTCGCAAGGGCACGCGGATTGTCCGTCTCTCGCTCTCCCGGAGAGTTTGACGAACTGC
GAGCCTGCTGGTGTTATGGGTTTCCCCAAACCCGTATTGCGAGTGGCGCATTGGACCGTCGCGGATAGCC
AGTCCCATCCATCTAGGAAGGGCTGGCATGTCTCCGTGGTCTTAAGTACAGGATGCGCCGTCGTCCTCCT
GTCCCTTTCGATGCTGCTAGCCTCGTCGGTGATCTCTCACCGCAGGGCAGGACCCGGCTCGTATGTTAAT
CTTTACGTATGGGCTTTGCGGACCTAGCCTTGCACTGCTCTCTTAAGGTGTGAGATGGGCCACATGGCGG
TTTTCGATTTGGGGCTTGCGGCGCCCGCCGTGTTGTATTAGCACCGGCGTATTTTTTTATTTTAATCTCC
GACCTCAGATCGGGCGAGATTACCCGCTGAATTTAAGCATATAACTAAGCGGAGGAAAAGAAACTAACAA
GGATTTCCTTAGTAACGGCGAGTGAAGCGGAAATAGCCCAGCACCGAATCCCCCAGTGTCATGCTGGCGG
GAACTGTGGTGTATGGGACGCCACCAGTCGTATGCGCGGGGCATCGAAGTCCTCCTGATCGGGGCCATCA
CCCAGAGTGGGTGTAAGGCCTTTGCAGGTGCCTTGCCGTGCGGCCGCGAGCGTCTCAGGAGTCGGGTTGT
TTGGGAATGCAGCCCAAAGCGGGTGGTAAACTCCATCTAAGGCTAAATACATGCACGAGTCCGATAGCGA
ACAAGTACCGTGAGGGAAAGTTGAAAAGAACTTTGAAGAGAGAGTTCAAGAGTACGTGAAACCGCTCAGA
GGTAAACGGGTGGTTCCGCAAAAGTCAGCTCGTGGAATTCAGCGCGGCGCTCGGCCTGAGGCTGCTGCGG
GTCGGGATCCCTCGGACCCACCTGCGGCGTCGCACTGGGCCTCGCCGCGTGCACTTTCCATGGGCAGAGC
GCTACGATCGGTTTCTCGGCGGCGATTAGCCCCGCTGCTTCGCGTGGCACGGCCTTCGGGTCGTGCCTCG

GGCGGCGGTGGTGAGCTGCTGGGGGACCGAGGAACGGCCGTTCGCTTCGAGGCTGCTCGGCCTCGCGGGG
GAGTTCGACTGGCGGGGACTGGGCAACCGTGCCCGCCGACCGCTCTTCTGTGTTGGATCGGGCTCGCCGG
GCGAGCATGGGGTCAGTGGCGAGTCTGTCGGCCTACCACCCGACCCGTCTTGAAACACGGACCAAGGAGT
CTAACATGCGCGCGAGTCATTGGGCTGTACGAAACCCAAAGGCGCAGTGAAAGCGAGGGTCGCCCCGGGC
TGACCCAGGTGGGATCCTTTTGCCCCGTGCAGGAGGCGCACCACCGGCCCGTCCCGTCTGCGTCGTCAGT
GGGGCGGAGCTAGAGCGCGCACGCTGGGACCCGAAAGATGGTGAACTATGCCTGAGTAGAACGAAGCCAG
AGGAAACTCTGGTGGAAGTCCGTAGCGATTCTGACGTGCAAATCGATCGTCAAACTTTGGTATAGGGGCG
AAAGACTAATCGAACCATTTAGTAGCTGGTTCCCTCCGAAGTTTCCCTCAGGATAGCTGGCGCTCGATCG
CAGTTTTATCCGGTAAAGCGAATGATTAGAGGCCTTGGGGACGAAACGACCTCAACCTATTCTCAAACTT
TAAATGGGTAAGAAGTCCGGCTCGCTCGATTGGAGCCGGGCCCTATTCGAATGCGTGTGCCAAGTGGGCC
ACTTTTGGTAAGCAGAACTGGCGCTGTGGGATGAACCAAACGCTCGGTTAAGGTGCCAAACGCAGACGCT
CATCAGACACCATAAAAGGTGTTGGTTGATTTAGACAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCT
AAGGAGTGTGTAACAACTCACCTGCCGAATCAACCAGCCCTGAAAATGGATGGCGCTAGAGCGTCGGACC
TATACCGGGCCGTCTCGGCAATCGAGTCTGGAAGACAGGCTGAGACGAGTAGGAGGGCCGCTGGGGTGAG
CGTGGAAGCCTGGGAAGCGATTCTGGGTGGAGCCGCCCCGGGTGCAGATCTTGGTGGTAGTAGCAAATAT
TCAAACGAGAACTTTGAAGACTGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCAGTC
GGTCCTAAGAGATAGGAAAACTCCGTTCTGACTCGGGGGCAAAGCTAATATATATACTATACGAGTAGTA
TTAATTGGTTGCTCGACTCTTATCGAAAGGGAATCGGGTTAATATTCCCGAACCTGGACACGGAGATTGG
TCCTTTGGGGCCACGTGCGGCAACGCAAACGAAGTGGGGGACGTCGGCGGGAGCCCCGGGAAGAGTTCTC
TTTTCTTTGTAAGGGGCTTCATCCCTGGAATCGGCTTGCCCGGAGATAGGGACAATGGCTCCGTAAAGCA
CCGCGGCTCTTGCGGTGTCCGGTGCGCTCCTGTCGGCCCTTGAAAACCCCACGGAGACGGTGTGAATTTC
GTGCCAGGCCGTACCCATATCCGCAGCAGGTCTCCGAGGTGAACAGCCTCTAGTCGATAGAACAATGTAG
GTAAGGGAAGTCGGCAAATTGGATCTGTAACTTCGGGAAAAAGATTGGCTCTGAGGGCTGGGCCAGTCGG
GCCGGGGTACGAAGCGGGACTGGGATGAGCCCGGGCTGGGCGAGGCCGCCCGTCGCTAGTCGACGGGGCC
GGCCGAGCTCGGAACGCGGCTTCAACCTTCCCGTGGACTGCCTCAGCTATGCGGCGGCTCCTCGCGGGTC
GTCCGCGTCGGCTGGCATTCAACAGCCAACTCAGAACTGGTACGGACCAGGGGAATCCGACTGTCTAATT
AAAACAAAGCATTGCGACGGCCGTCACCCGGTGTTGACGCAATGTGATTTCTGCCCAGTGCTCTGAATGT
CAAAGTGAAGAAATTCAACCAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAGGTAGCCAA
ATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATTAACGAGATTCCCACTGTCCCTATCTACTATC
TAGCGAAACCACAGCCAAGGGAACGGGCTTGGTAGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGAC
TCTAGTCCGACTTTGTGAAGACACATGAAGGGTGTAGCATAGGTGGGAGCGCAAGCGCCATTGAAATACC
ACTACTTTTATCGTTTCTTTACTTATTCAGTAAAGCGGAGAGCGGGGTGCAAGCCCCTCGATTCTGGAGT
TAAGCGGCCGCATGGCTCCGTTCGTTAGCCCCTCTTCACCGGGGGGCGCGGGCGGGGTCGGTCGCGATCC
GCTCTGAAGACAGTGTCAGGCGGGGAGTTTGACTGGGGCGGTACATCTGTCAAAAGGTAACGCAGGTGTC
CTAAGGCGAGCTCAGCGAGGACGGAAACCTCGCGTAGAGCAAAAGGGCAAAAGCTCGCTTGATTTTGATT
TTCAGTACGAATACAGACCGTGAAAGCGTGGCCTATCGATCCTTTTGACTTTAAGAGTTTTAAGCAAGAG
GTGTCAGAAAAGTTACCACAGGGATAACTGGCTTGTGGCAGCCAAGCGTTCATAGCGACGTTGCTTTTTG
ATCCTTCGATGTCGGCTCTTCCTATCATTGCGAAGCAGAATTCGCCAAGCGTTGGATTGTTCACCCACTA
ATAGGGAACGTGAGCTGGGTTTAGACCGTCGTGAGACAGGTTAGTTTTACCCTACTGATGACAAGTCGTT
GCTACGGTAATCCTGCTCAGTACGAGAGGAACCGCAGGTTCAGACATTTGGTTCACGTGCTTGGCTGATA
AGCCAATGGTGCGAGGCTACCATCTGAGGGATTATGGCTGAACGCCTCTAAGTCAGAATCCCGCCCGGAT
ATGCGACGATATCTTTCGCACCCCTGCGTCGGGAGGCAACGATATGGCTGGGTTGTTTGGGGCCGCTACT
CCTCGGGGTGGTGGCACCGGCCCGGCCTGGTCGAGCCGCAGTAAGAGGTCTTTGCCGTTCCGCCGATCAG
TGCGGGCGGGGACCTGGACGATATTCACCCCATGCGGCGTGGAGGTGTTAAATCATTTGTAGACGACTTA
GTTCTCGGTCGGGGTGTCGTAATCAGTAGAGCAGCCACCTCACTGCGATCTGTTGAGACTCAGCCTTTGA
CCAGGAGATTTGTTCGCTTCAGACGGACAACATCTCCGACTCGTGGTCATTGGCCAGCCTCTGTACTGGC
ACGTACGTAGGCGAACGAATCTTGCAGGTTCGACTCCTGCCCCGGGCAATCATTACTTTACGGGACAAGC
TTTTCTATGGCAGTGGGTAAATTTAGCCCGGTGACCCACGATTCGTTCCTTACAACTTGCAAGTCAGGGG
ACTTGGTGTGAGCGTGGCGCTGTCTGGCTGATATCTCTCGCTCGGTATCAGCCGCGACACAATWTTTTTT
TTTTTTCTCTAAGTCCAAAAGACATAGTGCACATGCACAAGCTTCAGCGCTGCGATGGCGTAGAGTCCAG
TCACACAGACCAAGACCAACTATAAGTGTCAGGCCAGCCAMGCCAGTATGCCAATGTATAATGTAACTTG
TGTTTTCTTTTTTTTCCCTCCAAGTCCGAAAGACATAGTGCATGAACAGAAAGACTGTCATGCGCGTCTT
TTAAAAAAATTTTATTTTCCAAGTCCAAAAGACATAGTGCATACATTTTTAAACATAGTGCATGAACAGA
AAGACATAGTGTATAGTCTCATACGCGCGCGCCTTTTGAATTTTTTTTTTTAAGTCCAAAAGACATAGTG

CATATACATTTCAAAATTGAAGACATAGTGCACAACATCCATCCTCGCGCGTTTTCAGTACCGCCAATTA
TACAGTTATCTGATTAATTTTCCTCCAAACTACATAACCCTTAACTCACCCAAACTCCTATCTACCCATC
CACCGCACCAGCGAAATAATATATGTC