

## Homework #6

Due April 30<sup>th</sup>, 11:59pm

---

Each homework submission must include:

- An archive (.zip or .gz) file of the source code containing:
  - The makefile used to compile the code on Monsoon **(5pts)**
  - All .cpp and .h files **(5pts)**
  - A readme.txt file outlining all modules (if any) needed for the execution of the code and the exact command lines needed to answer homework's questions **(5pts)**
- A full write-up (.pdf or .doc) file containing answers to homework's questions **(5pts)** – screenshots of code output are ok.

The source code must follow the following guidelines:

- No external libraries that implement data structures discussed in class are allowed, unless specifically stated as part of the problem definition. Standard input/output and utilities libraries (e.g. math.h) are ok.
  - All external data sources (e.g. input data) must be passed in as a command line argument (no hardcoded paths within the source code).
  - Solutions to sub-problems must be executable separately from each other. For example, via a special flag passed as command line argument **(5pts)**
-

### Problem #1 (of 2): Suffix trie

Create a class called ***suffix\_trie***. Use the **suffix trie (not tree)** data-structure to store a given genome (see Appendix A). At minimum, the class must contain:

- A default constructor
- At least one custom constructor to build a trie from a single sequence (here genome in appendix A)
- A function to traverse (*search*) the trie using a sequence fragment of a given length.
- A destructor

#### A. Implementing a basic (perfect match only) search function

- Read in the SARS-COV2 genome sequence (Appendix A) and store it in the suffix trie.
- Implement a perfect match search function, which would take in a sequence fragment, search for presence of that fragment in the suffix trie and report presence/absence of the fragment in the genome (finding a single hit is ok – no need to find all possible hits).
- Use read fragment in appendix B to test your search function.

#### B. Stress testing your search function

- What is the size of the trie (# of nodes)?
- Generate 5K, 50K, and 100K random 36-mers from the SARS-COV2 genome sequence and use them to search the suffix tree.
  - i. How many of your 36-mers had a match? Does it make sense? Explain why.
  - ii. How long did it take (big O notation estimate).

### Problem #2 (of 2): Suffix tree – extra credit (50 points)

Create a class called ***suffix\_tree***. Use the **suffix tree (not trie)** data-structure to store a given genome (see Appendix A). You must construct the suffix tree without building the trie first (i.e. use the trick shown in class). At minimum, the class must contain:

- A default constructor
  - At least one custom constructor to build a tree from a given genome sequence
  - A function to traverse (*search*) the tree using a sequence fragment of a given length.
  - A destructor
- A. Implementing a basic (perfect match only) search function
- Read in the SARS-COV2 genome sequence (Appendix A) and store it in the suffix tree.
  - Implement a perfect match search function, which would take in a sequence fragment, search for presence of that fragment in the suffix tree and report presence/absence of the fragment in the genome (finding a single hit is ok – no need to find all possible hits).
  - Use read fragment in appendix B to test your search function.
- B. Stress testing your search function
- What is the size of the tree (# of nodes)?
  - Generate 5K, 50K, and 100K random 36-mers from the SARS-COV2 genome sequence and use them to search the suffix tree.
    - i. How many of your 36-mers had a match? Does it make sense? Explain why.
    - ii. How long did it take (big O notation estimate). Compare with estimates in Problem #1

## Appendix A: SARS-COV2 Genome

>NC\_045512.2\_SARS\_COV2\_complete\_genome

ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAA  
CGAAGCTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAAC  
TAATTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTG  
TTGCAGCCGATCATCAGCACATCTAGGTTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTC  
CCTGGTTTCAACGAGAAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTCGCGACGTGCTCGTAC  
GTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG  
CTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGAT  
GCTCGAAGTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAAGTGAAGGCATTTCAGTACGGTC  
GTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGGCGAAATACCAGTGGCTTACCGCAAGGTTCT  
TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA  
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTTCAAGAAAACCTGGAACACTAAACATAGCAGTGGTG  
TTACCCGTGAAGTCAATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTTCTGTGG  
CCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTG  
TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTG  
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAAATTAAATTGGCAAAGAA  
ATTTGACACCTTCAATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAA  
CCAAGGGTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCAC  
CAAATGAATGCAACCAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTGGTGAAACTTCATGGCA  
GACGGGCGATTTTGTAAAGCCACTTGCGAATTTTGTGGCACTGAGAATTTGACTAAAGAAGGTGCCACT  
ACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTTATTGTCCAGCATGTCACAATTCAGAAGTAG  
GACCTGAGCATAGTCTTGCCGAATACCATAATGAATCTGGCTTGAAGAACCTTCTTCGTAAGGGTGGTCG  
CACTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAGTGTGCCTATTGGGTTCCA  
CGTGCTAGCGCTAACATAGGTTGTAACCATAACAGGTGTTGTTGGAGAAGGTTCCGAAGGTCTTAATGACA  
ACCTTCTTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGACTTTAACTTAATGAAGA  
GATCGCCATTATTTTGGCATCTTTTTCTGCTTCCACAAGTGCTTTTGTGGAACTGTGAAAGGTTTGGAT  
TATAAAGCATTAACAATTTGTTGAATCCTGTGGTAATTTTAAAGTTACAAAAGGAAAAGCTAAAAAAG  
GTGCCTGGAATATTGGTGAACAGAAATCAATAGTACTGAGTCTCTTTATGCATTTGCATCAGAGGCTGCTCG  
TGTTGTACGATCAATTTTCTCCCGCACTCTTGAAACTGCTCAAAATTTCTGTGCGTGTTTTACAGAAGGCC  
GCTATAACAATACTAGATGGAATTTACAGTATTCACTGAGACTCATTGATGCTATGATGTTACATCTG  
ATTTGGCTACTAACAATCTAGTTGTAATGGCCTACATTACAGGTGGTGTGTTTTCAGTTGACTTCGCAGTG  
GCTAACTAACATCTTTGGCACTGTTTATGAAAACTCAAACCCGTCCTTGATTGGCTTGAAGAGAAGTTT  
AAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTGTTAAATTTATCTCAACCTGTGCTTGTGAAA  
TTGTGCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGGAGAGTGTTTACAGACATTCTTTAAGCTTGT  
AAATAAATTTTGGCTTTGTGTGCTGACTCTATCATTATTGGTGGAGCTAACTTAAAGCCTTGAATTTA  
GGTGAACATTTTGTACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAAGAACTGGCC  
TACTCATGCCTCTAAAAGCCCCAAAAGAAATTATCTTCTTAGAGGGAGAAACACTTCCACAGAAGTGTT  
AACAGAGGAAGTTGTCTTGAAAACCTGGTGATTTACAACCATTAGAACAACCTACTAGTGAAGCTGTTGAA  
GCTCCATTGGTTGGTACACCAAGTTTGTATTAACGGGCTTATGTTGCTCGAAATCAAAGACACAGAAAAGT  
ACTGTGCCCTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAAAGGCGGTGCACCAACAAA  
GGTTACTTTTGGTGATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGGAATATCACTTTTGAACCTT  
GATGAAAGGATTGATAAAGTACTTAATGAGAAGTGCTCTGCCTATACAGTTGAACTCGGTACAGAAGTAA  
ATGAGTTTCGCTGTGTTGTGGCAGATGCTGTCTATAAAAACTTTGCAACCAGTATCTGAATTACTTACACC  
ACTGGGCATTGATTTAGATGAGTGGAGTATGGCTACATACTACTTATTTGATGAGTCTGGTGAGTTTAA  
TTGGCTTTCACATATGTATTGTTCTTTCTACCTCCAGATGAGGATGAAGAAGAAGGTGATTGTGAAGAAG  
AAGAGTTTGAAGCATCAACTCAATATGAGTATGGTACTGAAGATGATTACCAAGGTAAACCTTTGGAATT  
TGGTGCCACTTCTGCTGCTCTTCAACCTGAAGAAGAGCAAGAAGAAGATTGGTTAGATGATGATAGTCAA  
CAAAGTGTGGTCAACAAGACGGCAGTGAGGACAATCAGACAACCTACTATTCAAACAATTGTTGAGGTTT  
AACCTCAATTAGAGATGGAACCTTACACCAAGTTGTTTACAGACTATTGAAGTGAATAGTTTTAGTGGTTATTT  
AAAACCTTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAAAAAGGTAAACCAACA  
GTGGTTGTTAATGCAGCCAATGTTTACCTTAAACATGGAGGAGGTGTTGCAGGAGCCTTAAATAAGGCTA  
CTAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCACTTAAAGTGGGTGGTAG  
TTGTGTTTTAAGCGGACACAATCTTGCTAAACACTGTCTTCATGTTGTGCGCCCAAATGTTAACAAGGT  
GAAGACATTCAACTTCTTAAGAGTGCTTATGAAAATTTTAAATCAGCACGAAGTTCTACTTGCACCATAT  
TATCAGCTGGTATTTTTGGTGCTGACCCTATACATTCTTTAAGAGTTTGTGTAGATACTGTTTCGACAAA

TGTCTACTTAGTCTGCTTTTGATAAAAAATCTCTATGACAAACTTGTTCACAGCTTTTTTGGAAATGAAGAGT  
 GAAAAGCAAGTTGAACAAAAGATCGCTGAGATTCTTAAAGAGGAAGTTAAGCCATTTATAACTGAAAGTA  
 AACCTTCAGTTGAACAGAGAAAAACAAGATGATAAGAAAATCAAAGCTTGTGTTGAAGAAGTTACAACAAC  
 TCTGGAAGAAACTAAGTTCCTCACAGAAAACCTTGTTACTTTTATATTGACATTAATGGCAATCTTCATCCA  
 GATTCTGCCACTCTTGTTAGTGACATTGACATCACTTTCTTAAAGAAAGATGCTCCATATATAGTGGGTG  
 ATGTTGTTCAAGAGGGTGTTTTAACTGCTGTGGTTATACCTACTAAAAAGGCTGGTGGCACTACTGAAAT  
 GCTAGCGAAAGCTTTGAGAAAAAGTGCCAACAGACAATTATATAAACCCTTACCCGGGTGAGGGTTTTAAAT  
 GGTTACACTGTAGAGGAGGCAAAAGACAGTGCTTAAAAAGTGTAAGAGTGCCTTTTACATTCTACCATCTA  
 TTATCTCTAATGAGAAGCAAGAAATTCCTTGAACCTGTTTCTTGGAAATTTGCGAGAAATGCTTGCACATGC  
 AGAAGAAACACGCAAAATTAATGCCTGTCTGTGTGGAACTAAAGCCATAGTTTTCTTCAACTATACAGCGTAAA  
 TATAAGGGTATTTAAAAATAAGAGGGTGTGGTTGATTATGGTGCTAGATTTTTACTTTTTACACCGATAAA  
 CAACTGTAGCTCACTTATCAACACACTTAACGATCTAAATGAAACTCTTGTTACAATGCCACTTGGCTA  
 TGTAACACATGGCTTAAATTTTGAAGAAGCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTT  
 TCTGTTTTCTTCACCTGATGCTGTTACAGCGTATAATGGTTATCTTACTTCTTCTTCTTAAACACCTGAAG  
 AACATTTTTATTGAAACCATCTCACTTGCTGGTTTCTATAAAGATTGGTCTATTCTGGACAATCTACACA  
 ACTAGGTATAGAATTTCTTAAAGAGAGGTGATAAAAGTGATATTACACTAGTAATCTTACCACATTCCAC  
 CTAGATGGTGAAGTTATCACCTTTGACAATCTTAAAGACACTTCTTTCTTTGAGAGAAGTGAGGACTATTA  
 AGGTGTTTACAACAGTAGACAACATTAACCTCCACACGCAAGTTGTGGACATGTCAATGACATATGGACA  
 ACAGTTTGGTCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCATAATTCACATGAAGGT  
 AAAACATTTTTATGTTTTACCTAATGATGACACTCTACGTGTTGAGGCTTTTGAGTACTACCACACAACCTG  
 ATCCTAGTTTTCTGGGTAGGTACATGTCAGCATTAATACACTAAAAAGTGGAATACCCACAAGTTAA  
 TGGTTTTAACTTCTATTAATGGGCAGATAACAACCTGTTATCTTGCCACTGCATTGTTAACACTCCAACAA  
 ATAGAGTTGAAGTTTAATCCACCTGCTCTACAAGATGCTTATTACAGAGCAAGGGCTGGTGAAGCTGCTA  
 ACTTTTGTGCACTTATCTTAGCCTACTGTAATAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAACAAT  
 GAGTTACTTGTTTCAACATGCCAATTTAGATTCTTGCAGAAAGAGTCTTGAACGTGGTGTTGATAAACTTGT  
 GGACAACAGCAGACACACCTTAAAGGGTGTAAGAAGCTGTTATGTACATGGGCACACTTTCTTATGAACAAT  
 TTAAGAAAGGTGTTCAGATACCTTGTACGTGTGGTAAACAAGCTACAAAATATCTAGTACAACAGGAGTC  
 ACCTTTTGTATGATGTCAGCACCACTGCTCAGTATGAACCTTAAGCATGGTACATTTACTTGTGCTAGT  
 GAGTACACTGGTAATTACCAGTGTGGTCACTATAAACATATAACTTCTAAAGAACTTTGTATTGCATAG  
 ACGGTGCTTTACTTACAAAGTCTCAGAAATACAAAGGTCTATTACGGATGTTTTCTACAAAGAAAACAG  
 TTACACAACAACCATAAAACCAGTTACTTATAAATTGGATGGTGTGTTGTTGTACAGAAATTGACCCTAAG  
 TTGGACAATTATTATAAGAAAGACAATTCCTTATTTTACAGAGCAACCAATTGATCTTGTACCAAACCAAC  
 CATATCCAAACGCAAGCTTCGATAATTTTAAGTTTGTATGTGATAATATCAAATTTGCTGATGATTTAAA  
 CCAGTTAACTGGTTATAAGAAACCTGCTTCAAGAGAGCTTAAAGTTACATTTTTCCCTGACTTAAATGGT  
 GATGTGGTGGCTATTGATTATAAAACACTACACACCCTCTTTTTAAGAAAGGAGCTAAATTTGTTACATAAAC  
 CTATTGTTTGGCATGTTAAACAATGCAACTAATAAAGCCACGTATAAAACCAATACCTGGTGTATACGTTG  
 TCTTTGGAGCACAAAACCAGTTGAAACATCAAATTCGTTTGATGTACTGAAGTCAGAGGACGCGCAGGGA  
 ATGGATAATCTTGCCTGCGAAGATCTAAAACCAGTCTCTGAAGAAGTAGTGGAATCCTACCATACAGA  
 AAGACGTTCTTGAGTGTAATGTGAAAACCTACCGAAGTTGTAGGAGACATTATACCTTAAACCAGCAATAA  
 TAGTTTTAAAAATTACAGAAGAGTTTGCCACACATGCTAATGGCTGCTTATGTAGACAATTTCTAGTCTG  
 ACTATTAAAGAAACCTAATGAATTTCTAGAGTATTAGGTTTGAACCCCTTGCTACTCATGTTGTTAGCTG  
 CTGTTAATAGTGTCCCTTGGGATACTATAGCTAATTATGCTAAGCCTTTTTCTTAAACAAAGTTGTTAGTAC  
 AACTACTAACATAGTTACACGGTGTTTAAACCGTGTGTGTACTAATTATATGCCTTATTTCTTTACTTTA  
 TTGCTACAATTGTGTACTTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAG  
 CAAAGAATACTGTTAAGAGTGTGCGTAAATTTTGTCTAGAGGCTTCATTTAATTATTTGAAGTCACCTAA  
 TTTTTCTAACTGATAAATATTATAAATTTGGTTTTTACTATTAAGTGTTTGCCTAGGTTCTTTAATCTAC  
 TCAACCGCTGCTTTAGGTGTTTTAATGTCTAATTTAGGCATGCCTTCTTACTGTACTGGTTACAGAGAAG  
 GCTATTTGAACCTCTACTAATGTCATATTGCAACCTACTGTACTGGTCTATACCTTGTAGTGTTTGTCT  
 TAGTGGTTTTAGATTCTTTAGACACCTATCCTTCTTTAGAACTATACAAATTACCATTTTCATCTTTTAAA  
 TGGGATTTAACTGCTTTTGGCTTAGTTGCGAGAGTGGTTTTTGGCATATATTCTTTTCACTAGGTTTTTCT  
 ATGTACTTGGATTGGCTGCAATCATGCAATTGTTTTTTCAGCTATTTTGCAGTACATTTTATTAGTAATTC  
 TTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCGATTTTCAGCTATGGTTAGAATGTACATC  
 TTCTTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGACGGTTGTAATTCATCAACT  
 GTATGATGTGTTACAAACGTAATAGAGCAACAAGAGTCAAGTGTACAACATATTGTTAATGGTGTTAGAAG  
 GTCCTTTTATGTCTATGCTAATGAGGTAAGGCTTTTGCACAACTACACAATTTGGAATTTGTGTTAATTGT  
 GATACATTTGTGCTGCTAGTACTATTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTTTAAAA  
 GACCAATAAATCCTACTGACCAGTCTTCTTACATCGTTGATAGTGTTACAGTGAAGAATGGTTCATCCC

TCTTTACTTTGATAAAGCTGGTCAAAAGACTTATGAAAGACATTCTCTCTCTCATTGTTAACTTAGAC  
AACCTGAGAGCTAATAACACTAAAGGTTCAATTGCCTATTAATGTTATAGTTTTTGTATGGTAAATCAAAAT  
GTGAAGAATCATCTGCAAAATCAGCGTCTGTTTACTACAGTCAGCTTATGTGTCAACCTATACTGTTACT  
AGATCAGGCATTAGTGTCTGATGTTGGTGATAGTGCAGGAAAGTTGCAGTTAAAATGTTTGATGCTTACGTT  
AATACGTTTTTCATCAACTTTTAAACGTACCAATGGAAAACTCAAAACACTAGTTGCAACTGCAGAAGCTG  
AACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCTACTTTTTATTTTCAGCAGCTCGGCAAGGGTTTTGT  
TGATTGAGATGTAGAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCAATCTGACATAGAAGTT  
ACTGGCGATAGTTGTAATAACTATATGCTCACCTATAACAAAGTTGAAAACATGACACCCCGTGACCTTG  
GTGCTTGTATTGACTGTAGTGCGCGTCATATTAATGCGCAGGTAGCAAAAAGTCACAACATTGCTTTGAT  
ATGGAACGTTAAAGATTTTCATGTCAATTGTCTGAACAACTACGAAAACAAATACGTAGTGCTGTCAAAAAG  
AATAACTTACCTTTTTAAGTTGACATGTGCAACTACTAGACAAGTTGTTAATGTTGTAACAACAAGATAG  
CACTTAAGGGTGGTAAAATTGTTAATAATTGGTTGAAGCAGTTAATTAAGTTACACTTGTGTTCCTTTTT  
TGTTGCTGCTATTTTTCTATTTAATAACACCTGTTTCATGTCTGCTAAACATACTGACTTTTTCAAGTGAA  
ATCATAGGATACAAGGCTATTGATGGTGGTGTCACTCGTGACATAGCATCTACAGATACTTGTGTTTGCTA  
ACAAACATGCTGATTTTTGACACATGGTTTAGCCAGCGTGGTGGTAGTTATACTAATGACAAAGCTTGCCC  
ATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTTGTCGTGCCTGGTTTTGCCTGGCAGCATATTACGC  
ACAATAATGGTGACTTTTTGCATTTCTTACCTAGAGTTTTTAGTGAGTTGGTAACATCTGTTACACAC  
CATCAAACTTATAGAGTACACTGACTTTGCAACATCAGCTTGTGTTTTGGCTGCTGAATGTACAATTTT  
TAAAGATGCTTCTGGTAAGCCAGTACCATATTGTTATGATACCAATGTACTAGAAGGTTCTGTTGCTTAT  
GAAAGTTTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTATTCAATTTCCCTAACACCTACC  
TTGAAGGTTCTGTTAGAGTGGTAACAACCTTTTGATTCTGAGTACTGTAGGCACGGCACTTGTGAAAGATC  
AGAAGCTGGTGGTTGTGTATCTACTAGTGGTAGATGGGTACTTAACAATGATTATTACAGATCTTTACCA  
GGAGTTTTCTGTGGTGTAGATGCTGTAAATTTACTTACTAATATGTTTACACCACTAATTCAACCTATTG  
GTGCTTTGGACATATCAGCATCTATAGTAGCTGGTGGTATTGTAGCTATCGTAGTAACATGCCTTGCCCTA  
CTATTTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTACTATTC  
CTTATGTCATTCACTGTACTCTGTTTAAACACAGTTTACTCATTCTTACCTGGTGGTTTATTCTGTTATTT  
ACTTGTACTTGACATTTTATCTTACTAATGATGTTTCTTTTTTAGCACATATTCACTGGATGGTTATGTT  
CACACCTTTAGTACCTTTCTGGATAACAATTGCTTATATCATTGTTATTTCCACAAAGCATTCTATTGG  
TTCTTTAGTAATTAACCTTAAAGAGACGTGTAGTCTTTAATGGTGGTTTCTTTTAGTACTTTTGAAGAAGTG  
CGCTGTGCACCTTTTTGTTAAATAAAGAAATGTATCTAAAGTTGCGTAGTGATGTGCTATTACCTCTTAC  
GCAATATAATAGATACTTAGCTCTTTATAATAAGTACAAGTATTTTAGTGGAGCAATGGATACAACCTAGC  
TACAGAGAAGCTGCTTGTGTGTCATCTCGCAAAGGCTCTCAATGACTTCAGTAACCTCAGGTTCTGATGTTT  
TTTACCAACCACCACAAACCTCTATCACCTCAGCTGTTTTGCAGAGTGGTTTTAGAAAAATGGCATTCCC  
ATCTGGTAAAGTTGAGGGTTGTATGGTACAAGTAACCTGTGGTACAACCTACACTTAACGGTCTTTGGCTT  
GATGACGTAGTTTACTGTCCAAGACATGTGATCTGCACCTCTGAAGACATGCTTAACCTAATTATGAAG  
ATTTACTCATTTCGTAAGTCTAATCATAATTTCTTGGTACAGGCTGGTAATGTTCAACTCAGGGTTATTGG  
ACATTCTATGCAAAATTGTGTACTTAAGCTTAAGGTTGATACAGCCAATCCTAAGACACCTAAGTATAAG  
TTTGTTTCGATTCAACCAGGACAGACTTTTTTCACTGTTAGCTTGTGTTACAATGGTTTACCATCTGGTGGTT  
ACCAATGTGCTATGAGGCCCAATTTCACTATTAAGGGTTTCACTTCTTAATGGTTTCACTGTTGTTGG  
TTTTAACATAGATTATGACTGTGTCTCTTTTTGTTACATGCACCATATGGAATTACCAACTGGAGTTTCA  
GCTGGCACAGACTTAGAAGGTAACCTTTTATGGACCTTTTGTGACAGGCAAACAGCACAAAGCAGCTGGTA  
CGGACACAACCTATTACAGTTAATGTTTTAGCTTGGTTGTACGCTGCTGTTATAAATGGAGACAGGTGGTT  
TCTCAATCGATTTACCACAACCTCTAATGACTTTAACCTTGTGGCTATGAAGTACAATTATGAACCTCTA  
ACACAAGACCATGTTGACATACTAGGACCTCTTCTGCTCAAACCTGGAATTGCCGTTTTAGATATGTGTG  
CTTCATTAAAAGAATTACTGCAAAATGGTATGAATGGACGTACCATATTGGGTAGTGCTTTATTAGAAGA  
TGAATTTACACCTTTTGATGTTGTTAGACAATGCTCAGGTGTTACTTTCCAAAGTGCAGTGAAAAGAACA  
ATCAAGGGTACACACCCTGGTTGTTACTCACAATTTTGAACCTTCACTTTTAGTTTTAGTCCAGAGTACTC  
AATGGTCTTTGTTCTTTTTTTTTGTATGAAAATGCCTTTTTACCTTTTGCTATGGGTATTATTGCTATGTC  
TGCTTTTGCAATGATGTTTGTCAAACATAAGCATGCATTTCTCTGTTTGTGTTTTGTTACCTTCTCTTGCC  
ACTGTAGCTTATTTAATATGGTCTATATGCCTGCTAGTTGGGTGATGCGTATTATGACATGGTTGGATA  
TGTTTGATACTAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT  
AATCCTTATGACAGCAAGAAGTGTGTATGATGATGGTGCTAGGAGAGTGTGGACACTTATGAATGTCTTG  
ACACTCGTTTTATAAAGTTTATTATGGTAATGCTTTAGATCAAGCCATTTCCATGTGGGCTCTTATAATCT  
CTGTTACTTCTAACTACTCAGGTGTAGTTACAACGTGCATGTTTTTGGCCAGAGGTATTGTTTTTATGTG  
TGTTGAGTATTGCCCTATTTTCTTCATAACTGGTAATACACTTCAGTGTATAATGCTAGTTTTATTGTTTC  
TTAGGCTATTTTTGTACTTGTACTTTGGCCTCTTTTGTGTTACTCAACCGCTACTTTAGACTGACTCTTG  
GTGTTTATGATTACTTAGTTTCTACACAGGAGTTTAGATATATGAATTCACAGGGACTACTCCCACCCAA

GAATAGCATAGATGCCTTCAAACCTCAACATTAAATTGTTGGGTGTTGGTGGCAAACCTTGTATCAAAGTA  
GCCACTGTACAGTCTAAAATGTGAGATGTAAAGTGCACATCAGTAGTCTTACTCTCAGTTTTTGCAACAAC  
TCAGAGTAGAATCATCATCTAAATTGTGGGCTCAATGTGTCCAGTTACACAATGACATTCTCTTAGCTAA  
AGATACTACTGAAGCCTTTGAAAAATGGTTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGTGTGA  
GACATAAACAAGCTTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTTA  
GTTCCCTTCCATCATATGCAGCTTTTGTACTGCTCAAGAAGCTTATGAGCAGGCTGTTGCTAATGGTGA  
TTCTGAAGTTGTTCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGAATTTGACCGTGATGCA  
GCCATGCAACGTAAGTTGGAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTG  
AGGACAAGAGGGCAAAAGTTACTAGTGCTATGCAGACAATGCTTTTCACTATGCTTAGAAAAGTTGGATAA  
TGATGCACTCAACAACATTATCAACAATGCAAGAGATGGTTGTGTTCCCTTGAACATAATACCTCTTACA  
ACAGCAGCCAAACTAATGGTTGTGCATACCAGACTATAACACATATAAAAAATACGTGTGATGGTACAACAT  
TTACTTATGCATCAGCATTGTGGGAAATCCAACAGGTTGTAGATGCAGATAGTAAAATTGTTCAACTTAG  
TGAAATTAGTATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTAACAGCTTTAAGGGCCAATTCT  
GCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTTGTGCTGCCGGTACTA  
CACAACTGCTTGCCTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTAGGTTTGTACT  
TGCCTGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGATGGAAGTGGTACTATC  
TATACAGAAGTGAACACCTTGTAGGTTTGTACAGACACACCTAAAGGTCCTAAAGTGAAGTATTTAT  
ACTTTATTAAAGGATTAAACAACCTAAATAGAGGTATGGTACTTGGTAGTTTAGCTGCCACAGTACGTCT  
ACAAGCTGGTAATGCAACAGAAGTGCCTGCCAATTCACTGTATTATCTTTCTGTGCTTTTGTGTAGAT  
GCTGCTAAAGCTTACAAAGATTATCTAGCTAGTGGGGGACAACCAATCACTAATTGTGTTAAGATGTTGT  
GTACACACACTGGTACTGGTCAGGCAATAACAGTTACACCGGAAGCCAATATGGATCAAGAATCCTTTGG  
TGGTGCATCGTGTGTGTGTACTGCCGTTGCCACATAGATCATCAAATCCTAAAGGATTTTGTGACTTA  
AAAGGTAAGTATGTACAAATACCTACAACCTTGTGCTAATGACCCTGTGGGTTTTACTTTAAAAACACAG  
TCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAAGCCATGCTTCA  
GTCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCCTGTAGTGCAGCCCGTCTTACACCGTGCAGGCA  
CAGGCACTAGTACTGATGTGCTATACAGGGCTTTTGACATCTACAATGATAAAGTAGCTGGTTTTGCTAA  
ATTCCTAAAACTAATTGTTGTGCTTCCAAGAAAAGGACGAAGATGACAATTTAATTGATTCTTACTTT  
GTAGTTAAGAGACACACTTTCTCTAACTACCAACTGAAGAAACAATTTATAATTTACTTAAAGGATTGTC  
CAGCTGTTGCTAAACATGACTTCTTTAAGTTTGAATAGACGGTGACATGGTACCACATATATACGTC  
ACGTCTTACTAAATACACAATGGCAGACCTCGTCTATGCTTTAAGGCATTTTGTGATGAAGTAATTGTGAC  
ACATTTAAAGAAATACTTGTACATACAATTGTTGTGATGATGATTATTTCAATAAAAAGGACTGGTATG  
ATTTTGTAGAAAACCCAGATATATTACGCGTATACGCCAAGTTAGGTGAACGTGTACGCCAAGCTTTGTT  
AAAAACAGTACAATTCTGTGATGCCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGATAATCAA  
GATCTCAATGGTAAGTGGTATGATTTGCGGTGATTTTACATAACACCGCCAGGTAGTGGAGTTCCTGTTG  
TAGATTCTTATTATTATTGTTAATGCCTATATTAACCTTGACCAGGGCTTTAACTGCAGAGTCACATGT  
TGACACTGACTTAAACAAAGCCTTACATTAAGTGGGATTTGTTAAATATGACTTCACGGAAGAGAGGTTA  
AAACTCTTTGACCGTTATTTTAAATATTGGGATCAGACATACCACCCAAATTGTGTTAACTGTTTGGATG  
ACAGATGCATTCTGCATTGTGCAAACTTTAATGTTTTATTCTCTACAGTGTTCCACCTACAAGTTTTGG  
ACCACTAGTGAGAAAAATATTTGTTGATGGTGTTCATTTGTAGTTTCACTGGATACCACTTCAGAGAG  
CTAGGTGTTGTACATAATCAGGATGTAACTTACATAGCTCTAGACTTAGTTTTAAGGAATTACTTGTGT  
ATGCTGCTGACCCTGCTATGCACGCTGCTTCTGGTAATCTATTACTAGATAAACGCACTACGTGCTTTTC  
AGTAGCTGCACTTACTAACAATGTTGCTTTTCAAACCTGTCAAACCCGGTAATTTTAAACAAAGACTTCTAT  
GACTTTGCTGTGTCTAAGGGTTTCTTTAAGGAAGGAAGTTCTGTTGAATTTAAACACTTCTTCTTTGCTC  
AGGATGGTAATGCTGCTATCAGCGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAG  
ACAACACTACTATTTGTAGTTGAAGTTGTTGATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAATGCT  
AACCAAGTCATCGTCAACAACCTAGACAAATCAGCTGGTTTTCCATTTAATAAATGGGGTAAGGCTAGAC  
TTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACCTTTTCGCATATACAAAACGTAATGTCTATCCC  
TACTATAACTCAAATGAATCTTAAAGTATGCCATTAGTGCAAGAATAGAGCTCGCACCGTAGCTGGTGTCT  
TCTATCTGTAGTACTATGACCAATAGACAGTTTCATCAAAAATTATTGAAATCAATAGCCGCCACTAGAG  
GAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTTGGCACAACATGTTAAAACTGTTTATAG  
TGATGTAGAAAACCTCACCTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAACATGCTT  
AGAATTATGGCCTCACTTGTCTTGTCTCGCAAACATACAACGTGTTGTAGCTTGTACACCGTTTCTATA  
GATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTATATGTTAAACC  
AGGTGGAACCTCATCAGGAGATGCCACAACCTGCTTATGCTAATAGTGTTTTTAACATTTGTCAAGCTGTC  
ACGGCCAATGTTAATGCACCTTTTATCTACTGATGGTAACAAAATTGCCGATAAGTATGTCCGCAATTTAC  
AACACAGACTTTTATGAGTGTCTCTATAGAAATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGC  
ATATTTGCGTAACATTTCTCAATGATGATACTCTCTGACGATGCTGTTGTGTGTTTCAATAGCACTTAT

GCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAGTCAGTTCCTTTATTATCAAAACAATGTTTTTA  
TGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAATTTTGCTCTCAACATAC  
AATGCTAGTTAAACAGGGTGATGATTATGTGTACCTTCCTTACCCAGATCCATCAAGAATCCTAGGGGCC  
GGCTGTTTTGTAGATGATATCGTAAAAACAGATGGTACACTTATGATTGAACGGTTCGTGTCTTTAGCTA  
TAGATGCTTACCCACTTACTAAACATCCTAATCAGGAGTATGCTGATGTCTTTCATTTGTACTTACAATA  
CATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGTTATGCTTACTAATGAT  
AACACTTCAAGGTATTGGGAACCTGAGTTTTATGAGGCTATGTACACACCGCATAACAGTCTTACAGGCTG  
TTGGGGCTTGTGTTCTTTGCAATTCACAGACTTCATTAAGATGTGGTGCTTGCATACGTAGACCATTCTT  
ATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAAATTAGTCTTGTCTGTTAATCCGTAT  
GTTTGCAATGCTCCAGGTTGTGATGTCACAGATGTGACTCAACTTTACTTAGGAGGTATGAGCTATTATT  
GTAAATCACATAAACCCATTAGTTTTCCATTGTGTGCTAATGGACAAGTTTTTGGTTTTATATAAAAA  
TACATGTGTTGGTAGCGATAATGTTACTGACTTTAATGCAATTGCAACATGTGACTGGACAAATGCTGGT  
GATTACATTTTAGCTAACACCTGTACTGAAAGACTCAAGCTTTTTGCAGCAGAAACGCTCAAAGCTACTG  
AGGAGACATTTAACTGTCTTATGGTATTGCTACTGTACGTGAAGTGCTGTCTGACAGAGAATTACATCT  
TTCATGGGAAGTTGGTAAACCTAGACCACCCTTAACCGAAATTATGTCTTTACTGGTTATCGTGTAACCT  
AAAAACAGTAAAGTACAAATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGTGATGCTGTTGTTTACC  
GAGGTACAACAACCTTACAAATTAAATGTTGGTGATTATTTTGTGCTGACATCACATACAGTAATGCCATT  
AAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATACCCAACACTCAATATC  
TCAGATGAGTTTTCTAGCAATGTTGCAAATTATCAAAAGGTTGGTATGCAAAGTATTCTACACTCCAGG  
GACCACCTGGTACTGGTAAGAGTCATTTTGCTATTGGCCTAGCTCTCTACTACCCTTCTGCTCGCATAGT  
GTATACAGCTTGCTCTCATGCCGCTGTTGATGCACTATGTGAGAAGGCATTAAAATATTTGCCTATAGAT  
AAATGTAGTAGAATTATACCTGCACGTGCTCGTGTAGAGTGTTTTGATAAATTCAAAGTGAATTC AACAT  
TAGAACAGTATGTCTTTTGTACTGTAAATGCATTGCCTGAGACGACAGCAGATATAGTTGTCTTTGATGA  
AATTTCAATGGCCACAAATTATGATTTGAGTGTTGTCAATGCCAGATTACGTGCTAAGCACTATGTGTAC  
ATTGGCGACCCTGCTCAATTACCTGCACCACGCACATTGCTAACTAAGGGCACACTAGAACCAGAATATT  
TCAATTCAGTGTGTAGACTTATGAAAACCTATAGGTCCAGACATGTTCCCTCGGAACCTTGTCGGCGTTGTCC  
TGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTTATGATAATAAGCTTAAAGCACATAAAGACAAATCA  
GCTCAATGCTTTAAATGTTTTATAAGGGTGTTATCACGCATGATGTTTCATCTGCAATTAACAGGCCAC  
AAATAGGCGTGGTAAGAAATTCCCTTACACGTAACCCCTGCTTGGAGAAAAGCTGTCTTTATTTACCTTTA  
TAATTCACAGAATGCTGTAGCCTCAAAGATTTTGGGACTACCAACTCAAAGTGTGATTATCACAGGGC  
TCAGAATATGACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTTGTAATGTAAACAGATTTA  
ATGTTGCTATTACCAGAGCAAAAGTAGGCATACTTTGCATAATGTCTGATAGAGACCTTTATGACAAGTT  
GCAATTTACAAGTCTTGAAATTCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATGTAACAGGACTC  
TTTAAAGATTGTAGTAAGGTAATCACTGGGTACATCCTACACAGGCACCTACACACCTCAGTGTTGACA  
CTAAATTCAAACTGAAGGTTTTATGTGTTGACATACCTGGCATACCTAAGGACATGACCTATAGAAGACT  
CATCTCTATGATGGGTTTTTAAATGAATTATCAAGTTAATGGTTACCCTAACATGTTTATCACCCGCGAA  
GAAGCTATAAGACATGTACGTGCATGGATTGGCTTCGATGTGAGGGGTGTCATGCTACTAGAGAAGCTG  
TTGGTACCAATTTACCTTTACAGCTAGGTTTTTCTACAGGTGTTAACCTAGTTGCTGTACCTACAGGTTA  
TGTTGATACACCTAATAATACAGATTTTTCCAGAGTTAGTGCTAAACCACCGCCTGGAGATCAATTTAAA  
CACCTCATACCCTTATGTACAAAGGACTTCCTTGGAATGTAGTGCGTATAAAGATTGTACAAATGTTAA  
GTGACACACTTAAAAATCTCTCTGACAGAGTCGTATTTGTCTTATGGGCACATGGCTTTGAGTTGACATC  
TATGAAGTATTTTGTGAAAATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTT  
TCCACTGCTTCAGACACTTATGCCTGTTGGCATCATTCTATTGGATTTGATTACGTCTATAATCCGTTTA  
TGATTGATGTTCAACAATGGGGTTTTACAGGTAACCTACAAAGCAACCATGATCTGTATTGTCAAGTCCA  
TGGTAATGCACATGTAGCTAGTTGTGATGCAATCATGACTAGGTGTCTAGCTGTCCACGAGTGCTTTGTT  
AAGCGTGTTGACTGGACTATTGAATATCCTATAATTGGTGATGAACTGAAGATTAATGCGGCTTG TAGAA  
AGGTTCAACACATGGTTGTTAAAGCTGCATTATTAGCAGACAAATTCAGTTCTTTCACGACATTGGTAA  
CCCTAAAGCTATTAAGTGTGTACCTCAAGCTGATGTAGAATGGAAGTTCTATGATGCACAGCCTTG TAGT  
GACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGACAAATTCACAGATGGTG  
TATGCCTATTTTGGAAATTGCAATGTGATAGATATCCTGCTAATTCCATTGTTTGTAGATTTGACACTAG  
AGTGCTATCTAACCTTAACCTTGCTGTTGTGATGGTGGCAGTTTGATGTAAATAAACATGCATTCCAC  
ACACCAGCTTTTGATAAAAGTGCTTTTGTTAATTTAAACAATTACCATTTTTCTATTACTCTGACAGTC  
CATGTGAGTCTCATGGAACAAGTAGTGTCAGATATAGATTATGTACCACTAAAGTCTGCTACGTGTAT  
AACACGTTGCAATTTAGGTGGTGCTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTCGATGCT  
TATAACATGATGATCTCAGCTGGCTTTAGCTTGTGGGTTTTACAAACAATTTGATACTTATAACCTCTGGA  
ACACTTTTACAAGACTTCAGAGTTTAGAAAATGTGGCTTTTAAATGTTGTAAATAAGGGACACTTTGATGG  
ACAACAGGGTGAAGTACCAGTTTCTATCATTAATAACACTGTTTACACAAAAGTTGATGGTGTTGATGTA



GAATTGTTTGAAAATAAAACAACATTACCTGTTAATGTAGCATTGAGCTTTGGGCTAAGCGCAACATTA  
AACCAGTACCAGAGGTGAAAATACTCAATAATTTGGGTGTGGACATTGCTGCTAATACTGTGATCTGGGA  
CTACAAAAGAGATGCTCCAGCACATATATCTACTATTGGTGTGTTCTATGACTGACATAGCCAAGAAA  
CCAACCTGAAACGATTTGTGCACCACCTCACTGTCTTTTTTGGATGGTAGAGTTGATGGTCAAGTAGACTTAT  
TTAGAAATGCCCCGTAATGGTGTCTTATTACAGAAGGTAGTGTTAAAGGTTTACAACCATCTGTAGGTCC  
CAAACAAGCTAGTCTTAATGGAGTCACATTAATTGGAGAAGCCGTAAAAACACAGTTCAATTATTATAAG  
AAAGTTGATGGTGTGTTCCAACAATTACCTGAAACTTACTTTACTCAGAGTAGAAATTTACAAGAATTTA  
AACCCAGGAGTCAAATGGAAATTGATTTCTTAGAATTAGCTATGGATGAATTCATTGAACGGTATAAATT  
AGAAGGCTATGCCTTCGAACATATCGTTTATGGAGATTTTAGTCATAGTCAGTTAGGTGGTTTACATCTA  
CTGATTGGACTAGCTAAACGTTTTAAGGAATCACCTTTTGAATTAGAAGATTTTATTCCCTATGGACAGTA  
CAGTTAAAAACTATTTTCATAACAGATGCGCAACAGGTTTATCTAAGTGTGTGTGTTCTGTTATTGATTT  
ATTACTTGATGATTTTTGTTGAAATAATAAAATCCCAAGATTTATCTGTAGTTTCTAAGGTTGTCAAAGTG  
ACTATTGACTATACAGAAATTTTCAATTTATGCTTTGGTGTAAAGATGGCCATGTAGAAACATTTTACCCAA  
AATTACAATCTAGTCAAGCGTGGCAACCGGGTGTGCTATGCCTAATCTTTACAAAATGCAAAGAATGCT  
ATTAGAAAAGTGTGACCTTCAAATTATGGTGATAGTGCAACATTACCTAAAGGCATAATGATGAATGTC  
GCAAAATATACTCAACTGTGTCAATATTTAAACACATTAACATTAGCTGTACCCTATAATATGAGAGTTA  
TACATTTTGGTGTGTTCTGATAAAGGAGTTGCACCAGGTACAGCTGTTTTAAGACAGTGGTTGCCTAC  
GGGTACGCTGCTTGTGATTGATCTTAATGACTTTGTCTCTGATGCAGATTCAACTTTGATTGGTGAT  
TGTGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGATATGTACGACCCTAAGACTAAAA  
ATGTTACAAAAGAAAATGACTCTAAAGAGGGTTTTTCACTTACATTTGTGGGTTTATACAACAAAAGCT  
AGCTCTTGGAGGTTCCGTGGCTATAAAGATAACAGAACATTCTTGAATGCTGATCTTTATAAGCTCATG  
GGACACTTCGCATGGTGGACAGCCTTTGTTACTAATGTGAATGCGTCATCATCTGAAGCATTTTTAATTG  
GATGTAATTATCTTGGCAAACACGCGAACAATAGATGGTTATGTCATGCATGCAAATTACATATTTTG  
GAGGAATACAAATCCAATTCAGTTGTCTTCCCTATTCTTTATTTGACATGAGTAAATTTCCCTTAAATTA  
AGGGGTACTGCTGTTATGTCTTTAAAGAAGGTCAAATCAATGATATGATTTTATCTCTTCTTAGTAAAG  
GTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATGTTCTTGTAAACAACTAAACGAA  
CAATGTTTGTGTTTTCTTGTGTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCA  
ATTACCCCTGCATACACTAATTCTTTCACACGTGGTGTGTTATTACCCTGACAAAGTTTTTCCAGATCCTCA  
GTTTACATCTCAACTCAGGACTTGTCTTACCTTTCTTTTCCAATGTTACTTGGTTCATGCTATACATG  
TCTCTGGGACCAATGGTACTAAGAGGTTTGATAAACCTGTCCCTACCATTTAATGATGGTGTGTTATTTTGC  
TTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTTGGTACTACTTTAGATTGGAAGACCCAGTCC  
CTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGAATGATCCAT  
TTTTGGGTGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTATTCTAGTGC  
GAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTTC  
AAAAATCTTAGGGAATTTGTGTTTAAAGAATATTGATGGTTATTTTAAATATATTCTAAGCACACGCCTA  
TTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTTCGGCTTTAGAACCATTTGGTAGATTTGCCAATAGGTAT  
TAACATCACTAGGTTTCAAACCTTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCA  
GGTTGGACAGCTGGTGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTAATAATA  
ATGAAAATGGAACCATTAACAGATGCTGTAGACTGTGCACCTTGACCCTCTCTCAGAAACAAAGTGACGTT  
GAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACCTTCTAAGTTTGAAGTCCAAACCAACAGAATCTATT  
GTTAGATTTTCTAATATTACAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCATCTG  
TTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCTATATAATTCCGCATC  
ATTTTCCACTTTTAAAGTGTTATGGAGTGTCTCCTACTAAATTAATGATCTCTGCTTTACTAATGTCTAT  
GCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACCTGGAAAGATTGCTG  
ATTATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAAATCTAACAATCTTGATTCT  
TAAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAAGAG  
GATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATGGTGTGGAAGGTTTTAATTGTTACT  
TTCTTTTACAATCATATGGTTTTCAACCCACTAATGGTGTGTTGTTACCAACCATACAGAGTAGTAGTACT  
TTCTTTTGAACCTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTGGTTAAAAAC  
AAATGTGTCAATTTCAACTTCAATGGTTTAAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTC  
TGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGA  
GATTCTTGACATTACACCATGTTCTTTTGGTGGTGTGAGTGTATAACACCAGGAACAAATACTTCTAAC  
CAGGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCTATGCAGATCAACTTA  
CTCCTACTTGGCGTGTTTATTCTACAGGTTCTAATGTTTTTCAAACACGTGCAGGCTGTTTAAATAGGGGC  
TGAACATGTCAACAACCTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACT  
CAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCCTACACTATGTCACCTTG  
GTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATACCCACAAATTTTACTATTAGTGTTAC

CACAGAAATTCTACCAAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCA  
ACTGAATGCAGCAATCTTTTGTGCAATATGGCAGTTTTTGTACACAATTAACCGTGCTTTAACTGGAA  
TAGCTGTTGAACAAGACAAAAACCCCAAGAAGTTTTTGCACAAGTCAAACAAATTTACAAAACACCACC  
AATTAAAGATTTTGGTGGTTTTAATTTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCA  
TTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAAACAATATGGTGATT  
GCCTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAAACGGCCTTACTGTTTTGCCACC  
TTTGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGG  
ACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAAATGGTATTG  
GAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAA  
AATTCAAGACTCACTTTCTTCCACAGCAAGTGCACTTGGAAAACCTCAAGATGTGGTCAACCAAAATGCA  
CAAGCTTTAAACACGCTTGTGTTAAACAACCTTAGCTCCAATTTTGGTGCAATTTCAAGTGTTTTAAATGATA  
TCCTTTACCGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTTGATAGGTTGATCACAGGCAGACTTCAAAG  
TTTGAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAGCTTCTGCTAATCTTGCTGCT  
ACTAAAATGTGAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTA  
TGTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCACTGTGACTTATGTCCCTGCACAAGAAAA  
GAACTTCACAACCTGCTCCTGCCATTTGTCTATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTT  
TCAAATGGCACACACTGGTTTGTAAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACA  
CATTTGTGTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACC  
TGAATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTTAAAGATCATACATCACCAGATGTTGATTTA  
GGTGACATCTCTGGCATTAAATGCTTCAGTTGTAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTG  
CCAAGAATTTAAATGAATCTCTCATCGATCTCCAAGAACCTTGAAAGTATGAGCAGTATATAAAATGGCC  
ATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATAGTAATGGTGACAATTATGCTTTGCTGT  
ATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAAATTTGATGAAGACG  
ACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAACGAACCTTATGGATTTGTTTATGAGA  
ATCTTCACAATTGGAACCTGTAACCTTTGAAGCAAGGTGAAATCAAGGATGCTACTCCTTCAGATTTTGTTC  
GCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTGGCGTTGCACTTCT  
TGCTGTTTTTTCAGAGCGCTTCCAAAATCATAACCCTCAAAAAGAGATGGCAACTAGCACTCTCCAAGGGT  
GTTCACTTTGTTTGAACCTTGCTGTTGTTGTTGTTGTTGTAACAGTTTACTCACACCTTTTGCTCGTTGCTGCTG  
GCCTTGAAGCCCCCTTTCTCTATCTTTATGCTTTTAGTCTACTTCTTGCAGAGTATAAACTTTGTAAGAAT  
AATAATGAGGCTTTGGCTTTGCTGGAAATGCCGTTCCAAAACCCATTACTTTATGATGCCAATGATTTTT  
CTTTGCTGGCATACTAATTGTTACGACTATTGTATACCTTACAATAGTGTAACCTTCTTCAATTGTCATTA  
CTTCAGGTGATGGCACAACAAGTCCTATTTCTGAACATGACTACCAGATTGGTGGTTATACTGAAAAATG  
GGAATCTGGAGTAAAGACTGTGTTGTATTACACAGTTACTTCACTTCAGACTATTACCAGCTGTACTCA  
ACTCAATTGAGTACAGACACTGGTGTGTAACATGTTACCTTCTTCATCTACAATAAAATTGTTGATGAGC  
CTGAAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTTGTTAATCCAGTAATGGAACCAAT  
TTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAAGCTGATGAGTACGAACCTTATGTAC  
TCATTGTTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCTGTTGAT  
TCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATTGTGTGCGTACTGCTGCAATATTGTTAACGT  
GAGTCTTGTAACCTTCTTTTTACGTTTACTCTCGTGTTAAAAATCTGAATTCTTCTAGAGTTCCTGAT  
CTTCTGGTCTAAACGAACATAATATTATATTAGTTTTTCTGTTTGGAACCTTAAATTTTAGCCATGGCAGA  
TTCCAACGGTACTATTACCGTTGAAGAGCTTAAAAAGCTCCTTGAACAATGGAACCTAGTAATAGGTTTC  
CTATTCCTTACATGGATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGGTTTTTGTATATAATTA  
AGTTAATTTTCTCTGGCTGTTATGGCCAGTAACCTTTAGCTTGTTTTGTGCTTGCTGCTGTTTACAGAAT  
AAATTGGATCACCGGTGGAATTGCTATCGCAATGGCTTGTCTTGTAGGCTTGATGTGGCTCAGCTACTTC  
ATTGCTTCTTTTCACTGTTTGCAGCTGTTTGCAGCTACGCGTTCATGTGGTCATTCAATCCAGAACTAACATTCTTC  
TCAACGTGCCACTCCATGGCACTATTCTGACCAGACCGCTTCTAGAAAGTGAACCTCGTAATCGGAGCTGT  
GATCCTTTCGTGGACATCTTCGTATTGCTGGACACCATCTAGGACGCTGTGACATCAAGGACCTGCCTAAA  
GAAATCACTGTTGCTACATACGAACGCTTTCTTATTACAAATTTGGGAGCTTCGACGCGTGTAGCAGGTG  
ACTCAGGTTTTGCTGCATACAGTCGTACAGGATTGGCAACTATAAATTAACACAGACCATTCCAGTAG  
CAGTGACAATATTGCTTTGCTTGTACAGTAAGTGACAACAGATGTTTCATCTCGTTGACTTTTCAAGTTAC  
TATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGTTTCCATTTGGAATCTTGATTACATCATA  
AACCTCATAATTAATAAATTTATCTAAGTCACTAAGTGAATAAATATTCTCAATTAGATGAAGAGCAAC  
CAATGGAGATTGATTAAACGAACATGAAAATTATTCTTTTCTTGGCACTGATAACACTCGCTACTTGTGA  
GCTTTTATCACTACCAAGAGTGTGTTAGAGGTACAACAGTACTTTTAAAAGAACCTTGCTCTTCTGGAACA  
TACGAGGGCAATTCACCATTTTATCCTCTAGCTGATAACAAATTTGCACTGACTTGCTTTAGCACTCAAT  
TTGCTTTTGTCTTGTCTGACGGCGTAAAACACGTCTATCAGTTACGTGCCAGATCAGTTTACCTAACT  
GTTTCATCAGACAAGAGGAAGTTCAAGAACCTTTACTCTCCAATTTTTCTTATTGTTGCGGCAATAGTGTTT

ATAACACTTTGCTTCACACTCAAAAGAAAGACAGAATGATTGAACTTTCATTAATTGACTTCTATTTGTG  
CTTTTTAGCCTTTCTGCTATTCCTTGTTTTAATTATGCTTATTATCTTTTGTTTCTCACTTGAAGTGCAA  
GATCATAATGAACTTGTCACGCCTAAACGAACATGAAATTTCTTGTTTTCTTAGGAATCATCACAAGTG  
TAGCTGCATTTACCAAGAATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCC  
GTGTCCTATTCACCTTCTATTCTAAATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAA  
TTGTGCGTGGATGAGGCTGGTTCTAAATCACCCATTTCAGTACATCGATATCGGTAATTATACAGTTTCCT  
GTTTACCTTTTACAATTAATTGCCAGGAACCTAAATTGGGTAGTCTTGTAGTGCGTTGTTTCGTTCTATGA  
AGACTTTTTAGAGTATCATGACGTTTCGTGTTGTTTTAGATTTTCATCTAAACGAACAAACTAAAATGTCTG  
ATAATGGACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCTCAGATTCAACTGGCAG  
TAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCCAAGGTTTACCCAATAATACT  
GCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCTCGAGGACAAGGCGTTC  
CAATTAACACCAATAGCAGTCCAGATGACCAAAATTGGCTACTACCGAAGAGCTACCAGACGAATTTCGTGG  
TGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCT  
GGACTTCCCTATGGTGCTAACAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCAA  
AAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACCTTCTCAAGGAACAAC  
ATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTTCCTCATCACGT  
AGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCA  
ATGGCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGG  
TAAAGGCCAACAACAACAAGGCCAAACTGTCTACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGG  
CAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAGAACAACCC  
AAGGAAATTTTGGGGACCAGGAATAATCAGACAAGGAAGTATTACAAACATTGGCCGCAAATTGCACA  
ATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTGCGCATTGGCATGGAAGTCACACCTTCGGGAACG  
TGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGC  
TGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAGAAGGC  
TGATGAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAAGTGTGACTCTTCTTCCTGCTGCAGAT  
TTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACTCAGGCCTAAACTCATG  
CAGACCACACAAGGCAGATGGGCTATATAAACGTTTTTCGCTTTTCCGTTTACGATATATAGTCTACTCTT  
GTGCAGAATGAATTCTCGTAACTACATAGCACAAAGTAGATGTAGTTAACTTTAATCTCACATAGCAATCT  
TTAATCAGTGTGTAACATTAGGGAGGACTTGAAAGAGCCACCACATTTTACCAGAGGCCACGCGGAGTAC  
GATCGAGTGACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAAT  
TTTAGTAGTGCTATCCCATGTGATTTTAATAGCTTCTTAGGAGAATGACAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAA

## Appendix B

>fragment1

TTTTAAGTGTTATGGAGTGTCTCCTACTAAATTA

>fragment2

TCTACCAGTGTCTATGACCAAGACATCAGTAGATTG

>fragment3

TTTACAAGACTTCAGAGTTTAGATAATGTGGCTTTT

>fragment4

TACCAATTTACCTTTACAGCTAGTTTTTTCTACAGG

>fragment5

CCTTACCGCAGAGACAGAAGAAACAGCAAACGTGA