**Zadanie1.**

Odpowiada sekwencji zupełnie losowej – każdy aminokwas występuje z tą samą częstością.

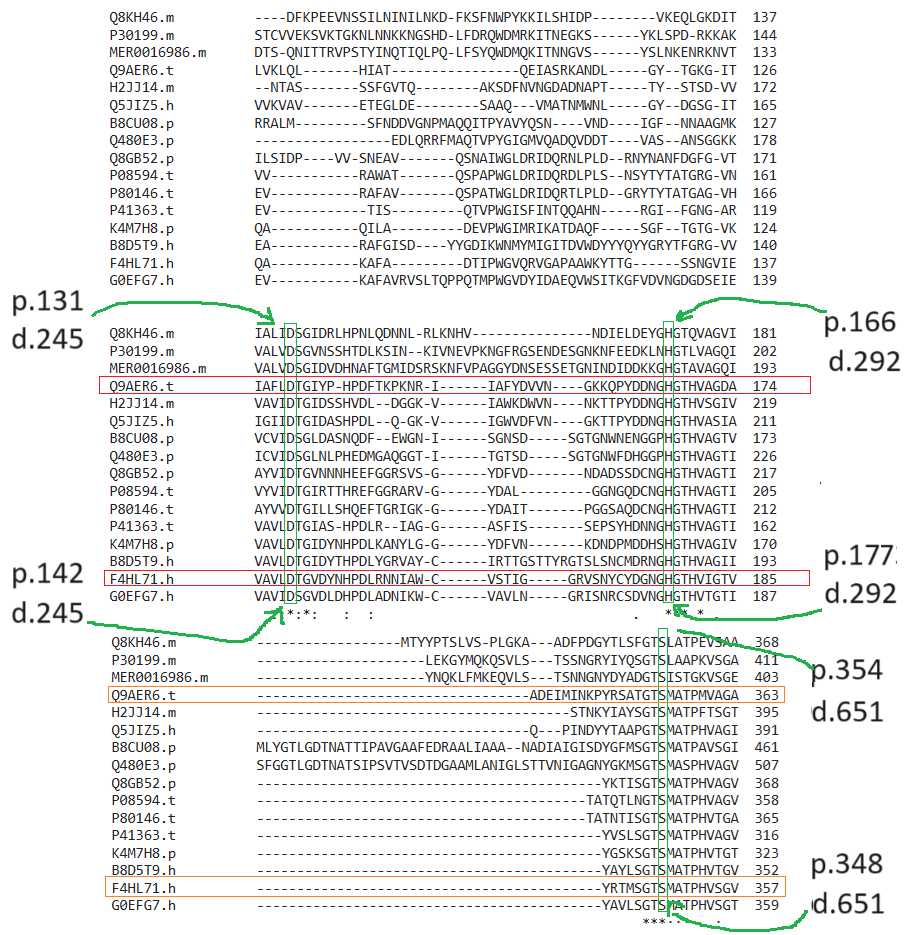
Odpowiada sekwencji złożonej z jednego aminokwasu – niska złożoność.

**Zadanie2.**

* Wg UNIPROT Q9AER ma reszty z centrum katalitycznego na pozycjach:

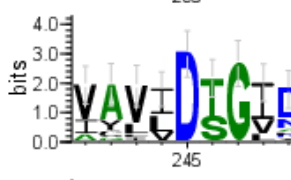
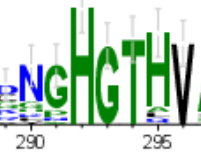


* Zlokalizuzjmy te reszty na dopasowaniu. Zobaczmy gdzie analogiczne reszty ma F4HL71.

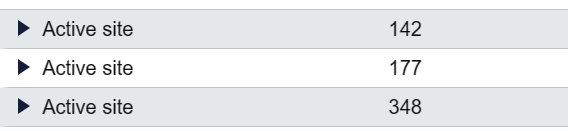


Obserwujemy wysoką konserwatywność na tych pozycjach.

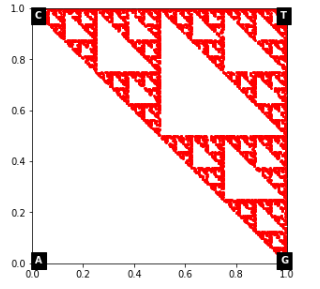
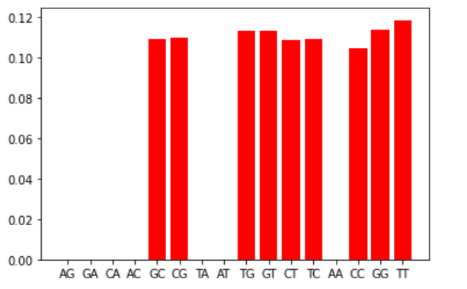
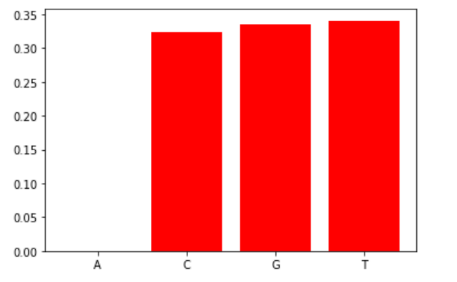
* Jak to wygląda na Logo?

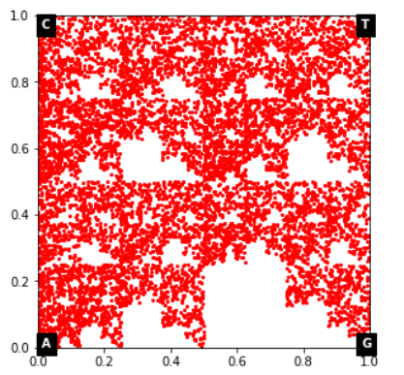
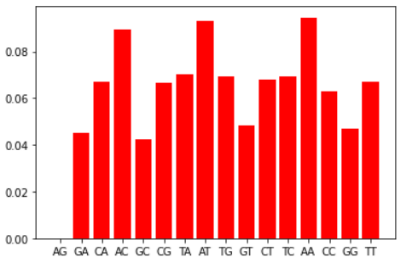
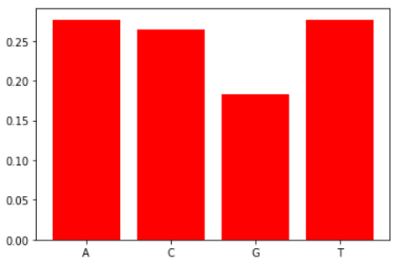
* Zobaczmy czy to jesto zgodne z F4HL71 wg UNIPROT.



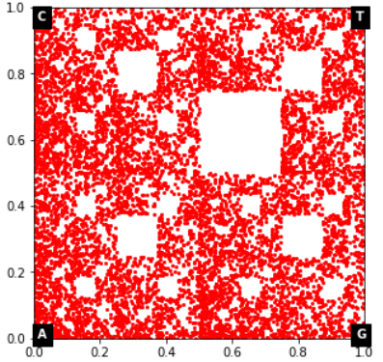
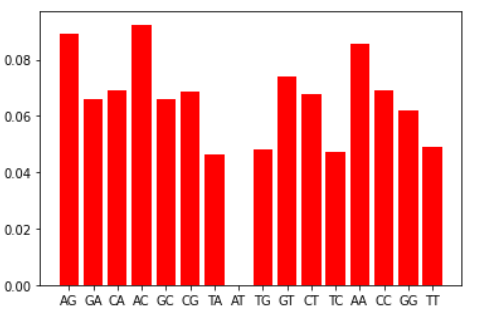
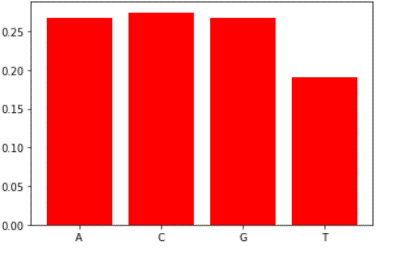
**Zadanie3.** Generujemy sekwencję o długości 10k w której nie ma A.



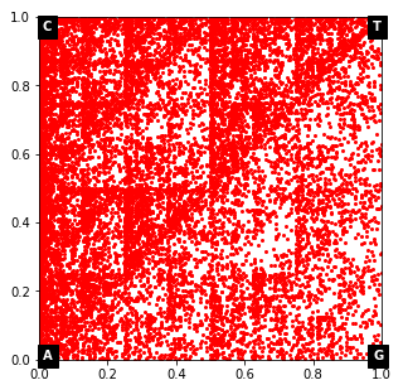
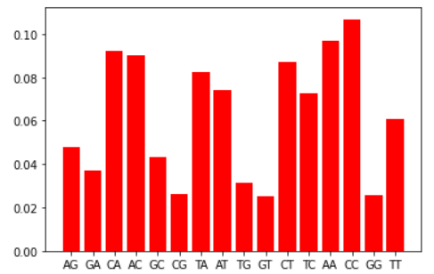
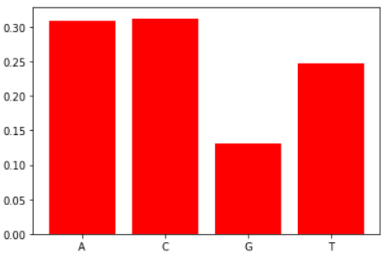
**Zadanie4.** Generujemy sekwencję o długości 10k w której nie ma AG.



**Zadanie5.** Generujemy sekwencję o długości 10k w której nie ma AT.



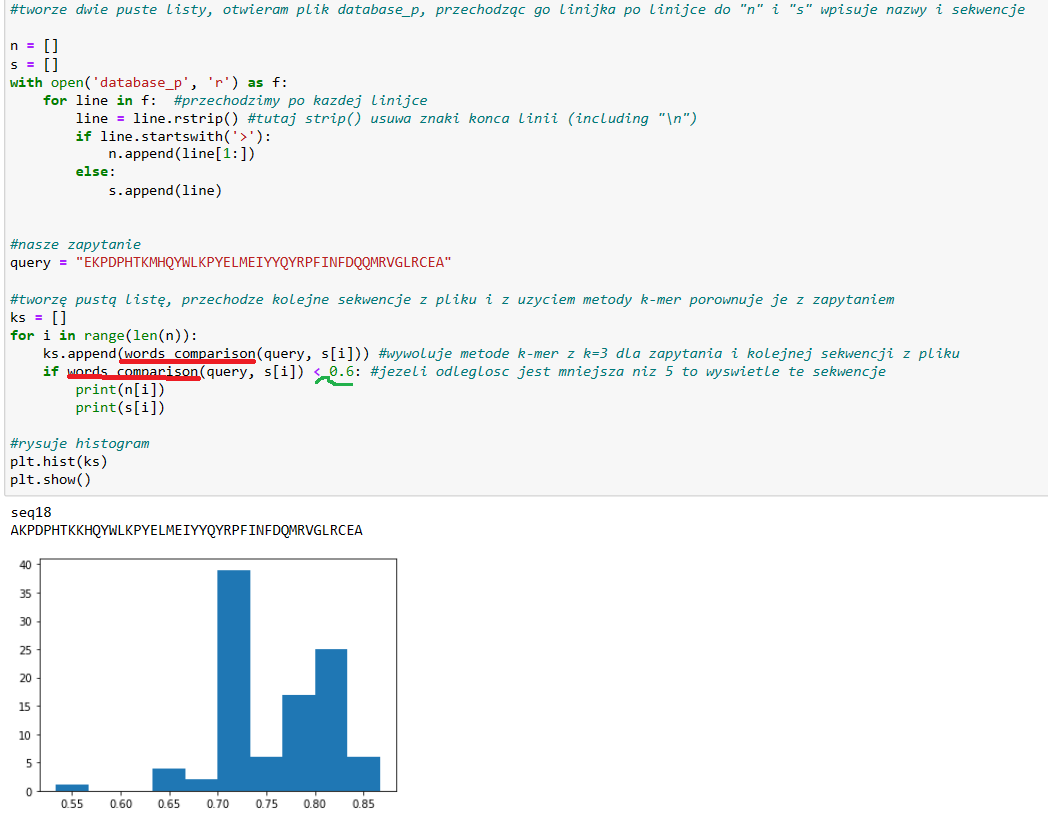
**Zadanie6.** Przeanalizujmy genom mitochondrialny. Skomentuj skład.



- obserwujemy znacznie mniej nukleotydu G

- CC wystepouje najczęściej, CG/GT najrzadziej - > widać też to na CGR

**Zadanie7.**  Porównywanie sekwencji zapytania z sekwencjami z bazy danych.



**Zadanie8.** Porównanie dwóch sekwencji

>s1

CCCGCAATAAGTCGCCCGCGCACACCAGCAGCTCCGCGCAAGGGGCCTTAATCAATTGCTACCCCTACTTGGGCTACCTCCTTGCTGGAACTTGAAGGGA

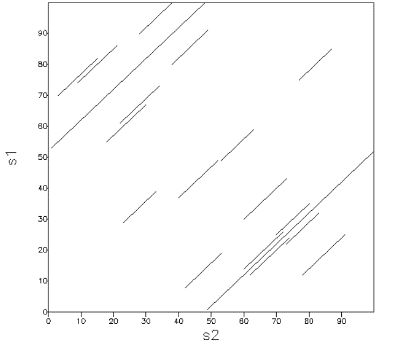
>s2

CAATTGCTACCCCTACTTGGGCTACCTCCTTGCTGGAACTTGAAGGGACCCGCAATAAGTCGCCCGCGCACACCAGCAGCTCCGCGCAAGGGGCCTTAAT

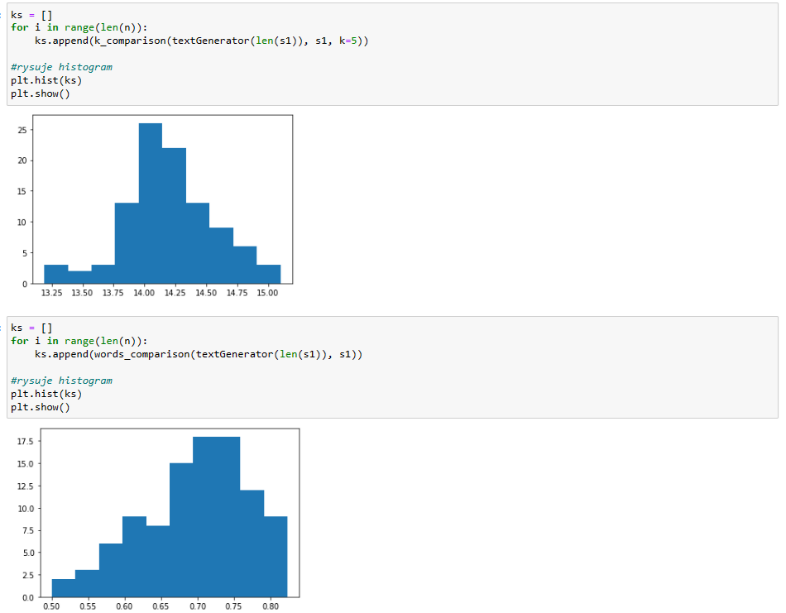
* Wykonujemy dopasowanie (Clustal)



* Rysujemy macierz kropkową



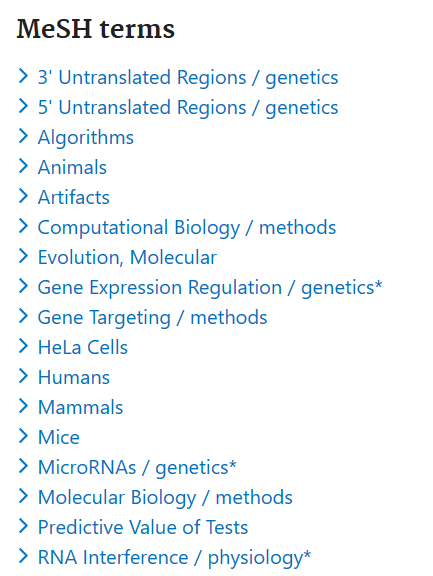
* Porównanie metodą k-mer (k=5): 2.83
* Porównanie metodą Lempel-Ziv: 0.59



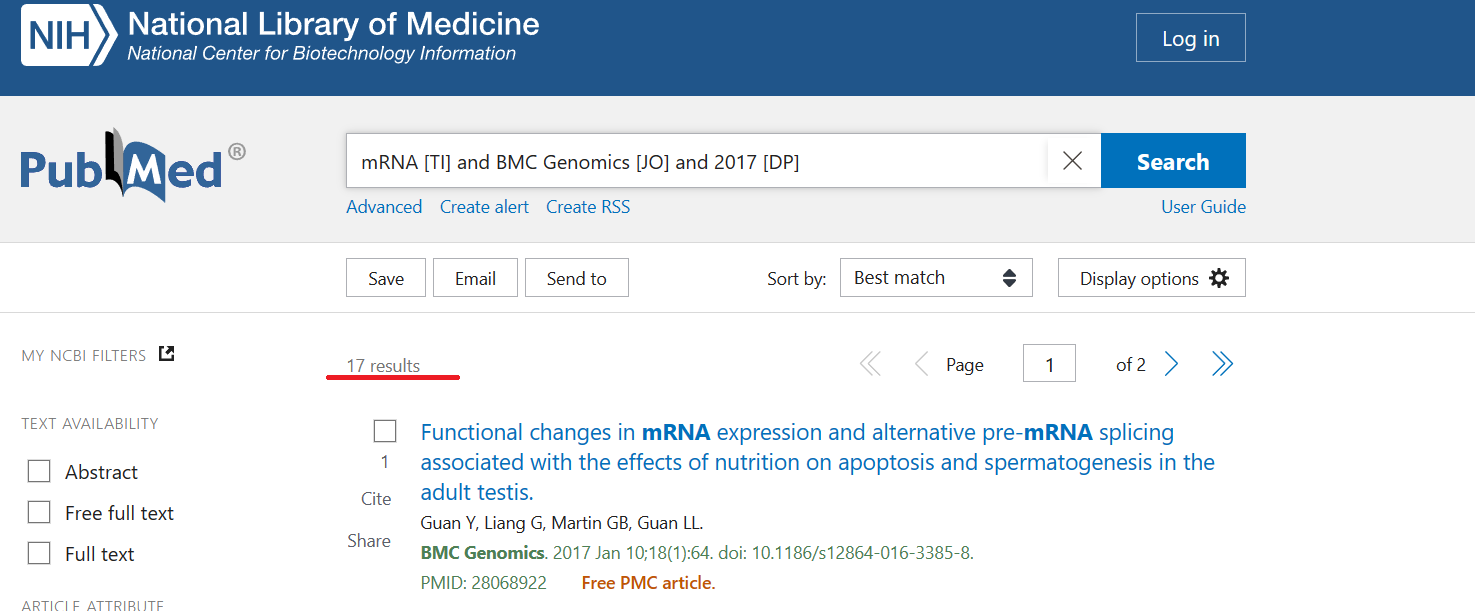
Sekewencje są do siebie podobne. Patrząc na dopasowanie/macierz kropkową widzimy, że początek pierwszej sekwencji jest taki sam jak koniec drugiej sekwencji.

**Zadanie9.**

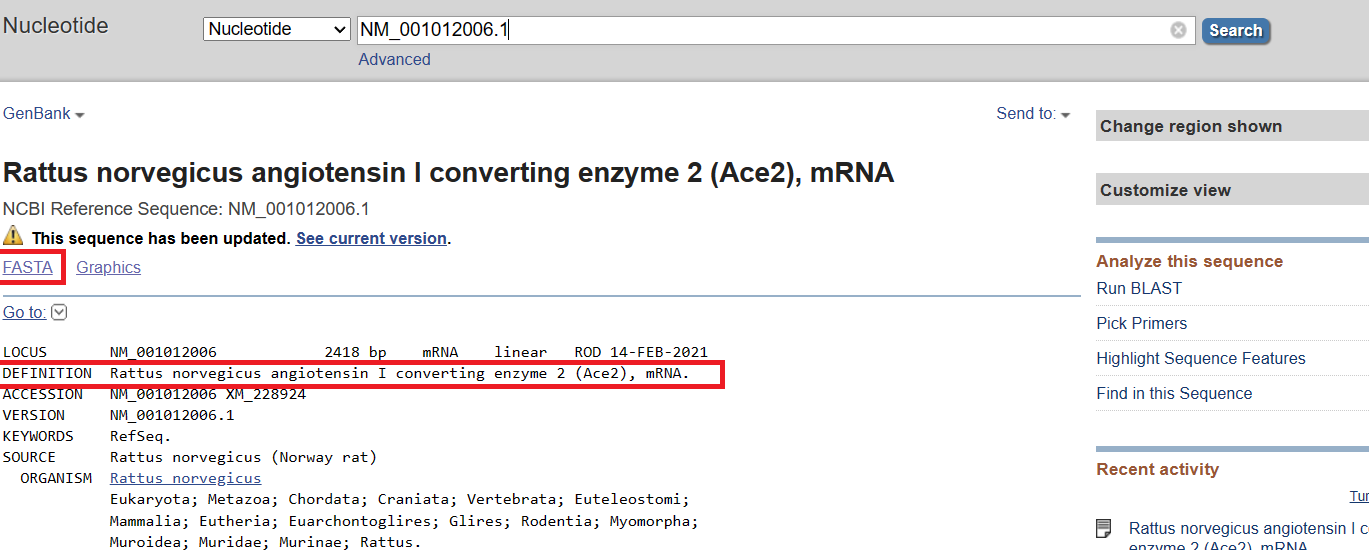
****

****

**Zadanie10.**

****

**Zadanie11.**

****

****

>NM\_001012006.1 Rattus norvegicus angiotensin I converting enzyme 2 (Ace2), mRNA

ATGTCAAGCTCCTGCTGGCTCCTTCTCAGCCTTGTTGCTGTTGCTACTGCTCAGTCCCTCATCGAGGAAA

AGGCCGAGAGCTTTTTAAACAAGTTTAACCAGGAAGCTGAAGACCTGTCTTATCAAAGTTCACTTGCTTC

TTGGAATTACAACACCAACATTACGGAGGAGAATGCCCAAAAGATGAACGAGGCTGCGGCCAAATGGTCT

GCCTTTTATGAAGAACAGTCCAAGATCGCCCAAAATTTCTCACTACAAGAAATTCAGAATGCGACCATCA

AGCGTCAACTGAAGGCCCTTCAGCAGAGCGGGTCTTCAGCGCTGTCACCAGACAAGAACAAACAGTTGAA

CACAATTCTAAACACCATGAGCACCATTTACAGTACTGGAAAAGTTTGCAACTCAATGAATCCACAAGAA

TGTTTTTTACTTGAACCAGGATTGGACGAAATAATGGCAACAAGCACAGACTACAATCGTAGGCTCTGGG

CTTGGGAGGGCTGGAGGGCTGAGGTCGGCAAGCAGCTGAGGCCGTTATATGAAGAGTATGTGGTCCTGAA

AAATGAGATGGCAAGAGCAAACAATTATGAAGACTATGGGGATTATTGGCGAGGGGATTATGAAGCAGAG

GGAGTAGAAGGTTACAACTACAACCGAAACCAGTTGATCGAAGACGTAGAAAATACCTTCAAAGAGATCA

AACCGTTGTATGAGCAACTTCATGCCTATGTGAGAACGAAGTTGATGGAAGTGTACCCTTCTTACATCAG

CCCTACTGGATGCCTCCCTGCTCATTTGCTTGGTGATATGTGGGGTAGGTTTTGGACAAATCTGTACCCT

TTGACTACTCCCTTTCTTCAGAAACCAAACATAGATGTTACTGATGCAATGGTGAATCAGAGCTGGGATG

CAGAAAGAATATTTAAAGAGGCAGAGAAGTTCTTCGTTTCTGTTGGCCTTCCTCAAATGACTCCGGGATT

CTGGACAAACTCCATGCTGACTGAGCCAGGAGATGACCGGAAAGTTGTCTGCCACCCCACAGCTTGGGAT

CTGGGACATGGAGACTTCAGAATCAAGATGTGCACAAAGGTCACAATGGACAACTTCTTGACAGCCCATC

ATGAGATGGGACACATCCAATATGACATGGCATATGCCAAGCAACCTTTCCTGCTAAGAAACGGAGCCAA

TGAAGGGTTCCATGAAGCCGTTGGAGAAATCATGTCACTTTCTGCAGCTACCCCCAAACATTTGAAATCT

ATTGGTCTTCTGCCATCCAATTTTCAAGAAGACAATGAAACAGAAATAAACTTCCTACTCAAACAGGCAT

TGACAATTGTTGGAACGCTGCCATTTACTTACATGTTAGAGAAGTGGAGGTGGATGGTCTTTCAGGATAA

AATTCCCAGAGAACAGTGGACCAAAAAGTGGTGGGAGATGAAGCGGGAGATCGTTGGTGTGGTGGAGCCT

CTGCCTCATGATGAAACATACTGTGACCCTGCATCTCTGTTCCATGTCTCTAATGATTACTCATTCATTC

GATATTACACAAGGACCATTTATCAATTCCAGTTTCAAGAAGCTCTTTGTCAAGCAGCTAAACATGATGG

CCCACTACACAAATGTGACATCTCAAATTCCACTGAAGCTGGGCAGAAGTTGCTCAATATGCTGAGTCTT

GGAAACTCAGGGCCCTGGACCCTAGCCTTGGAAAATGTGGTAGGATCAAGGAATATGGATGTAAAACCAC

TGCTCAATTACTTCCAACCATTGTTTGTCTGGCTGAAAGAGCAGAACAGGAATTCGACTGTGGGGTGGAG

CACTGACTGGAGCCCATATGCCGACCAAAGCATTAAAGTGAGGATAAGCCTAAAATCAGCTCTTGGGAAA

AATGCGTATGAATGGACCGACAACGAAATGTACCTATTCCGATCATCTGTTGCCTATGCCATGAGAGAGT

ATTTTTCAAGGGAAAAGAACCAGACAGTTCCTTTTGGGGAGGCAGACGTATGGGTGAGTGATTTGAAACC

AAGAGTCTCCTTCAACTTCTTTGTCACTTCACCCAAAAATGTGTCTGACATCATTCCCAGAAGTGAAGTT

GAAGAGGCCATCAGGATGTCTCGGGGCCGTATCAATGATATTTTTGGTCTGAATGATAACAGCCTGGAGT

TTCTGGGGATCTACCCAACACTTAAGCCACCTTACGAGCCTCCTGTCACCATATGGCTGATTATTTTTGG

TGTCGTGATGGGAACGGTAGTGGTTGGCATTGTTATCCTGATCGTCACTGGGATCAAAGGTCGAAAGAAG

AAAAATGAAACAAAAAGAGAAGAGAATCCTTATGACTCCATGGACATTGGCAAAGGAGAAAGTAACGCAG

GATTCCAAAACAGTGATGATGCTCAAACTTCATTCTAA