

BIOINFORMATIKA V PRAXI

CVIČENÍ 3 – DRUHÁ ČÁST

IDENTIFIKACE GENŮ, PROTEINŮ A JEJICH FUNKCE

STUDIJNÍ MATERIÁLY

Studijní materiály předmětu C2130 Úvod do chemoinformatiky a bioinformatiky, přednáška **Predikce genu, Sequence-evolution-function: Computational Approaches in Comparative Genomics.**

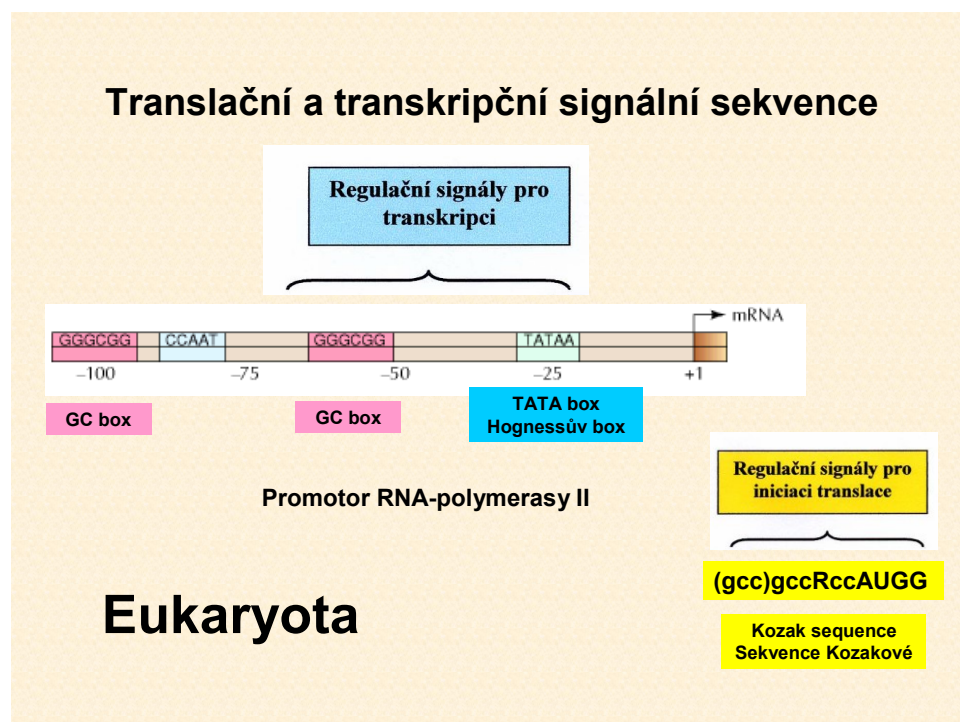
ÚKOL 1

Charakterizujte část genomu *neznámého organismu* pomocí predikčního programu **GeneMark** (<http://exon.gatech.edu/GeneMark>). Využijte *heuristický* predikční program.

```
ATCTGGCCGGTTTTATTACGCGCTGTTACTCGCGGCAGCCGAGCTTGCCCGCGCGTGTATGAAAGATGTTATTGCCGAACCATACCGCGCGG
TTTACTGCCGGGCTTTAGCCAGGCGCGGCAGGCGGTGTCGGAGATCGGCGCGCTGGCGAGCGGGATTTCCGGGTCGGGGCCGACGCTGTTTGGC
CTATGCGATAAACCGGAGACGGCGCAGCGCTGCGGACTGGCTGAGCAAACATTATCTGCAAAATCAGGAAGGCTTCGTTTCATATTTGCCGGC
TGGATACGGCGGGCGCACGAGTAGTGGGATAATCAATGAACTCTATAATCTGAAAGACCATAATGAGCAGGTCAGCTTTGCCGAGGCGGTCAC
GCAAGGACTGGGCAACAGCAGGACTTTTTTTCCGACGATCTGCCGGAGTTAGCCTGACGGAAATGATGAGATGCTCAACCAGGACTTT
GTCAGCCGTAGCGAAAGATCCTCTCGGCATTTATTGGCGATTGAAATACCGCAGCAAATCTGGAAGAGCGCGTCCGCGCGCGTTCGGCTTC
CGGCACCGGTAGCGCAGGTAGAAAGCGATGTCGGCTGCCTGGAGCTGTTCATGGTCCGACGCTGGCCTTTAAAGACTTCGGCGGGCGTTTTAT
GGCGCAAATGCTGACGCATATCAGCGCGCAGAAACCGGTGACGATTTCTGACTGCAACGTCAGGCGGATACCGGCGCGCGGTGGCTCACGCGTTC
TATGGCCTGGAATGTCGGGTCGTCATTCTCTACCGCGCGGTAAGATCAGTCCGTTGACGAAAAACTGTTCTGTACGCTGGGCGGCAACA
TTGAAACCGTGGCGATCGACGGCGATTTCGACGCGTGCCAGGCGCTGGTGAAACAGGCAATTTGATGACGAAGAAGCTGAAACCGGCGCTGGGGCT
GAATTCGGCTAATTCGATTAATATCAGCGCGCTGTGGCGCAAATTTGCTACTACTTTGAAGCCGTGGCGCAACTGCCGAGGGGGCGGTAAC
CAACTGGTGATCTCCGTCAGCGGCAACTTTGGCGATTTGACGGCAGGGCTGCTGGCGAAGTCGTTAGGCCTACCGGTGAAACGTTTTATCG
CCGCCAACGTCACGACAGGTCGCGCTTTCTGATGACGGAAGTGGGCGCGAAAGCGACGCGAGGCACTGTCGAATGCGATGGA
TGTGAGCCGCGGAAATAACTGCGCGCGTGGAGGAGCTATTCGCGGCTTAACTGCGCGCTGACTGAGCTGAGCGCGGTGGATGAT
ACTACGACACAGCAGACGATGCGCGAGCTGAAAGCGAAAGGTTATATCTCGGAACCTCATGCGGCGGTAGCGTATCGGGCATTACGCGACCAGT
TAAACCTGGCGAGTATGGCTTGTTCGGAACGGCGCATCCGGCGAAGTTTAAAGAGAGCGTGGAGTCCATTTGGGAGAAACGCTGCCTTC
GCCTGAAGCGCTCGCGAACGCGCCGACCTGCCGCTGCTTTCACATCATCTGCCTGCGGATTTTGGCGCCTGCGTAAGCTGATGATGACCCGC
CAATAACCATTCGCGCGGTGGCGCTGTCGCTTACCGGCGCTATGGGGTGGTTCGATTTGATGAGCCGGATAAGCGGTAACCGCATCCGGCGA
TGCCGTTACTGCTCGTAGCGCTTAAAGACCAGCTCGTCTGTGTGGAGTTTCTTCATCAAAGAAATACCCTTCACGGTCAAACGCGGTAAGCT
GTTCCGGCTTCGTTAAGCGGTTTTCAATAATAAACCGGCTCATCAGTCGCGCGCTTTTTTGGCGTAGAAGCTTACCACCTTAAACTTGCCGTT
TTTCTCATCAAGGAACACGGGCTTAATCAGTTCGGCATTGAGTTTCTCGGCTTACCGATTTAAATATTTCTCGGAGGCCAGATTACCACC
ACCGGATCGCCCTCGCTGAGCTTTCGTTGAGCTTATCGGTAATGATATCCGCCAGAAATGATAAAGATCTTTGCCGCGCGGATTTCCCA
GGCGAATCCCATCTCCAGACGATAAAGGCTGCATTAATCCAGCGGGCGCAATACGCCATAACAAGCCAGAGAGCATACGCAGATGTTGTTGAGC
AAAATCAAATCCGCGCTGTTGAACTTTCCGCTGTAGGCGCGTATAAACAATCGCCTTGAACGCCAGAATCGCCTGGCGTGCATATCCGGC
GTAATGAGGCTGCCAGTATGAAACCGCGTGGCGTTGAGATCCGCCAGTTTGTGCTAATTTCCCATCAGCTGGAATTTGCGGCGCGGAA
GCTGGCGCGGCTGTTGAATAAGCTGCTGCTGATCCAAACAGCTCCGCTGGTATAGCGGGTTCGTTGGCCAGCGGGCTTTGATAATCAAGCGT
TTTTGACGGTGAATCAGAATCAGCATAATCAGTCTTGCAGGAATTTTCTGCGACTTTAGCAAAAAACGCGCCAGAGTTGACCGATGGTTG
CGATTGTCGGCTTAATCGCGCGATGCCGCTCCAGGATCTGGCGCAGTTGTGTTGATATCCGGGAAGCGCGCGGATCGAACTGCGGTC
TTACGCCAGTTTCCGTTGTGCGAGATAATCGCTGGCGAGGTTATACACCACCGGTGAGAGCAACAAAATCGCCGTGAGATTGGTAATGGCCAT
ACAGGCCATGATCATGCTCAGCGAGCTGCCATATCAGCGGAAACTGATAAGCGTACCAGCGGATAACCATGCCAAGCGTCCGAAAGCGTAATAGC
CAGATAGCCTTTGCGTTATGTAACCGCAGAAAAACAGATTGTTTTCGGCGTAAATATAGTTGGCGACGATAGAAGTGAACGCGAACAGAAATGA
CGATAAGCGCGACAAAATGGCGCCCATTCACCGGTCAACGAAACCATGTCATGTTGGAGAAGCTGAATGCCCTTCTGTTGACGAGTGGGACGC
GTGATTTCCCGCCAGCAGGATAATCATCGCGCTGGCGGTACAGATGATAATGGTGTGCTGAATACGCCAATCATTTGCAACAATCCCTGCGCG
ACAGGGTGAGGGGATACGAGCTGCAAGCTGCGGCGCGCATTAGCGTGTGACCCCATTCGCGCTTCAATTAGAGAACATCCACCGTGAACACCGC
TGGTAATAGCTGGGTGAGCGTATATCCGGCTGCGCTGCCGCGGCTTCTGCCAGCCAAATGCGCTTTTACTGACTGAGGGGATAACGCCAGG
CATTTGCTCAATATGCCAGAGGCAAATGAATACGCTGCCGCGGACCCACAATAACCGCATGAGGGGAATCAGCCATTGATCAGACGGCGGACG
CCTTTGATGCGCGAATGATAATTAACAGGGCAGAACGCGCAGAGCAATGCCGGAGATAAGCGGCGGAATGTTGAAGGCGAAATGGAGCGCGC
GTGAGAGGAACTCGCCTGACAGCTATTAATAAATCAACCCGTAGGCGACGAGCAGAGAAAGAGGGCGAAAAACACCGCCATCCAGGCAATCCAG
CCCACGCGCCATATACCAGCGCGGCGCGCGGAACTGGCCTGTGCGGTACGTTTCTTTATAAAGCTGGGCAAGCGAACACTCGGCGAAGGAG
GTGCCATGCCAATGATGGCCGAGACCCACATCCAGAATACCGCGCGGGACCGCTGCGGCGATAGCCAGCGCCACGCGCGCCAGGTTACCGC
TGCCAATCCGCGCGCGAGGCTGGTACACAGAGCTGAAATCAGCTCAGGCGCGCTGGCTGCGGGCTAAGGCTGCCTTTCAGACTGCGCGTAAA
TTGGCAATATAACGAAACTGAATGAATCCGATCAGGTAACCGAAGTAAACCAACATCTGCCGCGAGCAGGTAATCATTTACCGGCGGAGGAT
ATTTGCTTAATAAACTGAAAACTCAGGCATTAACGTCCTCTTGTGATGCCGGCACGCTTTGATAATCCTGTATAAGCGTGACCCATGATG
TAGATGACCTTGTGACAGTAATATAACGGCAGTTTACCATAAATACGGTGGTATCCTTTAATTGCGCATCAACCGTCGGCAGATACGCAACA
GTGCACAAGGGCAGCCAGGTGATGTAGGGCGTTGCGCTGTGAGTGCCTGCTGTTATCATCAGGGTAGACCGGTTACATCCCTAAACAAGCTGT
TTAAAGAGAACTCTATCATGACGGACAAATGACCTCCTTGTGAGTTACCACCGTAGTGGCTGATACCGGAGATATCGCGCAATGAAAC
TGTATCAGCCGAGGATGCTACAATAACCTTCTCTCATTCTTAACGACGCGCAAATCCCGAATATCGTAAGCTGATTGACGATGCTGTGCGC
CTGGGCGAAACAGCAGAGCAGCGACCGCGCGCAGCAGTTGTTGACGCGACCGATAAGCTGGCGGTGAATATTGGCCTGGAGATCCTGAAGCTG
GTGCCGGGCGTATTTCTACCGAAGTTGACGCGCTGTCTTATGACACTGAAGCTCTATCGCCAAAGCAAACGATCATTTAACTCTACA
```


AATTGAAGTTATATAATGAGTTTTTGAATAAATATCTTACATGGAAGAGAATATTGATGTGTTTTTTTTTGGTTTTTATATAGGTGTTTGACAG
TTATAAGTAAGACCAATATCATACTGGAATATATAAATTACAAAAAATAATTAATTTACAAGCGCGTATAAAGTAGTTTGTAGATTTTGAAC
TGATTTTTTTTTTCAAATATAAACAAACCACATCTTTCTTTTGGACTTCTATTACCAATGTCCAGGATAGAGATATATCGTCGTTGATAATA
AATCAAACATTTAAGATCTATAAATCCAATTCATCACTATCGTCATCTATAAACTCACAAAAACAATTGGAGAAATTTGGAATAACCTTAT
GTAAATACCTTGTAACCAAGATATCAACAACTTTACAATAGAGAAGATGATTTACTAGTTCAATATGAATATATATCGTGATATTGTACCTGAACT
CACTATTTATAACAATTCAATTGAAGAGCATTACAGTATAAAATCCAATATCTAATAATCAAAATGTGCCAATAATCGATTGATATCTCGTCAA
GTGTTCAATTTGTTAGTAAATGATGATGATTTCACTTTGGTATGAAATGAATTACCAATCACCTAGCATAAATGAATATATTTTGAATAT
TAGAAGATATTAATGATTATCCATTGATTATAAAACCCTTGTATTACCTGATTATCTAATACCCAACCTGGTGGATCAATTGGGTAATAAGTT
GGATTCCTTGGTGATTACAATTAGTATATTCATCTCTACATCTAATAGGTTGAGAAATTTATAAATGATAATCCGAAAAACGATTACCA
AAATTGCTAGTGATGGGAAATATATCAAGATATTGTTAAATTCATGTAATAATAACAAAAATAAAATTTGAAAAATTAAGAATAGAAAAAT
TTATTAATAAATTAATTAATATGCTAGTGATGGGAAATTTAAATTAATAAATTAATGAATAATCAATTAATTTGGTTTTTGGATTGAATCTATAAA
ATCATCTTTACAATAAGTAGTAAGTGTACGTGGTCAATAAAGTAATGTATATATGTGCGTGTACTGAGTGATGATTTTCTATTTTATCTCGCGC
GTGTTATTTTTGTGTTATTTGTTGTTGTGCTGAGACTTAATTAATAACTTACGTGTTCACCTACATTTATTGCTTATGAGCCTAGGCGTT
ATTTTTTTTTTCTCTCTCTCCCACTTTCTACTCTTTGAACAATGTGGTGAIAAAAAAAAAAAAAATAGAACCAGATCTTTTTTATTCCATATAT
TTCAAAACCGACTTGTCTTTTTTTCTTTTACAACCAATTAATAACAAAAGAAAGACAACAATCATAATGTCTACTTCTGTGAAACCAATGAA
ACAGAAGCTTGTGAGAAAGCAGAATGATCTTTCCACAACCTGCCTCAATTGAAGAAAAATATCTCCACAACAAGGAGAGGCTGCAGAAGACG
ATGACGACACTCTAAAAGAACCAATATGATGAAGCTAAAAGAACCCTGAATCTTTAAAACAAGTTGAATCGATATTAGCACCTATTGTTTT
CACTGCATTGTCAATTTTCTGTGAGATTTTATCGTATTTCACTGAATGACCATGTTGTTTGGGATGAAGCTCAATTTGGTAAATTTGGATCCTAT
TATTTACGACACGAATTTTATCATGATGTTCACTCCCATTTGGTAAATGTTAGTTGTTTATCTGTTATTTGGCTGGATACAAATGGATCCTT
GGGATTTCCCAAGTGGTGAIAAATACCCTGATTATATGATTACTAAAATGAGATTGTTTAAATGCCACTTTCTCTGCCTGTGTGTACCATT
GGCTTATTTCACTGGGAAAGAAGTTGGATTTTCCATGTTTACTTACTTGGTATTTACTTTGATGGTGGCTCTTGATCAAGTTATGTCACTTTA
GGTAAATTCATTTTGTGGATTCATGTTGTTATTTCTCACCGTGTCTACTGTTTCTGTTTTTCACGTTTCAACAATTTTAAACAATAAATCAC
AAGAATTTCTAGAAAAATGGTGAIAAATGGATTCTTTTAACTGGTGTTCATTTGTTGTTGTTTCTGTAIAAATGGTTGGATTGTTTTGTTAC
CACATTTGGTGGGATTTACACTGTTGTGACCTTTGGAATAAATGAGTGATAAATCTATTTTCATGGACAAAAATACATTCACAAATTTGGTTGCT
AGAATTGTTGCTTTGATTTCTTGCCCAATTTTCAATTTTTCATGCTTTCAATTAAGTTCAATTTTGGATTTGTTGTATAAATCGGGTACTGGTGATG
CCAATATGTCACTACTTTTCCAAGCTAATTTGGCTGGTCCGATGTTGGTGGTGGCCCCACGTGAAGTATCCATGTTCCACTCGGTTATCACTTT
AAAGAATCAAGGTTTAAAGTGGTGGCCTTTTACACTCCACGTTTCAAACATTTCCAGAAAGTTCAAACAACAACAAGTTACTACTTATGGTCCAC
AAAGATTCAAACAACAATTTGGATTTTCCAAAGAGCTAGAGGACAACCTTATATGATACTTCTGGTAACACCACCTGACATTTGAATATATTTTTG
ACGGTATGCATGTAAGATTGATGCATCCACAACTGGTAGAACTTACATACTCATGATATCCAGCTCCAGTGTCTAAATCTGAATATGAAGT
TGCATGTTATGGTAAATTTGACTATTGGTGCATCTAAAGATAATTTGGACTGTTGAAATTTGGAACAAGCAAGTATGAAGATAAATGAGATTA
CATCTTTGACTTCGTCTATTAGATTGAAGAATGAAGTGAATGTTATTGGGGGTCCTGGTACTACATTACCTCAATGGGGTTCAGAC
AAGGTGAAGTTGTTGTTACAAGAACCATTTAAAAAAGACAAGAGAACTTG

PREDIKCE GENŮ U MNOHOBUNĚČNÝCH EUKARYOT



Mnohobuněčná eukaryota se vyznačují komplexní organizací genomu, geny jsou separovány dlouhými intergenovými úseky, geny obsahují mnoho intronů, i velmi dlouhých. Exony/introny jsou identifikovány pomocí míst sestřihu (GT na 5'konci intronu, AG na 3'konci). Vzniká velké množství chyb! Dlouhé introny jsou určeny jako intergenové úseky, krátké intergenové úseky jako introny, krátké exony nemusí být identifikovány.

ÚKOL 4

Analyzujte část genu mnohobuněčného eukaryotického organismu (*Arabidopsis thaliana*) pomocí programu **GeneMark** (<http://exon.gatech.edu/GeneMark>), verze pro eukaryota.

TATTTTATGAAAAACATTAATAACAAAATAAATTTAGTATTAGTACAATTAATATGTAAAATTAATAAATAATATATCATTAATCTATAAAATAAAAGTTGAA
GAGCTTACATTAATTTAGTGGACACATGGCAAAATGTTAAAGTGATGATGTGCAATCATATGGAGAGAATGGTCAACTTTCATATATGATTTATCTCACAA
AAATACTATTTTTATTTTACCTGTGTTTTAAATATAAATTTTTATTTAGAAATTTTTAACTCGTCGAGTACATCTTTATAGTTAAATAATCTTACATTTTTAC
AATTTGGTCCAAATATATTTTCTAAAAATTTGGGAAATGAAAGATTTTAAATTTTCGGTTTTAACTGGATTATTGATGTGAATTTTTTTATTTATTTCTT
AATTAACAAATTAATAATATTAGCTAATCAAAGGTGTTAATTTAATTTACACAAATCTCAAATAATATGAGTAACTTTTAAAGAGTGTATGATTTCTAGTA
ATATTCACTTTTCTATAACATCTAAGTAATTAATTCGTTTTTGTGACTGAATTTTTAAATAAAATTCAGATATAACAAAATGGAGAAGAAAAGAAATAGTAACAA
AAAAACGTAGACAAACAAAATATTACACTTAACTTGGTGGTGTACCAAAAGTCACTAAGCAATTTAGATTGTTCCACCTATTTTACCTGTTTTGTTTTCCAG
GAATCAATATATAATCAACGAGGATCTAGTTCGAAATTTCTTTTAGTACAAACATTTATAATCTTAAATCCAAAAGAAACCGATGATGACGATTTTATTGTATA
ATCAAGTCAATGTAATGTTTTTAAATTTTTATCCACAGAGATGTAATTAATCATATTTAGATATTTCCCGGACCACCGTAAAAGATTATATGACCGGCTTCTGG
TATGGGGAATATTTAGTCTGTACAACTTGGTACACCGTCCAGTATCCGTCACGAATTTTTCCGGTATATGCCACTTTGGCGGTAAAGGCTTATTAACACCGTATAC
AGATATCCGTTTTGATATATGCAATCTAGCGCATACAGAAAATATCCATATCCCAATTTGAGGACTCGGCTGTGATGCACTATACGTTCTCCACCCCACT
CCACCAAATTTCCAGAGCTTTGATGAAATCTACTCGACGGCCAAAATCCAACCTTCTCTGCGACATTACCAAATGCAACCACCAATTTCCATTTATTTGGTCTGT
TACCCGAAAATTAAGGTGAAATTTTTTTCATTTAGTTTAAAGAAACACAAATATGATGATAAAAATTTGGAATGCTGTGTCATCTTAATTTAAACCTTGTAT
TAGACCGAAGATTGCATAATGAGTTGCTTTAGCACCACGAAACAGACACTGGCAATTAATGCCATACCAAGATGAAAATCTTGACGAACGAGTATCATTCTACGTCGC
AGTCATTTATGTAACATGATTTCCACCTGACTGTTCCATCAATATACATTTTCTTACTAACATGATACAAAATATAAAAATTAAGAAAAAATTTACTAGCAT
GATATGCGTGACTCACTATTGTTATGACGAAAAGTCCGAGGTTGGTCTGCTTGTACAAATGTTGAGTTTACCTGAAATAGTACATATAATGTTGTTTCAACAAAT
ATAATACATTTTGTAGTTACTAAAATAAATTAAGTAACTAGAGCTCGTTGACAGGTTGTTTTATCTAGTCAAGTTGTAGATGGCAACAAATTAATCAATA
ACTTACAGTAAAACCGTTTGGAGAAAATCACTTCCAATCTGAATGTAAGTCGAGAAGCAGTGAATGGTTTGGCTAACCTTAGGAGCGGTTAAACAAAGTTCCA
TCGTTGCTTCACTTGAACCTTGTCTTGGCCAAACCGCTCCAGCCACCGCAAACTAATGATGATAAAAATTTCAATTTCAATTAACAACTTCTTAATAACCAATATA
AATATATCAAAACATTTAATTAATAAAGAAAGCAACGTCATGATGTTGGTTGTCATATTAGTATCATAGTGGTAAAATTTCAAGAGCCACGAGGTTTATGATTA
CGCCAAATGACTGCGAATCAAAAAGATCTTCTTTGGTGACTCTTTGATGGGACGGTACCAATGGGCAACCCACACCTTCCCATAGGTAACCAAAATTTGGTA
TTATTTGCTTTTCTTCTCTTTTCCATTTGGGCGCATTAACAACTGTCGACACCAATAAATATATTTACTGTAGTTTCTTTTAAATTTAGTAACTGAAAT
ATAAGTATATAGAACTGATATATGAAATTTATATATACCTTGTAAATGGTATGTTGATTTCTTCAAGGGTGTATCAAAATGCTGGTTGCTGAAAGAAATCCACAT
CCGTATCTTTCCATCTATCTATCTATCATTTACAGCAATATAAAGAAGTTAACTTTTACTTAATTTGATAATATATAGAGAACAGTAACTGGTTATGAGA
TATATAAAAACCTTGTAGTCTTTGACAGCGTTTGTGTAGTAGCTTCT
ATTGTTATCGTTTATAATATGTTACCAAAAGTGAACACATTAAGAACAATCCATCACTTCCATTTGAAATTTCAAAATTTCAAAATTTCAAGTTCACGAATTTG
GAAGCTGGAAATAACCTTTAATAATAAAGCTGGAGGAAACAAACATTTTCAATTTCAAGACCAATTTGACCAAAATTTTGGGAAAGAAATTTTGGTATG
TAGATTAATCAATTAGCATTGACAAAAGTAATTAGTGTGTTGTTCAACAATTTCTGAGATAAAAATCTGTTAATATTTAGAAATTAACACATAATACCATTG
ATCATTTGGTAGATTAAACAATAGGCTAATAAGTATGTTACTTCTCTGTAATTAATTTTCAATATCATGTTATATATCTATAGTACCCATATTTTCCCGGAT
TTCTCTGGAGCAATAACATTAACAAAATAAATCCATGAATCAATGTTTGTGTTTCAAAATTTCAAAATTTCAAGTTCACGAATTTGCAAAATTTTGTAGCTCT
CCGATTGAGTTTCCGCAAAAGGTTGTTTTCTGTGTTAGTGGCCTCTTGTGAGGGTTTTCGTCGATTGGTCTCTCCGAGTTGGGACTCATGGGTTGGTCTGTG
GGCTTAAATGTTTATGGTGTCT
AATATGTTTACATCTCCATCTCTATCTTCCAGGACTTCCATCT
ATCATTTAGCTGAAGCTTAAACTTCCAGAAAATTAATTTATATCATATGAACGTAATACATATATATCTCATAAGCAACATCCATTTATTTCTTGTCTCTAGTT
TCATTTCAAGAAACCAATAAAGAGTGAACAAAGGTTCCCAAAAATAAACAATATAAAGACCAAAAGAAACACAAACCTTAGGAAGAAACAAACCAATAT
ATGTTATCGATCGATGGAGAATGACTCTTTGATTTCTCGATGTCGAAATATCCAGAGATTTGAATTTGTTTCTTAAGCAGCGCAGCATTTGTTCTTGG
CGCATTTCT
CTCAATTTGGATAACCGGTTTGTGCGAACCAACGCAATTTCTCTCTGTTTCCACCAACATATTCGAAACTTTGGTCCAGCAGGCTCTGTTAAACTTCTCTCATGT
TAAATTTTAAAAAATTTCAAGAACTCAAAAACAGAAAAGGTTTGTGTTCTGCTGCTCGAAAGTACCTCTCCCTCCAAAGGCTTATCGAGTTTCCCAATCTTGGT
GTCCCTGTACATATATGAACTCAAACTTTAGATAGATAAATTTCAAAATTTGAAATTTCAAAATTTCAAAATTTCAAAATTTCAAAATTTCAAAATTTCAAAAT
GAAGCTTTTCTCAAGACTTCCAGTTCTCTTGAAGTTTGTGATGAAATGTTGCTGCAAGTTCGATGAGAACTTTTCAAAATTTGATGTTTGGTATTTACCTT
AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTCAAACTTTAGACCAATTCAGTTTCCATAGCCAAAATTTCAAGATTTCTACTTAATTTGATTTG
ACATATATCTACT
GCTTTTGGTAGACCCGTTATCGCGGTTAACTTGTGCGACAAGATAAACTGTTTGTGCACATGCCATGCATTTCTGGTGTGTTCTGCAACGCCATTTCTGTAAG
AGATCAGACTAAAAGACTCAGGAAAAGAAAGAAAGAAAGTCAATAGTCTTGAGAGAGAGAGAGATAAAGAAAGAAATAGATGAAGATGAAAGAAAGAGATTTGCAAT
ATTGTTGCTTGGTTTTAATATGGGACAGATATATTGAAAATGAACATGACTTCTCTTTTAAATATTTAAAGTAAATTTCTTAAAGTAAATTTCTTAA
TAAATGTCACAACAAAAGTTCAAGATGAGTATCGAATTTGGATGCAAAATTCGAATATTTGGTGTCTGACACCCCTAGCATGTTTTTCTGTTGTTGCTAA
TCCACAAAACAGATAAGCGATGAATAAATAAATAATTTGAACGTTAGCAACAAAACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
TAAATATTTTGGCCAACTTTGAAGTAACTTTGATGTTTATTTTCCCTAGATGTTTATTTAGCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
CGTATATTTGTTCAATTTTATAACAACTGCAAAATGGTAAGCAATCTACAAAACACGGTATGTCAAATTTTGTATATGAAAGAAACATATCGTTTTCTTGGTCTG
ACATAACATATCTTTACCTTTTATCAAATCTCTTGTCTTTTTCCTCCAACTAATTTCCGTACTCTAATTTGGATATCTGTCGTGTGATAAATAAATAAATAA
AGTGGTAAATGAACATATGATGAACAAACACACACAGCTGCGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
CATGTTCCACACACTCTCGTAGATTCCACAGTCTTTTAGATTATTTTACATAGAAACAAACAAAATATGCAATGCAATACTCTATCAAAATTAATAACTT
TTTTTTTTTCTATTTCTGTTAGTTTGGTACGTTAGAAATATTTTGAAGTGGCAAGATACAAGTGTGTTGATCATTTGGCGTTACACATTTGATCTTAA
CATGAAATCCAGCTATACTAAAAGACATATGGGCTATGTAAGAAAGAAAGTCCAGGAAAATAAATGATGCGAACCGCAACCTTTGAGGTAATTAAGGAAATCCATA
ATGAGCTTTGAGTGGTACTTGTGTCATGCAAGACGATTTGGCACCAGGATATAGATGTTAGCAAGTGTGCAAAATCGGTTAAAATAATGCTTCAACAAAGT
AAAGTAAAAGACATCGCTGTAATTTG
AGTCAAGTGGCTTTATGCGCAAGATGAGTGAATGGCTTCCAAATTTCCGAACTGATGAGTTTGGTACTTTGGTATGATCGGCACTTTAAGACTAATTTTAGTGT
TACATAAATAAAGAGTGAAGTATGATAAATAAAGAAATCGGTTTGTAGTAAATTTGGAAGAAAATAATAAAGCAATTAATAAAGAGGTAAGGAACTTAACT
CCCAACAAAACCAACTCACATGAGGTAATTTTACTTTTACTTTAATGAAATATCAATAAATAATCAATTTGCTTAATTTGCTTAATTTGCTTAATTTGCTTA
CTATTGCTTTCTTGGGATTTAATACTTATATGATATAAATTTTGGACTTATGTAAGTTTAAATAGATGTTTATGTTAACAACAACTCCATTTAGGCCATACCTCT
TCCCTCACTGTTTTAAGGCTCTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
GCCACAGAAAAGAAAACAGTAAAGCTTGAAGTTTCTTGAAGTATTTCTGACATGATCAACGAGAGATGTTGATGAGAAATTTGCTACTTGTCTGCTGCTGCTCT
TTGAATGTAATATTTGGATGGCCTTACATTTCTTCTATAGTTGAGCTTTATGATTTGAAAATGCTTTTCAAGTACCTAGTGAATTTAATAAAGCAAGGATG
ATATATGTTCCAAAGCTGCCCCTCATCTGAGTACTATGACCATGATCAGTTTTTCTTGTGTTGACTTAAACAGTCACTGTTGTTGTTGTTGTTGTTGTTGTTG
GTTAATTTTGGAGACTGCTTCTGGGGAAGTCCAGGATATTTGAGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
GTGACGATCTGGCAGTCAAGATATCGGAAATCACAAATCGACTACCAAGTATGTTGGAAACCTATAAATGAGCTTGTCTTTGATGTTGTTGTTGTTGTTGTTG
AGTTATGCTACTGCTGCAAGGATGTTGGTATTTTCTTAAATAATTTGTTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT
TTTCCCATGCAATGCTGTTGAGTTGAGTTCAGAAATACCTTGGTAAAAGCACTCAATGACCACTTTGGTATCTGTCCTTCAAGGAGTGAAGATTAATCT
ATTCAGTCAATGCGCAAAACAGTATGAAGAAAGCTTGTTCGGTGTGGAGTACAGGATGATCTGGTGTGTTGATGACAAATTTGTTATCAACTTGGTATA

TCAGCACGCGTTCCTCGGCGGGTGACGGGACAAAGAGGACGATCTGCAGACACGTAAGATGTGAAACCGTGCCAAAGCTATGTGAGCGAGTGCATCGCGGAGCCGG
AGGAGTCTTCAGGGTACATGAACCAGTGGTGGGCTGTGGCTCGGAGCCCCACCCAAGTCTCACGGGAACTGTGATCCTGAGTGTGGAGCAGGGCTCGCTGGG
GGGTGACTGGGTGATGAGGGTGGACAGCCCTTGTGCTCTCGTAATCCTGAGTCTTGTGAGATCCGATGGTCTAAAAGTGGTGGCAGTGGCCGGGTGCGGTGG
CTCAGGCTGTCATCCCAGCCTTTGAGACGCCAAGGCGGGTGGATCACCCGAGGTCAGAGTTTCGAGACCAGCATGGCCAAACATGGTGAACCCCATCTCTCTTAA
AAATACAAAAATTAGCCGGGTGTGGTGGGCATGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGAGGAGAAATGGCTTGAACCCAGGAGGCGGAGGTTGCAGTGG
GTGAGATGGTGCCATTGCCTCCAGCTTGGGTGACAGAGCAAGACTCTGTCTCCAAAAAAGAGAAAAAGATGATACAGACAGCCTGTGTATTATGAAACAATTC
ATGCTGTATTAAAAACCTTCCCAGAGAAAACTATCCCAATGGCTTCACCTGATGAGTTCACCAAACATTCAAATAAAAAATAACACCCATTCATAACAATGGC
TCCAGAAAAGTGAATTTCTAAAGGACATTTTCCAAGCAACTCTCCGAGGCGACATCCCCCGATGCTAAACAGACCAAGCTATCACAAAGAAAAGGAGCCGAGGC
CAGTCTTCCAGAGCGACAGCAACCCCTGCCGAGGCTGGCAATGTGGACCAACACATAGGAGAAGTGTGACACTGTGTGGTGGACCAGGACCGGGGCTGGGG
GCTACCCAAAGACCGGGGCTCACCCAAAGACCGGGGCTCACCCAAAGACTGGGGCTCACCCAAAGAGCAGGGGCTCGCCAAAGACCGGGGCTCGCCAGGGACC
GGGGGCTCACCCAAAGACCGGGGCTCACCCAAAGACCGGGGCTCACCCAAAGACTGGGGCTCGCCAAAGGACCGGGGCTCACCCAGGGAGCAGGGGCTCACCCAA
AGACCGGGGCTCACCCAAAGACCGGGGCTCATCCAAGACCGGGGCTCACCCAGGAGCAGGGGCTCGCCAAAGACCGGGGCTCACCCAAAGACCGGGGCTCGCC
CAAAGACCGGGGCTCACCCAGGGACCGGGGCTCGCCAAACACAGGGGCTCACCCAAAGACTGGGGGCTCACCCAAACACCGGGGCTCACCCAGGGACTAGGGG
CTCACCCAAACACCGGGGCTCACCCAGGACCGGGGCTCACCCAGGAGCGGGGCTCACCCAAACACTGGGGGCTCACCCAAACACCGGGGCTCACCCAAACAC
CGGGGCTGCCTTACACTCAGACGCTCAGTCAAGTCAAGACAGTCTTATGACACAGAGCCTACCCCTGCGTGGCCGGGCAATGCAGCTCGGTGGAAACAGG
ACCGGGGCTCACCCAAAGACCGGGGCTGCCTTACACTCAGACGCTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT
GCAGAAAATACACATTTAGCATCCCAGCAAATTTCAAGGCACTTTTATAATTTTAAAAAGTAAACTGAGGCAAGGCTGATGGCTCATGCTTATCATCCCAGCA
CCATGAGAGGCTGAGCAAGAGGATCCTTCCAGCCAGGAGTCCAGACAGGCTGGGCAACATAGCAGAACCTGTCTTACCAAAAAGAAAATCAGCCAGGT
GTGGTGGCGGGTGCAGTAGTACGGCTACTCGGGAGGCTGAGGTGGGTGGATCACTGGAGCCCAGGAGTGGAGGCTCGAGTGCAGTGCAGTGCAGTGCAGTGC
TCCAGCCCGGGGCGAATGAGAATCTGCTCAAAAAAACAAGGCTCCAGCAAAGCAAGAAACAAAAGGGAAGGGGCTCCTGAGTCTGATGAAGGACGCCA
GGCAGGGCCGCTGCCTCAGCGCAGCTTGCAGCCGAGGCGGGAAGGGGCTGCTCTCCAGGCTTCCAGCACTGCAGTGGCAGGCGGAGCCTGTGTACAGAAATC
AGGGAGGAGACCCGATGCAGACGAGAAGAGAAAGGCTGAATGTGTCTCACCTGAAGATGACCTGACCCGTCGTTAGAGACTTCCCGGACTTAAAGGTTGAGCTGG
TAGAGCCGATGACTGAGTAAACTGGTCAGAAAATGCAAAATCCACATATAAACTTACAGTGCCTTCAACATTCACAAATGACAAATCATATTAAGTAAAAAA
ATACCAGTACAGTGCATCAGAAACAGGAAATGCTTGGGTATCAATGTGATGAACCGTGAACAGAACAACTACAAGGCGTACTGAAAGAGTGAAGATGCCTG
AAATAAGGAGAGCTGGACCGGCTCATGAGACCGAAGCCTCAATATGGTCAAGATTTCTTCAAGTTTCACTGGAGTCAACAGGATGCCATGTAATCCCAGCAGGC
TGTCGTAGCAACCGGAGGACCGGCTCAAAAATTACACGTTGGGAACGAAACACAGCTGACGCACTCGGGGAAAGCAGCGTGCAGGTTCCCGAGAAAGCCGGAAGTGG
AACTGCAGCGCACAGCAACACCGCAGTCCGCCCCCACCACACCGCAGTGAACACCGTGTGATATATGCCACAAACGCTTCCAGGTTGTGTCTCAAAATGG
TTACGGTGGTGAATCTGTTGTGTAATTTCCCAATAAAAAAAGTGCCGTGTCTGATTTCTGTTTACATAAAATCCAGAAAATCCGGCTAATCTATATTTG
ACAGAAAAGGAGGCTGAGGTAGCGTGGGATGAGGGCAGGTCGCGCTCAGAAAAGCAGCAGAAAAAGTCCAGCGCAGTGCCGCTCAAGCGCAAGGATGCAGGGACG
CCTATTGCTTGGTGTGTGGATAGTTTTCATGGTGGGTACAGACGCTCAAAAGTCAATGTAATTGTACTTTCCACAGCACCATGTCATCCCACATCAGTATATCTCA
ACACAGCTGCCCCAAATGAAAAAAGGCAAACTGATGAGAAGAAACAGTAAGTGGTCAACAGTACATCAGGAAATGAGGAGTGACAGCCAGGACCGGGGACGCC
TGCACACGCACAGAACGGCTAAAGGCAAGGCTGTGGCCACGTCGCCGACGGGGTCAACGCTGCTCCCTGGAAATGCCTCACTCAGGGCCACCCCGGAGCAGGGTT
TTTTGTTTTTTGTTTTTGCCTAGAGACACAGTCTTGTCTAGTCGCCAGGCTGGAGTGCAGTGGTGCAGTCTGGCTCACTGCAACCTCCACCTCCCAGGTTCAA
GCGATTTTCCTGTCTCAGCTCCCAGTAGCTGGAATGACAGGTCGCCGCCACCATACCTGGCTAATTTTTGTAATTTTAGTAGAGATGTGGTTTCATCATGTTGGC
CAGGCTGGCCCTGAACTCCTGACCTCAGGTGATCTGCCACCTCAGCCCTCCGAAAGTGTGGATTAGTGGAGTGAGCCACCACTCCTGGGCCCTGGAAGCAGTTTT
ATAGCTTCTACAAAATAAATATGTTGGCCACAGATGAGGACAGCTCACCTTCTGCGGATTAATCACCTCAGGAAAATGCAAGCATGCTGCGGTAATCAGTGCA
CAGACACAGCAGCTTTACTACAAGAAAGCAAAAACAAAGCAACCAGAGCTTATCAACCTGGAACAGGGGAGAGTGGAGCACACACAGCAGAAGCTACTC
AGTGATAGAAAAGAACTGCCATACGGCCAGGACACATAGTGCAGCCCAAGTATCCACACCTGGCCAGCCACAGGGAGCCCAAGGTAACACAGGCTTTGGCC
AGCCACACAGCAGCCACAGTATCCACACCCCGCCAGCCACACAGCAGCCCAAGGTAACCAGCCCTTCCGACGCCACAGCAGGCCACAGTATCCACACC
ACAGACTTAATCAGTCTCAACACACACATAGCGAAACATCTTGTGTACCTGTATACAAAATATTACTTGTCCATTCAAAATAAA