

# 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**.

#### CHYBY PŘI PREDIKCI GENŮ

Velmi zjednodušený přístup k predikci prokaryotických genů (genem je nejdelší ORF) vede k chybám, ale jejich množství je poměrně malé. Chyby mohou také vznikat při sekvencování DNA. Přidání/odstranění startovního a/nebo stop kodonu může vést ke zkrácení, prodloužení nebo úplnému vynechání genu.

#### ÚKOL 1 – příklady chyb vzniklých při sekvenaci

Pomocí predikčního programu **GeneMark** (<http://exon.gatech.edu/GeneMark>) identifikujte geny v sekvenačních výstupech a porovnejte je s původní sekvencí z databáze (část genomu *E. coli*). Určete, k jaké chybě došlo. Pro porovnání rozdílně predikovaných genů využijte program Align (<http://www.ebi.ac.uk/Tools/emboss/align/>).

#### SEKVENCE *E. COLI*

```
TTAAGAAATTCGGTATCAACTTCGAGGCCCTTTCAGGTACCCTGGCGTATTTTCGGGGCGATAGATTTGTATTTGTAGCCGCAACTG
CGATCGCCCCGGCGAAAAACGCAGATCCGGCGGTACTTTGTGAAATCACCGTTTCAATCCAGCCAGCGTGCCGGGTGTAACGAAACGAT
GGCGATCGTACAGTGTGCGGATCCAGACGTTTTCGCGCTGGAACCGGAGAAATCGACATCTGAGCCCCAGGCCGTGTGGAGTAGTCGATAGA
GTAGCGTTGCGAGTGCGCCCTGGTTGGCATCAGGCCACCACGAGAACCGGTGCGGCTAATCATACCCCAGGATAAAACAGCATCGTGGTGTG
GTAATTTACCCTGAGTAAAGTGGTCGAGACTCCAGCGCAGGTTAATGGCAGCGTGCCAGCCGCTGGAGAGATCCAGTAGCGAGAAGCCACCA
CGGTAGTGGAGTCAGATTCGGTATCGTTTCAGGTTCAGTGCCTTAAACCGCCCTGCACCAATAATATTGTTCCAGTGGATTCTTCAGCAGCGG
CATTTTATAGCTGAAGTCGAGGGTCTGTTCCGGCGCGAAATACTGGTACTGGTGGTTCAGACTGTGACCATAAGAGTTTCCACGGCTTTTTC
CAGTAGCTTTACGCGCGGTCCACGTCGGTAGAGTAACCGACCCCGTTTCGATGGTGTTCAGTTCGCGCGGAAACCACGCCCGTCAAGG
GTAATACTTTTCGTTTCGCGCGCTTTATCAAATGTGGAGCCACTACGACCGAGTTAAACCAGCCGGTAGCAGAAAGTCGACGGTTAAGTTCGGC
CAGATCTTTCGATTCGATCATCGCCCTTTTAAACGGCACCAGATTTGACGTTATTCATCGCGGATTTGTGATCCTTCAAAGGTCACATGC
CCAAAGCGGTAACGTTCCGCACTGTTATAATCAATATCCAGAAGGCTTTATGCAGGCCGAGCGCAATGCCAGCTGCGCTTGGTAAATTCGC
TATCGAAATAACCTTTAGCAACGCAATGCTGGTTAAGGACTTTTGAATTTTCATAATCGCCCTGGTTCAGTACCGTGCCAATAGCCGGGCG
AGTATCGAGCAATTTCAAATAGTCTTTATCGGTCGCGCGCGCGCGCAATACCACATCGGTGCCGCCAATTAACACCGGCACGCCCTGGCGTG
ACTTTGGCGGATCAATACCTGCCGCCCTTTCTTTGGCGGTGGACGGAGATCAAATTCATGGTTCGGTGGTAATAACCCAGCGCTTCAGACCTT
CGCGGATGGCATCATCGACGCTGCGCGAAAGCGACGGTCTGGCGTCACTTTCATCAATTCGATAGAAAGTGCAGCAGAACGTTCTTTTC
CAGCTGTCCCATAACCCCTCGACTGTAGACGGACGTTTCGCGGGACGGCAGATCCGTTAAGCAGAGTAAGCTTACACAGCATAACTGTCCG
ATATAGCGCATATTTCTCCTGAATATCCTTTTTCTCCTGCCCTTGAACCGCTTAAACAAAATCCAGTAATATGGATTAATAAAG
CAGACTAAACCCAAATATTTCTATGTTTTACTTTAGACCTATTCACGGTGGTATTGTTGTGCAAATACGCCCTTGTGTACAACCTTAACCCC
AATGACCGAATTTTCGGGAGAGCGACACCATGAGTTTATTTGATAAAAAGCATCTGGTTTTCCCGCCGATGCCCTGGACGTAAACACCCCG
ATGCCCTGAGCCAGCTGCATGCGGTCAACGGTCACTCAATGACCAATGTACCTGACGGAATGGAGATTGCCATTTTTCGGATGGGTTGTTTCT
GGGGTGTGGAGCGTCTGTTCTGGCAGTTACCCGGCGTTTACAGCACCGCCGACGGCTATACCGCGGCTATACGCCAAATCCGACTTATCGGGA
AGTGTGCTCCGGTATGCGGATGACCGAAGCGGTACGCATTGTTTACGATCCTTCTGTGCATCAGTATGAGCAGTTGTACAGGATTTTGG
GAGAATCAGATCCCGCCAGGGAATGCGTCAGGGCAATGACCGCAATGACCGCATCGATTCGTTTCAGCGATTATCCGCTGACCCAGAACAGGATG
CCGACGCTCGCGCAGTCTGGAACGTTTTCAGGCGCGATGCTTCCGCCGATGACGATCGTCACATCACCCAGGAAATCGTAAACGCCACACC
GTTTTATATGCGGAAGATGACCACCAGCAATATCTGCATAAAAACCCGATATGGTTACTGTGGAATTGGCGGAATTGGCGTCTGTGCCACCG
GAAGCATAGCGTTACGGTACAAATGTAGATTGTTGATAAAGTGCCTTTGTTTATGCCGAATGCAGCGTGAATGCCTTACCAGCCCTACAAAA
TCGTCCAATATCAATATATGCGAGGACTGCGTAGGCCATGACCGGATGACCGCATAGGCAATTCAGGCAATTTGCAATTTACGAGCTGGAATTAACC
CTCTGGCGACTTTACAGTACCTTACGCTATAC TAGCCACTGAAAATGCCGGATCACTTTCTTCGAATCGGCTTTCAATGTGATTTACACAAA
TTAATCAACTTCCCTTCCGAGGATCTGGCCTGAAAGTTCGGATAAGATATGTTAAACAGTATTTTAGTCATACTCTGCTTGATCGCTGTAAAGT
CGCTTCTCTCGATGCTGAGATCTCACTTGCCTCCTCAGCAAAATCAAATTAACCTGCTGGCTGATGAAGCAATATAAATGCCAACCGG
TTCGTAATATGCGAAGAAATCCCGCATGTTCTTACGGTGGTCCAAATCGGTTGACCGGCTGAAACGCAATTCGATTTTCGCGGATTCGTCGGTGTG
GGCATTCTTCCAGCTTTTACAGCCTGTTCTCCGCTATATGTCGGCAGAACCTCTGAGCAACTGAGCTTTATCTCTCTTTCTCGTTAGTG
ACTGGCATGTTTATCCTGTTTGGGATTTAACCCGAAACGCATCGGTATGATTGCGCCAGAAGCGGTGGCTTTGCGTATCATCAACCCGATGC
GCTTTCGCTGTACGTTTGCACCCCGCTGGTGTGGTCTTCAACGGCTGGCGAACATGATCTCCGATTTTCAAACCTGCCAATGGTACGTA
AGATGACATCACTCTGATGACATCTACCGGATAGTGAAGCCGGTGCCTGGCGGCTGTTACGTAACACAGGAACACGAGCTGATTTGAAAC
GTCTTTGAGCTGGAATCCCGTACCGTCCGCTTCAATGACACCGCTGAAAACGTGATTTGGTTTGTATCTCCACGAAGATGAGCAAAGCCTGA
AGAATAAGGTGGCGGAACATCCGCACTTAAGTTCCTCGTCTGTAATGAAGATATTGACCACATCATCGGTTATGTGATTTCAAAGACCTGCT
```

GAACCCGCTGCTGGCTAACCAAAGCCTGGCACTGAACAGCGGCGTACAAATTCGCAACACGCTGATTTGTGCCGGATACGTTAACCCCTTTTCAGAG  
GCGTTGGAAAGTTTTAAAACCGCAGGTGAAGACTTCGCGGTGATCATGAACAGTAGTACGCGCTGGTGGTGGGGATCATCACCCCTCAACGACGTGA  
TGACACCGCTGATGGGCTCGGTTCGTCAGGGCTGGAAGAGCAGATTTGTCCGCCGTGATGAGAACCTCATGGCTGATFGACGGCGGCCACCC  
AATTGACGACGTCATGCGCGTGTGGATATTGACGAGTTCCCGCAGTCCGGCACTACGAAACCATCGGAGGCTTTATGATGTTTATGCTGCGT  
AAGATCCCGAAACGCACCGATTTCGGTGAATTCGCCGGCTACAAATTTGAAGTGGTGGATATCGATAACTACCGCATCGACCAGCTGCTGGTGA  
CCCGGATCGACAGCAAGGCCACCGCCCTTTCCGCAAACTGCCTGACGCTAAAGATAAAGAAGAAAGCGTTCGGTAAACCCCGAGAACATCAAC  
GGCTCCTGAATCAGGACCGCTTATTACTCGATAGCACTTTGGTTAAGCCATCTCGTTTGCAGATAAAGCATTCAATTTTTCATATGATTTGTCGGCTGA  
TCACTGACAAATGCAGCTTATCCTTCACTTTTGGGATAAGAATCTTACCCTTATCAAATTCAAAAGCGCCAACGTCCTTGATGTATAACCGTCC  
ACGGAACAGGATCTTTACGTACTTCGCCACCTGAAGTGGGTTTACCGTTGAAAATTTTCATTTGTTATCTCCTGCTGAGTATTACGCCCTGTGC  
GGTGCCACAATCGGCCAACCTATTATGAGGCGCAATTTTAATGCCTAGTGACTATAGACTATCCGGGCAATGTTTCCACCGCGTATAACTTTT  
TTTACCTAAAGGTTAACCAATATTACAGAAATATCTTTTTACCGGAAGCGCGGTCTTCAGTATAAAGCATTCAATTTTTCATATGATTTGTCGGCTGA  
CCGCAAACTGGCATCACACTTGCAGGAAATTCGATAAATAGCACATATGATTAATAACTCAGACCCAAGTGGTCCGATCACCTGCATATCATAAG  
AAGGAAACACCATGACCTTACGCAAGATTCTGGCACTACCTGCCTGCCTGTCGCGATGATGGCTTCCGCACATCAGTTGAAAACCGGTCAGCG  
AGTGCCGCGGATTGGCATCACCGATTGGCGGAGTGGTGTGATAAAGATCAGTTTAGCTACAAAACCTGGAACAGCGCGCAGTTAGTGGGA  
AAAGTCGGACTGATGCAACATTTGCTGGTACCTCTGCAAAAAGAGAAAACCGCAGCGTGTATTGAAAGCATTAACCGCAAGCAGCGAATCCCG  
ACGATCGTTACCAGACCACCACCTTGTAAACACCGACGAGCAATTCGGGGTCCGGCATGTTTGTGCCGACGAGTCTGGAGAGTAATAAAAA  
GCTTTATCCCTGGTGCAGTTTTATTGTTGATAGCAATGGCGTGCACGCGGTGCTGGCAGCTGGATGAAAAAGTTCCGCTGTAGTGGTGTG  
GATAAAGACGGTCCGCTGCAATGGGCCAAAGACAGGGCGCTTACTCAGAAAGAGGTGCAGCAAGTGTGGACCTGCTGCATAAATTAATTAATA  
AATAGACTCTGATACCAAGATTTCAGGAACGACTCGCGCGGATTCAGTGCAGTAATGTTAACCGGTAATTCCTGCGAGTTCACCGAGCTGATGTTGCGGCTGA  
CGGGCCACAGCATGTCCAGCGCGGTGTCCTAAATATTCGTTGGTCCGAAAGCGCGGTTACAGCTGCGCCTGTCTTCCGCCACCAGGCAGAAAT  
TCAGCGAAGAGCGGATGACGTTGGTCTGATGTTGCCAAGCTGTTGCAGATACCTTTAGCTCCGCATCTGCATGGGAACGGCTGATAACCAC  
CAGCGCGGGCGCGCATCGCGCACTGAATCTGCTTCGCGCACCGCAGCTCTCTTCCAGGCCTTGCCTTCCGCGCGCTGTACATTAACGTTT  
ATTACCGCGCGCTGATACCAACCACTAAATCGGTTTGCATGGTCAATGTTAACCGGTAATTCGCGGTAATTCGCGGTAATTAATAACTCTT  
TAGTACCATCCAGCGGGTCTACCGCCAGTAACGCTGCCAGTGTGACGGACTTCCCAACCGGGAGGATCTTCTTCAGAAAGGACCGGAATATC  
CGGTGTACGCGTACGTAACCGTCCATGATAACGGTGTGAGCGGCAATATCCGCTGCCGTACCGGGAGATTGTCGCTTGTGCTGACGAGCTCC  
ATCGGTTCTGCTCCGCTGTAGACCTGCATAATGGCATCGCTGCATTCGGTGCAGCTGGCATACTTGATCTAACATTTCTCCACCTCGTCTCT  
GTGAGCGGTCTAATTTATGTTTACTTATACCCTATCGTTAATGAATGAAAGCGCAACTGTGATAGTGTATCATCTTTCAAAGCGTAAAATGTCG  
GCATTTCTTCACTGTTCTATAAGTAAGACGTTTTATCTTCTCTTTTCTTTCGATTCCCGATGATAAAGGATGTCCCTGATGATTAAGTTTAGCG  
CAACGCTCCTGGCCACGCTGATTGCCGCGAGTGTGAATGCAGCGACGGTGCATCTGCCATCATGAAACCACCTGATCTGCATAGCAACATGAT  
GGATTCGATTATTACAAGACACCGCCACGAAAAATTCGGACTGGTACGTACGGCAAGCTGATTAACGATGCCCGCAATGAAGTGAAAAAAC  
AGCTACTGGTTCGATAACCGCGATTTGATTCAGGGGATCCGCTGGCGGATTACATGTCGGCGAAAGGATTAAAAGACGGTGTATGTTTCATCCCG  
TTTATAAGGGCGTGAATACGCTGGATTATACGGTCCGTACACTCGGCAATCATGAATTTAACTACGGTCTGGATTACCTGAAAAATGCGTGGC  
GGGAGCGAAATCCCTTATGTAATGCCAACGTCATGAGCAGCAAGCAACAGCCTGATGTTTACACCGTATTTAATTAAGAGTACCGAAGTG  
TGCGATAAAGACGGAAAAAAACAGACGCTGAAGATTTGGCTATATTTGGCTGCTGCCCGCAAAATCATGGGCTGGGATAAAGCTAAATTTACCG  
GAAAAGTGACGGTGAATGATATTACCGAAACCGTTCGCAAAATAGCTGCCGTAATGTCGGCGAGAAAGGTTGCCGATGTCGTTGTCGTTTCCGCGCA  
TTCGCGGTCTGTGCCGATCCGTATAAAGTAATGGCGGAAAACTCAGTTTATTACCTCAGTGAATTCGGCGGCTTAACGCCATTATGTTTGGT  
CATGCTACGCGGTTTTCCCGGGTAAAGATTTTGTGATATCGAAGGGGTGATATCACCAAAGGCACGCTGAATGGTGTTCGGCGGGTAATGC  
CGGGCATGTGGGGCATCATCTTGGGGTGGTGCACCTTAACTCAGTAAATGACAGCGTAAATGGCAGGTGACCGAGCGGAAAGCGGAAGCTCCG  
GCCGATTTACGACATCGCAATAAAAAATCCCTCGCGGGGAGACAGCAAGCTGGTAGAAAACACTCAAAGCCGATCAGCATGCCACAGCCGAC  
TTTCGTACGAAGCAATCGGTAATCTGCCGACAATATGATAGCTATCTGGCGCTGGTGCAGGACGATCCGACCGTGCAGTGTGAACAACG  
CGCAAAAAGCGTATGTCGAACATTACATTCAGGGCGATCCGGATTCGGCAAACTGCCGGTGCCTTTCAGCTGCCGCACCGTAAAAGTTGGTGG  
TCGCAAAAATGACCCCGCAAGCTATGTGGAGGTGGAAAAAGCGCACTGACCTTCCGTAATGCCCGCATCTTTTATCTTACCCCAATACGCTG  
ATTGTGGTGAAGCCAGCGTAAAGAGTGAAGAGTGGTGGAGTGTCTCCCGGACAGTAAAACAGATGATCTTAACAGCAGCAAGAAACCGC  
AATCACTCATCAACTGGGATGGTTTTCCGCACCTATAAATCTTGTGATGTTGATGGTGTGAATTAACAGATTGATGTTTACCAACTGCCCGTTA  
TGACGGGAGTGCAGATGATTAATGCCAATGCGGAAAGGATTAAGAACCTGACCTTTAATGGCAGCCGATTGATCCGAAACGCGATGTTCCCTC  
GTTGCCACCAATAACTATCGCGCTTACGGCGGCAATTTGCCGCTACGGGCGACGCCATATCGCTTTTGCCTTACCAGGATGAGAACCGCTCCG  
TGCTGGCAGCTGGATTGCTGATGAGTGCAAACGTCGCGGGGAAATTCACCCGGCGCAGATAACAACCTGGCGTTTAGCACCAGTAGCTGGCGA  
TAAGAACTGGATATCCGTTTTGCAAAATCTCCGTCAGATAAAGCCCGCAGCTTTATAAAGAGAAAAGGCAATATCCGATGAATAAAGTCCGCG  
ACCGATGATATCCGGTTTGCATTTATCAGGTGGATTTGAGTAAAGTAAAACACTTCTTTTCCGCTATAAATCATCAACCGCATCCGGCATTT  
ATTTGGCGGATGCGATGCTGGCGCATCTTATCCGCCCTACAAGCCATGACCCGTGACCCAGATAAGCTCAGCGCATCCGGCAGTTATGCCGCAG  
TTCATCCCGCACCGCAACCTCGGGCAATTCAACTCACTCAGCTCCGCGAGTCAGCAGCACTTTTCGCTCACCTGCCAGCGCGCGGT  
AGCTATATTTCCACATCGCGCGCACCCAGGATACCGGATACCGTTAGGAAAACCATCTGTTCAACGCCCTGTAACGACTGCGCAAGCATCT  
TTTCGCTCACCCCGCGGATTTTGGCGCGAGGTCGCTAAAGCGATGAGTACCTTCGCGTAGCGCCACCAGAATCAACACCCCCCAACGGCTGGT  
GACGTGTTTCAACACCTCGCGGACGGGCACTGTTCCGCAAGAGATTACCCTCTTTCAGTTGTGTCAGCAGGCTAACCTGACTCATTTTACAT  
TTACTTTTGTGCGTACTACTAAAAGTAAGTTAGTGTGTAGCGTATTTAAACACAAGCAAAAACGATGGAGACTTCCCATGATCTGCTATTA  
CTGGTCCACTGGCAACTTGGTCACTATGTTTGAATCCTTGAAGAAAACCGTTCCTGTGACCCAAATAGTGCCATTACTCTGCTGCTTGAATCA  
AAAAGCCAGGCTCTGGCAGCACAGGCATTACCGTGCCTCAGGCTGACTACGGCGATGAAGCCGCATGACATCTGCACCTCAGGGAGTGGAA  
AAACTACTGCTGATCTTCCAGCGAAGTGGGTCAACGTGCCCGCAGCATCGTAATGTTATTAATGCCGCAAGGCGGTGGTGTGAATTTA  
TCGCTTATACAGCCTGCTACATGCAGATACCTCCCGCTCGCCCTCCCAATGAGCACATCGAGACGGAGAAAATGTTGGTGTATTCTGGCAT  
AGCCGATTTCCCGCGGCACTGAAAAGCGTCCGACTGCCCGCAGGACTGGCGGATATGCTGGCGGATTTGACGTTGGCGCATGAAAAGTGGT  
CTGTTTGTATAACAGCAAAACGTTAGCAAAATTGATTGGCCGCAACAGCAACGTTAGCCGAAAGCGTAAAGCACTTTTTTAAAGTAAATAACT  
AGTTAATTAAGTGGCATCCTCCCGCATCTCTCTAATAATGACGGGATGCCGGGAGCAATCATGCTGCTTCTTGAACCTTCTTCTGACAGAC  
CAATGGATGCCAGTAATGATTAGCGCGTGCTGTACGCCCTGTGACAGGGTGTAGTGGGGCTTATTTTTGTGGCGCGTGTGATCGTACCGG  
AATACCCGGCATGTTGACGTCGATGGGGCTTATCTGGCGTATAGGTTAATGGCGTGGCAATTTGCCTGGCTGGGACGCGTGGTCTCGTCA  
GTTGGCGCTGGCACTGGCTTACCGCCTGATGCTACTATGATGGCAACCTCATTTTATTACTTCTGCTGCTGCGTACGCAATGATTAAGTAACT  
GGCGCGCTGTTTCCAGCATGATTATCGGCACCTGCCGGTGGTCACTTCTGCTTTGCAATCTGCTTTATAGCCAGCGGACGGCAAACTCG  
CGTGGGAAAACTCGCCCGGCACTGATTTGATTTGGCATCGGCTGGCGAGTGTGAATATTGCTGAGTTAAACCACGGACTCCCGGATTTTGA  
CTGGGCAGCTTAACTCTGGCATCTGTAGCGTTAGTTCCGTTGCTGCTGGCATGGTATGCCCTGGCAACCGCCGCTGGCTGGCGGAA  
AATCCCGCAACACATCCGATGATGTGGCGGACGGCGCAGGCGTGGTGCCTGACTCGCTCCGCTTCTCATCGGCTATCTCGTCTGTTACTGGC  
TGAATATAACAACCGCGACTTCTCTTACCTTTTGGCCCGCTCGCTGGTGTATTATAGTCTGATGGTTGCGATAGCCGTGCTTGTCTCATG  
GTTGGCGCACTCTGCTGGAACCTGCCAGCCAGCGATTACCGCATGAGGATTCCTGGGGCGCTGATTGTTTTGAAACCGTGGCAGGTTTGTCT  
TACACTTTTTTACTCCCGCAGAAAATGCCCGGCTAATGACGCTGAGCGGTATCGCGTGTAGTGTATGGCTGCTGCTAGCTGAGGTTGCGGTCAGAGCAA  
AACAGAAAACCTTTAACTGAATCTGCTCAGAAAAGTTGACACGCTGGCAGTGTAGTTAAAATAAGCCTCTGCTACGTAAGGGTTATAGCTTTTG  
CCTTAAAGATGCATTTAAAATACATCTTATCTTATTAAGAAATGAGGTATCAGCTATGGCTTATCGCGACCAACCTTTAGTGAACGCGCTCT

CTATCCCTCGCGCTTCGGCTCTGTTTCGTAATATGATATGGATTACTGCTGTGGCGGTAAGCAGACCGCTGGCGCGCGGGGGCACGTAAGA  
ACTGGATGTTGAGGTCAATGAAGCTGAAGCTGAACTGGCAAAGCTTGCTGAACAACCGATTGAGAAAGACTGGCGTAGCGCCCGCTGGCAGAATCATT  
GACCATATCATCGTGGCGTAC

## ПРИКЛАД 1

TTAAGAAATTCGGTTTCAACTTCGAGCCCTTTCAGGTACCCTTGGCGTATTCGGGGCGTAGATTTGTATTTGTAGCCGCGAACTACTG  
CGATCGCCCCCGCGGAAAAACGCAGATCCGGCGGTACTTTGTCGAAATCACCCGGTTTCAATCCAGCCAGCGGTGCCCGGTGTAACGAAACGAT  
GGCGATCGTACAGTGTGGGATCCAGACGTTTTCGCGCTGGAAACCGGAGAAATCGACATCTGAGCCCCAGGCGGTGTTGGAGTAGTCGATAGA  
GTAGCGTTGGAGTTCGCCCTGGTTGGCATCAGGCCACCACGAGAACCGGTGGCGTAAATCATCACCCAGGATAAAACAGCATCGTGGTGTG  
GTAATTTACCCCTGAGTAAAGTGGTCGAGACTCCAGCGCAGGTTAATGGCACCGTACCACATCGGTCGCCCAATTAACAGCCGAGGAAGCCACCA  
GCGTAGTGGAGTCAGATTCGGTATCGTTCAAGTTCAGGTCAGTGCCTTAAACCCGCCCTGCACCAATAATATTTGCCAGTGGATTTCTCAGCAGCGG  
CATTTTTAGCTGAAGTCGAGGTCTGTTCGGCGCGGAAATACTGGTACTGGTGGTTCAGACTGTGACCATAAGAGTTCATCCAGGGCTTTTTTC  
CACGTAGCTTTCAGCGCGGTCCACGTCCTGTAGAGTAAACCGCCGGTTTCGATGGTGTTTTCAGTTCGGCGGAAACCCAGCCCGTCGAAAG  
GTAATACTTTGTTTCGGCGCTTATCAAATTTGGAGCCACTACGACCGAGTTAAACCCAGCGGTAGCAGAAAGTCGACGGTTAAGTTCCCGC  
CAGATCTTTGATTCGTACTCATCGCCCTTTTTAAACGGCACCAGATTTTGCAGGTATTCATCGCGGATTTGTGATCTTCAAAGGTCACATGC  
CCAAAGCGGTAACGTTCCGCACTGTTATAATCAATATCCAGAAGGCTTTATGACGGCCGAGCGCAATGCCAGCTGCGCTTTGGTAAATTCGC  
TATCGAAATAACCTTTACGCAACGCAATGCTGGTTAAGGACTTTTTGAAATTTTCATAATCGCCCTGGTTTCAGTACCCTGCAATAGCCGGGG  
AGTATCGAGCAATTTTCCTGAAATCTTTTCCTGCTCCGGCCCGCGCCGATTAACCAAAATTCAGTAAATTAACAGCGGTAAACGCGGTAAG  
ACTTTGGCGATCAATACCTGCCGCCCTTTCTTTGGCGGTGGACGGAGATCAAATTCATGGTTCGGCTGGTAATAACCCAGCGCTTTACAGACCTT  
CGCGGATGGCATCATCGACGCGTGGCGGAAAGCGGACGCTGTCGGGTCACTTCACTTCATTCATTCGTAAGAGCTGCGCAGCAAGCTTCTTTTC  
CAGCTGTCCGATAAACCTTCGACTTCAGCTGTAGACGGACGTTCCGGCGCAGCGAGATCCGCTTAAAGCAGATAAGCTTACACAGCATAACTGTCCG  
ATATAGCCGCAATTTTTCTTCTGAAATCTTTTCTTCTGGTGGAAACCGGATTAACCAAAATTCAGTAAATTAACAGCGGTAAACGCGGTAAG  
CAGACTAAACCCCAATATTTCTTATGTTTTACTTTAGACCTATTCACGGTGGTATTGTGTGCAAAATACGCCCTTGTGTACAACCTTAAACCC  
AATGACCGATTTTCGGGAGAGCCACCATGAGTTATTTGATAAAAAGCATCTGGTTTTCCCGCGCGATGCCCTGCCTGGACGTAACACCCCG  
ATGCCCGTAGCCAGCTGCATCGGTCACCGTCAACGGTCAATGACCAATGTACCTGACGGAAATGGAGATGGCCATTTTTCGATGGGTTGTTTTCT  
GGGTTGGAGCGTGTGTTTCGGCAGTTACCCGCGCTTTACAGCACCAGCGGTATACCGCGGCTATACCGCGGCTATACCGCAAATTCGACTTATCGGTA  
AGTGTGCTCCGGTGATACGGGTATGCCGAAGCGGTAGCATTTGTTACGATCCTTCTGTATCAGCTATGAGCAGTTGTACAGGATTTTTGG  
GAGAATCAGATCCCGCCAGGAATGCGTCAGGGCAATGACCAGGCAAGCAGTATCGTTCAGCGATTTATCCGCTGACCCAGAACAGGATG  
CCGACGTCGCGCCGATGTCGAAACGTTTTTCAGGCGGCGATGCTTGCAGCGATGACGATCGTTCACATACCACGGAAATCGTAAACGCAACCC  
TTTTTATATGCGGAAGATGACCACAGCAATTCGCATATAAAAACCGGATAGTTACTGTGGAATTTGGCGGAAATGGCGCTGTGCTGCCACCG  
GAAGCATACCGTTACGGGTACAATGTAGATTTTGAATAAGTCGTTTTGTTTTATGCCGAATGCAGCGTGAATTCCTTACCAGGCTTACAAAA  
TCGTCAAATTAATATATTGCAGGACTGCGTAGGCTGATCGGCATAGCGCATCAGGCAGTATTGCATTTATCAGCAGTCTGAATTTTAAACC  
CTCTGGCGACTTTACAGTACTTACGCTATACTAGCCACTGAAATCGCGGATCACTTTCTTCGAATCGGCTTTTCAATGTGATTTTACACAAAA  
TAACTCAACTTCCTTCCGAGATTCGGCTGAAAGCTCGGATAAGATATGTTAAACAGTATTTAGTACATCTGCTTGTGATCGCTGTAAAGT  
CGGTTCTTCTCGATTTGAGATCTCACTTCCGCGCTCACGATAAATCAAACTGACTGGCTGTGGAAGCAATATGATCTTCCGTATTTAAACCTTCC  
TCTGAAATATGCAGGAAATCCCGGATGTTCTTTACGGTGGTCCAAATCGGTCTGACGCGATTTGCGATTTCTCGGCGGTATCGTCCGGTATGC  
GGCATTTCTCCAGCTTTTACAGCCTGTTCTCCCGCTATATTCGCGAGAACTCTCTGAGCAACTGAGCTTTATTTCTCTTTCTCGTTAGTGC  
ACTGGCATGTTTATCTGTTTGGCGGATTTAACCCGAAACCGCATCGGATGATTTGCGCCAGAAAGCGGTGGCTTTGGCTATCATCAACCCGATGC  
GCTTCTGCTTATACGTTTGCACCCGCTGGTGTTGTTTTCTTCAACGGCTTGGCGAATGATCTTCCGTATTTTCAAACCTGCAATGGTATGGTAA  
AGATGACATCACTTCTGATGACATCTACGGGTAGTGGAAAGCGGTGGCGTGGCGGGCGTGTACGTAAACAGGAACACGAGCTGATTGAAAC  
GTCTTTGAGTGGAAATCCCGTACCGTTCGCTTCAATGACACCGCGTGAACAGCTGATTTGGTTGATCTCCAGCAAGATGAGCAAAAGCTGA  
AGAATAAGGTGGCGGAACATCCGCACTTAAGTTCTCGTCTGTAATGAAGATATTGACCACATCATCGGTTATGTGATTTCTAAAGACCTGCT  
GAACCCGCTGCTGCTTAAACCAAGCTTGCCTGACGCGCGTACAATAAATTTGAAGTGGTGGATATCGATAAATCAACCCGATCGACCGTTCCGTGA  
CCGGATTGGAAAGTTTTAAACCGCAGGTGAAGACTTCGCGGTGATCATGAACGAGTACGCGCTGGTGGTGGGATCATCACCTCAACGACGTGA  
TGACCACGCTGATGGCGCATCTGGTCGGTCAGGGGCTGGAAAGACAGATTTGTCGCCCGTATGAGAACTCATGGCTGATTGACGCGCGCACCC  
AATTGACGAGCTCATCGCGGTGCTGGATTTAGCAGATTCCCGCAGTCGGGCACTACGAAACCATCGGAGGCTTTATGATGTTTATGCTGCGT  
AAGATCCCGAAACGCACCGATTCGTTGAAATTTCCGCGCTACAATAAATTTGAAGTGGTGGATATCGATAAATCAACCCGATCGACCGTTCCGTGA  
CCCGGATCGACAGCAAGGCCACCGCCCTTTCGCCAAAACCTGCTGACGCTAAAGATAAAGAAGAAAGCGTCCGCTAACCCCGAAACATCAAC  
GGCTCCTGAATCAGGACCGCTTATTACTGCATAGCACTTTGGTTAAGCCATCTCTGTTTGCAGACGCATAAATGACGGTTAATTCGGGACA  
TCACTGACAAATGCAGCTTATCTTCACTTTTGGGATAAGAATCTTACCCTTATCAAATTCAAAAGCGCCAACCTCTGTATGTAATAACCGTCC  
ACGGAACAGGATCTTTTACGTACTTTCGCCACCTGAAGTGGGTTGTACCGTTGAAAATTTTCATTGTTATCTCCTGCTGATTTACGCTGTGC  
GGTGCCCAATCGGCCAACTATTTAGAGGCGCAATTTTAAATGCCTAGTGACTATAGACTATCCGGCAATGTTTCCACCGCTATAACTTTT  
TTTACCTAAAGGTTACAATATTCAGAAATTTCTTTTACCGAAGCGCGCTTTCAGTATAAGCATTCATTTTTTCAATGATTTTGTGCGGTTGA  
CCGCAACTGGCATCACACTTCGGGAAATTCGATAAATAGCACATATGATTAACACTCAGACCAAGTGGTGGATCACCTGCATATCATAAG  
AAGAAACACCATGACCTACGCAAGATTTCTGGCATTACCTGCTGTTGCGGATGATTTCCGCATGATTTCCGCATCAGTTCGAAACCGGTACGCG  
AGTGGCCGATTTGGCATCACCGATCGTGGCGAGTTGGTGTGATAAAGATCAGTTTGTGCTACAAAACCTGGAACAGCGCGCAGTTAGTGGGA  
AAAGTGGCAGTACTGCAACATATTGCTGGTCGACCTTCGAAAAGAGAAAAACGCGACGCTGATTGAAGCGATTAATCAGCGAAGTTACC  
ACGATCGTTACCAGACCACCATTTGTTAACCCGACGAGCAATTTCCGGTTCGGCATGTTTGTGCGCAGCAGCTGGAGAGTAATAAAAA  
GCTTTATCCTTGGTCGAGTTTATTGTTGATAAGCAATGGCGTGGCAGCGCAATATCCGCTTACCGGAGAAATTTCCGCTGTGATGGTGTGCTG  
GATAAAGACGGTTCGCTGCAATGGGCAAGACAGGGCGCTTACTCAGGAAGAGGTGCAGCAAGTGTGACCTGCTGCATAAATTAATTAATA  
AATAGACTCTGAAACCGGATTCAGGAACGACTCGCGCGGAGTGAATTCAGCGGTTTACCCTGCCAGTCTGTAACGCTGCGCTCCGGCAGT  
CGGCCACAGCATGTCAGCGCGGTGTCCTTCCAAATATCTGTTGGTCCGAAAGCGGGTACAGCTGCGCTGCTTCCGCCACCGAGCAGAAAT  
TCAGCGAAGAGCCGATGACGCTGGTCTGATGTTCCGCAAGCTGTTCAGATACCTTTTCAGTCTCGCATCTGCATGGAAACCGGTGATAACCC  
CAGCGCGGGCGCATCGGGACCTGAATCTGCTTGCGCACACCGCACTTCTTCTTCCAGGCTTCCCTTCTGCCGCGTGTACATTACGTTT  
ATTACCGCGCATAACACCACTAAAATCGGTTTGCATGGTCAATGACGCGCAATGTTAACCGTGAATTCGCCATTCGCTTTAATAAACTCTT  
TAGTACCATCCAGCGGTCTACCAGCCAGTAACGCTGCGAGTCTGACGGACTTCCCAACCGGGAGGATCTTCTTCAAGAAAGGACCGGAAATC  
CGGTGCTCAGCGTACGTAACCGTCCATGATAACCGGTGAGCGCGCAATATCCGCTTCCGCTTACCGGAGAAATTTCCGCTTTGCTGAGTGGTCTC  
ATCGGTTTTCGTCCCGTCTGACCTGCATAATGGCATCGCTGCATTCGTTGCAAGCTGGCATACTTGTATCTAACATTTCTCCACTCGTCTCT  
GTGAGCGGTGTTAACTTATGTTTACTTATAACCTATCGTTAATGAAATGCGCCAACTGTGATAGTGTGATCATTTTTCAAAGCGTAAAAATGTG  
GCATTTTCACTGTTCTATAAGTAAGACGTTTATTTCTTCTTTCTTCTGATTTCCCGATGATAAAAAGGATGTCCCTGATGATTAAGTTTAGCG  
CAACGCTCCTCGCCACGCTGATGTCGCGCAGTGAATGCAGCGACGCTGATCGGATCATGGAACCCACTGATCTGCAGCAACATGATG  
GGATTTGATTTTACAAAGACACCGCCACGAAAAATTCGGACTGGTACGTACGGCAAGCCTGATTAACGATGCCCGCAATGAAGTGAAGAAAC  
AGCGTACTGGTGCATAACGGCGATTTGATTCAGGGGAGTCCGCTGGCCGATTCATGTCGGCGAAAGGATTAAGAGCAGGTGATGTTATCCGG  
TTTATAAGGGCGTGAATACGCTGGATTTACGGTCCGTACACTCGGCAATCATGAAATTAACCTACGGTCTGGATTACCTGAAAAATTCGGTGGC  
GGAGCGAAATTCCTTTATGTAATGCAACGCTGATGACCGCAACGCAATGTTTTACCGGATTTAATTAAGAGTACCGAAGTGTGCGATAAAG  
GTCGATAAAGACGGAAGAAAAACAGACGCTGAAGATTTGGCTATATTTGGCGTCTGCGCGCCAAATCATGGGCTGGGATAAAGCTAATTTATCC

GAAAAGTGACGGTGAATGATATTACCGAAACCGTGCAGCAAACTACGTGCCTGAAATGCCGAGAGAAAGTGCCGATGTCGTTGTCGTTTCGGCGCA  
TTCGGGCTGTCTGCCGATCCGTATAAGTAATGGCGGAAAACACTCAGTTTATTACCTCAGTGAATTCGGGGCGTTAACGCCATTATGTTTGGT  
CATGCTCAGCCGTTTTCCCGGTTAAAGATTTTGTGATATCGAAGGGCTGATATCACCAAAAGCACGCTGAATGGTGTTCGGCGGTAATGC  
CGGGCATGTGGGGCGATCATCTGGGGTGGTCGACTTACAACCTCAGTAATGACAGCGGTAAATGGCAGGTGACGACGGCGAAAGCGGAAGCTCG  
GCCGATTTACGACATCGCCAATAAAAAATCCCTCGCGGGGAAGACAGCAAGCTGGTAGAAAACACTCAAAGCCGATCACGATGCCACACGCCAG  
TTCGTGACGAAGCAATCGGTAATCTGCCACAATATGTATAGTACTCTGGCGCTGGTGCAGGACGATCCGACCGTGAAGTAGTGAAACAACG  
CGCAAAAAGCGTATGTCGGAACATTACATTCAGGGCGATCCGATCTGGCAAAACTGCCGCTGCTTTCAGCTGCCGACCGTTTTAAAGTTGGTGG  
TCGCAAAAATGACCCGGCAAGCTATGTGGAGGTGAAAAAGGCCAGTTGACCTTCCGTAATGCCGCGATCTTTATCTTACCCCAATACGCTG  
ATTGTGGTGAAAGCCAGCGGTAAAGAGGTGAAAGAGTGGCTGGAGTGCCTCCGCGGACAGTTTAAACCAGATTGATCTTAACAGCACGAAACCGC  
AATCACTCATCAACTGGGATGGTTTCCGCACCTATAACTTTGATGTGATTGATGGTGTGAATATCAGATTGATGTTACCCAACTGCCCGTTA  
TGACGGCGAGTGCAGATGATTTAATGCCAATGCCGAAAGGATTAAGAACCTGACCTTTAATGGCAAGCCGATTGATCCGAAACCGCATGTTCTC  
GTTGCCACCAATAACTATCGCGCTTACGGCGCAAAATTTGCCGTTACGGGCGACAGCCATATCGCTTTTGCTTACCAGGATGAGAACCGCTCGG  
TGCTGGACGCGTGGATTGCTGATGAGTGCAAAAGTGCAGGGGAAATTCACCAGCGCGAGATAACAACCTGGCGTTAGCACCGGATAGCTGGCGA  
TAAGAACTGGATATCCGTTTCGAACTTCTCCGTCAGATAAAGCCGAGCGTTTATTAAGAGAGAAAGGGCAATATCCGATGAATAAAGTCGCG  
ACCGATGATATCGGTTTGCAGTTTATCAGTGGATTGATAGTAAAGGATTAAGAACTGACCTTTTTCGCGCTATAAATCATCAACGCCATCCGGATTT  
ATTGGCGGATGCGATGCTGGCGCATCTTATCCGCCCTACAAGCCATGCACCGTAGACCAGATAAGCTCAGCGCATCCGGCAGTTATGCCGCACG  
TTCATCCCGCACCGCAACACTCGGGCAAATCAACTCAATCCAGTCCGCCAGTGCAGCAACCTTTTCGCTCACTGCTCGCCAGCGCGGTG  
AGGCTATATTCACATCGCGGCGCACCACCGGATACCGGATACCGTTAAGAAAACCATCTGTCTAACCGCTGTAACGACTGCCAAGCATCT  
TTTCGCTCACCCCGCGATTGTCGCGCAGGTCGCTAAGGATGATGAGTGCAGCGCAGCTGCTGCGTAGCCACAGAATCAACCCAGCCAGCTGGT  
GACGTGTTTAAACACCTCGCGGACGGGCACTGTTCCGCAAAGAGATTACCCTCTTCAGTTGTGCGACAGGCTAACCTGACTCATTTCATAC  
TTACCTTTTTGTGCGTACTTACTAAAAGTAAGTTAGGTGTTAGCGTATTTAAACACAAGACAAAACGATGGAGACTTCCCATGATCGCTATTA  
CTGGTGCACCTGGCAACTTGGTCACTATGTTATGAATCTTGAAGAAACCGGTTCTGTCAGCCAAATAGTGGCTATCGTTGCTAATCCGCG  
AAAAGCCAGGCTCGCAAGCAAGGATTACCGTGCAGCTGAGTGCAGCGGATGAAGCGCATGACATCTGCATGACGATGACGATGACGATGAA  
AAACTACTGCTGATCTCTCCAGCGAAGTGGGTCAACGTGCCCGCAGCATCGTAATGTTATTAATGCCGCAAAGGGCGTGGTGTGAAATTTA  
TCGCTTATACAGCCTGCTACATGCAGATACCTCCCGCTCGGCCTCGCCAATGAGCAGATCGAGACGGAGAAAATGTTGGCTGATTCTGGCAT  
CGTTTACACCTGCTGCCAACGGCTGGTACAGCGAAAACCTCCTGCCAGCGCCCGCGCAGCACTGGAACACGGCGTATTTATCGGTTGGCGG  
GGCAGTGGCAAAAATCGCTCAGCAACCGGGGAGATTTGCGGCAGTGCAGCAGCGCTGATTAGCGAAGCCGGTCACGAAGGCAAGGTTTACG  
AACTGGCGGGCGATAGTGCCTGGACATTGACACAGTTAGCGGCAGAGTGACCAAACAGAGCGCAAACAGGTTACCTATCAAAATCTGAGCGA  
AGCCGATTTCCGCGCGGCACTGAAAAGCGTCCGACTGCCCGACGGACTGGCGGATATGCTGGCGGATTCTGACGTTGGCGCATCGAAAGTGGT  
CTGTTTGATAACAGCAAAAACGCTTAGCAAAATGATTGGCCGCCAACAGCAACGTTAGCCGAAAGCGTAAGCCATCTTTTTAATGTTAATAACT  
AGTTAATTAAGTGGCATCTCCCGCATCTCTCTAATAATGACGGGATGCCGGGAGCAATCATGTCTGCTTCCGAACTTCTCTGACAGAC  
CAATGGATGCCAGTAATGATTAGCGCGTGTCTGACGCCCTGTTAGCAGGGTTGATGTTGGGGGCTTATTTTGTGCGGGCGTGTATCGTACACCG  
AATACCAGCGCATGTTGAGTGCATGGGGCGTTTACTGGCGTTAGGGTTAATGGCGCTGCCATGTCCTGGCTGGGACCGGTGCGTCTGGTCA  
GTTGGCGCGTCCGGACTGGCTTACCOCCTGATGCTCACTATGATGGGCAAACTCATTTTATTACTTCTGCTTGGCAGTGGCATTCAACGTA  
GGCGCGCTGTTTCCAGCATGATTTATCGGCACCCTGCCAGCTGCTTCCGTTGCTTGGCAATCTGGCTTATAGCGAGCGGACCGCAAACCTC  
CGTGGGAAAACCTCGCCCGGCACTGATTTGATTGTCATCGCCCTGGCGAGTGTGAATATTGCTGAGTTAAACCACGGACTCCCGGATTTTTGA  
CTGGGACGTTTATACCTCTGGCATGCTGCTAGCGTTAGTTCCGTTGGTCTGCTGGGCATGGTATGCCCTGCGCAACGCCCGCTGGCTGGCGGAA  
AATCCCAGCAAACTCCGATGATGTGGGCGAGCGCGAGCGTGGTCAAGCTGCCGGTTTCTCTCATCGGCTATCTCGCTGGCTGTTTACTGGC  
TGAATATACAAAACGCCGACTTCTCTTACCTTTTGGCCCGCTCCGCTGGGTTTATTAGTCTGATGGTGTGTTTATTAGTCTGATGTTTGGCTGCT  
GGTTGGCGCACTCTGCTGGAACGTCGCCAGCCAGCATTACCAGACGTTTCTCGGGCGCTGATTGTTTTGCAAAACGCTGGCAGGTTTGTCTG  
TACACCTTTTTACTCCGCCAGCAAAATGCCCGCGCTAATGACGCTGAGCGGTATCGCGCTGTTAGTGATTGGCGTGGTCAATGGCTCAGAGCAA  
AACCAGAAAACCTTTAACTGAACTCTGTCTCAGAAAAGTTGACACGCTGGCAGTGAATTAATAAGCCCTGCTGACGTAAGGGTTATAGCTTTTG  
CTTAAGATGCATTTAAATACATCTTATCTTATTAAGAATGAGTATCAGTATGCTTATCGGCAAACTGAGTGTGTTTACCGCAACCTGAGTGGCTCT  
CTATCCCTCGCGCTTCGGCTCTGTTTTCGTAATAATGATATGATAGTACTGCTGGCGGTAAGCAGACGCTGGCGCGCGCGCGCCGACGTAAGA  
ACTGGATGTTGAGGTCAATGAAGCTGAAGTGGCAAGCTTGTCTGAACAACCGATTGAGAAAAGCTGGCGTAGCGCCCGCTGGCAGAAATCAT  
GACCATATCATCGTGGCTAC

## PRÍKLAD 2

TTAAGAAATTCGGTTATCAACTTCGAGGCCCTTTACAGTCCACCGTATTTGGCGTATTTCCGGGGCGATAGATTTGTTATTTTGGCCGGAATACTG  
CGATCCCGCCCGGCAAAAACGCGATCCGGCGGTACTTTGTGCAAACTACCGGTTTCAATCCAGCCAGCGTCCGCGTGTAAACGAACGAT  
GGCGATCGTACAGTGTCCGATCCAGACGTTCTGCGCTCGAATAACGGAATAATCGACATCTGAGCCCGAGCCGCTGTTGGAGTAGTCTGATAGA  
GTAGCGTTGCGAGTCCCGCTGGTTGGCATCAGGCCACCAGAGAACCGCTGGCGCTAATCATCACCAGGATAAAAACAGCATCGTGGTGTG  
GTAATTTACCCCTGAGTAAAGTGGTGCAGACTCCAGCGCAGGTTAATGGCAGCTGCCAGCCGCTGGAGAGATCCAGTAGCGAGAAGCCACCA  
CGGTAGTGGAGTCAATTCGGTATCGTTCAGTTCAGTGGCTTAAAACCGCCCTGACCAAAATAATTTGTTCCAGTGGATTCTTCAAGCAGCG  
CATTTTTAGTCAAGTGCAGGTTGTTCCGCGCGGAAATATCTGTTGCTGACGATGACCAATAAGAGTTTCAATCCAGGCTTAAACCGGCTTTTC  
CACGTAGCTTTACGCGCGGTTCCACGTCGCTAGAGTAACCGACCCCGGTTTCGATGGTGTTTTCAGTTCCGCGGCAAAACCGCCCGTCAAGG  
GTAATACTTTGTTTTGCGCGCTTTATCAAATTTGGAGCCACTACGACCGAGTTAAACCAGCCGCTAGCAGAAAGTGCAGCGTTAAGTTCGCG  
CCAAGCGGTAACGTTCCGACAGTGTATAATCAATATCCAGAGGCTTATGTTGCGGAGCGCAATGCCAGCTGGCTTGGTAAATTCGCG  
TATCGAAATAACCTTTACGCAACGCAATGCTGGTTAAGGACTTTTTGAAATTTTCATAATCGCCCTGGTTACGATACCGTGCCAATAGCCGGCG  
AGTATCGAGCAATTTCAAATAGTCTTTATCGGTCGCGCGCGCCGCGCAATACCACATCGGTTGCCGCAATTAACACCGGACCGCTGGCGTG  
ACTTTGGCGATCAATACCTGCCGCCCTTTCTTTGGCGGTGGACGGAGATCAAATCAATGGTCCGCTGGTAATAACCCAGCGCTTTCAGACCTT  
CGCGGATGGCATCATCGACCGTGCAGCAAGGAGCGTCTGCGGCTCACTTCACTACTTTTCAATCGTAGAAAAGCTGCGCACGAAGCTCTTTTC  
CAGCTGTCCCGATAAACCCCTCGACCTGTAGACGGAGCTTCCGCGGACGGCAGATCCGCTTAAGCAGAGTAAGCTTACACAGCATAACTGTGCG  
ATATAGCCCATATTTCTCTGAAATATCTTTTTCCTGCCCTGGAACCGCGTTAAACGGCTTAAACAAAATCCAGTAATATGATTAAGAAAG  
CAGACTAAACCCCAATATTTCTTATGTTTTACTTTAGACCTATTCACGGTGGTTATTGTTGTAACAAACCGCTCTGTTTAAACCTTAAACCC  
AATGACCAGATTTTCGGAGAGGACACCATGATGTTTATTGATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG  
ATGCCCGTAGCCAGCTGCATGCGGTCACCGGTCACCTCAATGACCAATGATACCTGACGGAAATGGAGATTGCCATTTTTGCGATGGGTTGTTTCT  
GGGGTGTGGAGCGTCTGTTCTGGCAGTTACCCGGCGTTTACAGCACCGCCGACGGCTATACCGCGCGCTATACGCCAAAATCCGACTTATCGGGA  
AGTGTGCTCCGGTGCATCGGGTCATGCCGAAGCGGTACGCAATGTTTACGATCTTCTGTGATCAGCTATGAGCAGTTGCTACAGGATTTTTGG  
GAGAATACGATCCCGCCAGGAAATGCTCAGGGCAATGACCAACCGCAGCAGTATGCTTTCAGCGATTATCAGCGATTATCAGCGATTACCCAGAACAGGATG  
CCGACGCTCGCGCAGCTGGAACGTTTTACGGCGCGATGCTTGGCGCGATGACGATGCTCACATCACACCGAAATCGTAAACGCCACACCC  
GTTTTATATGCCGAAGATGACCACAGCAATATCTGCATAAAAACCCGATGTTGTTACTGTGGAATGGCGGAAATGGCGTCTGTCTGCCACCG  
GAAGCATAGCTTACGGGTACAAATGATGATGTTGATAAAGTGGCTTTGTTTATGCCGAATGCAGCGTGAATGCCTTACCAGGCTTACAAA  
TCGTTCAAATTAATATATGCAAGGACTGCGTAGGCTGATGCGCATAGCGCATCAGGCAATGATGCAATATCAGCAGTGTGAATTTTTAAC  
CTCTGGCGACTTTACAGTACTTACGCTATACTAGCCACTGAAAATGCCGGATCACTTTCTTCGAATCGGCTTTCAATGTTGATTTTCACACAAA

TTAATCAACTTCCCTTCCGAGGATCTGGCCTGAAAGGTCGGATAAGATATGTTAAACAGTATTTTGTAGTCATACTCTGCTTGATCGCTGTAAAGT  
CGCTTCTTCTCGATGTCAGATCTCACTTCCCGCTCAGCAAAATCAAACCTAAACCTGCTGGCTGATGAAGGCAATATAAATGCCAACCGG  
TTCGTAATATGAGGAAATCCCGCATGTTCTTACGGTGGTCCAAATCGCTGTAACGCAAGTGGCATTTCCGGCGGTATCGCTGGTATGC  
GGCATTCTTCCAGCTTTTACAGCCTGTTCTCCCGCTATATGTCGGCAGAACCTCTGAGCAACTGAGCTTTATTTCTCTTTCTCGTTAGTG  
ACTGGCATGTTTATCCTGTTTGGGATTAACCCCGAAACGCATCGGATGATGTCGCCAGAAGCGGTGGCTTTGCGTATCATCAACCCGATGC  
GCTTCTGCCTGTACGTTTGACCCCGCTGGTGTGGTCTTCAACGGCTGGCGAACATGATCTCCGATTTTCAAACCTGCCAATGGTACGTAA  
AGATGACATCACTTCTGATGACATCTACGCGGTAGTGAAGACTTCGCGGTGATCATGAACGAGTACGCGCTGGTGGTGGGATCATCACCTCACACGTGA  
GTCTTTGAGTGGAAATCCCGTACCGTTCCTGCTTCAATGACACCGGTGAAAACGTGATTTGGTTTGATCTCCACGAAGATGAGCAAAGCTGA  
AGAATAAGGTGGCGGAACATCCGCACTTAAGTTCCTCGTCTTAATGAAGATATTGACCACATCATCGGTATGTCGATCTAAAGACCTGCT  
GAAACCGGTGCTGGTAAACAAAGCCTGGCACTGAACAGCGCGTACAAATTCGCAACACGCTGATTTGGCCGATACGTTAACCTTTCCAGAG  
CGCTTGGAAAATTTAAACCGCAGGTGAAGACTTCGCGGTGATCATGAACGAGTACGCGCTGGTGGTGGGATCATCACCTCACACGTGA  
TGACCACGCTGATGGCGATCTGGTCCGTCAGGGCTGGAAGAGCAGATTGTCGCGCGTATGAGAACTCATGGTGAATGACGGCGGCACCCC  
AATGACGACGTATGCCGCTGTGGATATTGACGAGTTCGCGAGTCCGGCACTACGAAACCATCGGAGGCTTTATGATGTTTATGCTGCGT  
AAGATCCCGAAACGCACCGATTCCGTTGAAATTCGCGGCTACAAATTTGAAGTGGTGGATATCGATAACTCCGCACTCGACCGCTGCTGGTGA  
CCCGGATCGACGCAAGGACATCGCCCTTTCGCAAAATGCCGCTGAGCTGACGCTGAAAGTAAAGAAAGAAAGCGTCCGCTAACCCCGGATGAAAC  
GGCTCTGAAATCAGGAGCCGCTTATTACTGCATAGCACTTTGGTAAAGCCTCTCTGTTTGCAGACGCATAACCTGACGGTTAACTTCGGACA  
TCACTGACAAATCGACGTATCTCTTCACTTTTGGGATAAAGACTTACCTTATCAAATTCAAAAGCGCAACCGTCTTATGATAAACCCTCC  
ACGGAACAGGATCTTACGTACTTCGCCACCTGAAGTGGTGTACCGTTCGAAAATTTTCATTTGTTATCTCTGCTGAGTATTACGCCCTGTGC  
GGTGGAAAATTTAAACCGCAGGTGAAGACTTCGAGCGCAAAATTTATGAGCGCAAAATTTTATGAGCGCAAAATTTTATGAGCGCAAAATTT  
TTTACCTAAAGGTTACAATTTATCAGAATTTATCTTTTACCGAAGCGCGTCTTCAGTATAAGCATTCAATTTTCATATGATTTGTGGCCTGA  
CCGCAAACTGGCATCACACTTGGCGGAAATTCGATAAATAGCAGATATGATTAAGCACTCAGACCAAGTGGTCCGATCAGCTGCATATCATAAG  
AAGGAACACCATGACCTACGCAAGATTTGCGCACTCACCTGCTGCTGTGCGGATGATGGCTTCGCGACATCAGTTGCAAAACCGGTGAGCG  
AGTGGCCCGGATTTAAACCGCAGGTGAGGACTTCGCGGTGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
AAAGTGGAGTACTGCAACATATTGCTGGTCCGACTCTGCAAAAAGAAAACCGCAGCCTGATTAAGCGATTAAATCAGCGAAGTTACCGC  
ACGATCGTTACAGACACCACCATTTGTAACACCGCAGCAGCAATTCGCGGTTCCGGATGTTTGTGGCAGCAGCTCGAGAGTAAATAAAA  
GCTTTATCCCTGGTCCGAGTTTATGTTGATAGCAATGGCCTGCCACGCGGCTGCTGGCAGCTGGATGAAAAGTTCCGCTGTAGTGGTGTCT  
GATAAAGCTGCGTGCATCGCGCAAAAGCAGGGGCTTACCTCAGGAAAGGTCAGCAAGTGAATGAGCTGAGTAAATGATGAAATTTTAAATA  
AATAGACACTCTGAACCCAGGATTCAGGAACGACTCGCGCGGAGTGAATCCAGCGGTTTACCTGCCAGTCTGAAACGCTGCGCTCCGGCAGCT  
CGGCCACAGCATGTCCAGCGGGGTGTCCAAATATTCTGGTTCGAAAGCGGGTACAGCTGCGCTGTCTTCCGCCACACAGGAGCAAT  
TCAGCGAAGAGCCGATGGACGTGGTCTGATGTTCCGCAAGCTGTTGAGATACTCTTTCAGCTCCGCTGCTGCATGGGAACGGCTGATAACCC  
CAGCGCGGGCGCATCGCGGACTGAATCTGCTTGGCAGCAGCAGCAATTCGCGGTTCCGGATGTTTGTGGCAGCAGCTCGAGAGTAAATAAAA  
ATTACCGCGCATAACCCACACCTAAATTCGGTTTGCATGGTCAATGACGCGCAATGTTAACCGTGAATTCGCCATTGCGTAAATAAATCTT  
TAGTACCATCCAGCGGGTCTACAGCCAGTAAAGCTGCGAGTGTGAGCGACTTCCCAACCGGGAGGATCTTCTCAGAAAAGGACCGGAAATCC  
CGGTGCTCAGCGTACGTAACCTCCATGATAACCGGTGTGAGCGCAATTCGCTGCGCTTACCGGAGAAATTTGCTGACGAGCTGC  
ATCGTTCGTCGCTGAGACTGCATAAATGGCATGCAGCTGCGTTCGTAAGCTGGCAGTAACTGAGTAACTGACATACTTGAATCAAACTTCACCTGCTCT  
GTGAGCGGTGTTAACTTATGTTTACTTTATACCTATCGTAAATGAATGCGCAACTGTGATAGTGTATCATATTTCAAAGCGTAAAATTTGTG  
GCATCTTCACTGTTCTATAAGTAAAGCTTTATCTTCTTTCTTTCTTTCGATTTCCCGATGATAAAAGGATGTCCCTGATGATTAAGTTAGCG  
CAACGCTTCTGGCCAGCTGATTTGCCCGCAGTGTGAATGACGCGCAGGCTGATGCTGCATATGAAAACCACTGATCTGCATCAACATGAT  
GGATTTCCGATAATTACAAACACCGCACGGAAATAATTCGAGTGGTGCATGCGCAAGCCTGATTAACGATGCGCCGAAATGAAGTAAAAT  
AGCGTACTGGTCGATAACCGCGATTGATTACAGGGAGTCCGCTGGCCGATTACATGTCGCGCAAAAGGATTAAGAGCAGTGTATGTTATCCGG  
TTTATAAGCGCTGAATACGCTGGATTATACGGTCCGTACACTCGGCAATCATGAATTTAACTACGCTGTTGATTTACCTGAAAATGCGCTGGC  
GGGAGCGAAATTTCCCTTATGTAATGCCAACGTCATTGACGCGCAGAACCAACAGCCTGTTTACACCGTATTTAAATTAAGATAACCGAAGT  
TGCATAAAGACGGAAAAAAGCAGCAGCTGAAGATTGGCTATATTTGGCTGCTGCCGCGCAAAATCATGGCTGGGATAAAGCTAACTTTATCCG  
GAAAAGTACCGGTGAATGATATTACCGAAACCGTGGCAAAATTCGCGCAAAATTCGCTGAAATTCGCGCGAAAGGTTGCCGATGTCTGTTGCTGCT  
TTCGCGGTGCTGCGGATCCGTATAAAGTAAAGGCGAAAACCTCAGTTTATTAACCTCAGTGAATTCGCGCGTTAACGCCATTATGTTGGT  
CATGCTCAGCGCTTTTCCCGGTAAGATTTTCTGTATATGCAAGGGCTGATATCCAAAAGCGCAGCTGAATGGTGTTCGCGCGGTAAATCG  
CGGCTATGTTGGCGCATCTTGGGGTGGTGCACCTACAACCTAGTAATGACAGCGTAAATGGCAGGTGACGCGAGGCAAGCGGAAAGCTGCG  
GCCGATTTACGACATCGCAATAAAAATTTCCGCGCGGAAGTCAACTCAACTCCGCGAGTGGTAGAAACTCAAAGCGATCATGATGCAACCGCAG  
TTGCTCAGCAAGCAATCGGTAATCTGCCACAATATGATAGCTATCTGGCGTGGTGCAGGACGATCCGACCGTGCAGTAGTGAACACG  
CCGAAAAGCGTATGTCGAACATTAATTCAGGGGATCCGGATTCGCGCAAAACTGCCGCTGCTTTCAGCTGCCCGACCGTTAAAGTTCGGT  
TCGCAAAAATGACCCGCAAGCTATGTGGAGTGGAAAAGCGGAGTGAACCTTCCGTAATGCCGCGGATCTTTATCTACCCCAATACCGTG  
ATTGTGGTGAAGACCGGTAAAGAGGTGAAAGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
AATCACTCATCAACTGGGATGGTTTCCGCACCTATAACTTTGATGTGATGATGGTGTGAATTTATCAGATGATGTTACCAACCTGCCCGTTA  
TGACGGCGAGTGCAGATGATTAATGCCAATGCGGAAAGGATTAAGAACTGACCTTTAATGGCAAGCGGATGATCCGAAACCGCATGTTCCCT  
GTTGCCACCAATAACTCTCGCCTACGGCGGCAAAATTTGCCGTAACGGGACAGCCATATCGCTTTTGGTTCACCGATGAGAACCGCTCGG  
TGCTGGACCGTGGATTCTGATGAGTGCAGATCGCGGGAAATTCAGCGGGAATACCAGCGCGGAGATAACAACTGAGCGGTGAGTACCGATGAGT  
TAAGAACTGGATATCCGTTTCGAAACTTTCCTCGTCAGATAAAGCGGACGCTTATTAAGAGAAAGGCAATATCCGATGAATAAAGTTCGG  
ACCGATGATATCGGGTTTGCATTTATCAGGTGGATTTGAGTAAGTAAACACTTCTTTTTCGGCCTATAAATCATCAACCGCATCCGGGATTT  
ATTGGCGGATCGCATGCTGGCGCATCTTATCCGCCCTACAAGCATGCACCGTAGACCAGATAAGCTCAGCGCATCCGGCAGTTTACCGCAGC  
TTCATCCGCAACCGCAACACTCGGGCAAAATTCAACTCAACTCCGCGAGTGGTAGCAACATCGAGCAACCTTTTCGCTACCTGCTGCCAGCGGCTG  
AGGCTATATTTCCACATGCGCGGCACCACCGGATACCGGATACGGTAAAGAAAACCTCTGTCTAACGCTGTACGACTGCGCAAGCATCT  
TTTCGCTCACCCCCGCATTTTTCGCGCGAGGTGCTAAGCGATGAGTACCTTCGCGTAGCGCCACCAGAAATCAACACCCCAACGGCTGGT  
GACGCTTTTCAACACTCGCGACGGGCACTTTCCGCAAGAGATTTACCCTTTTCAGTTGTTGGCAGGCTAACCTGACTCATTTTCATAC  
TTACTTTTTGTGCTACTTACTAAAAGTAAAGTTAGTGTATTAGTGTATTAAACACAAGACAACCGATGGAGACTTCCCATGATCGTTATTA  
CTGGTGGCACTGGCAACTTGGTCACTATGTTATTTGAAATCTTGAAGAAAACGGTTCCTGTGACGCAAAATAGTGGCTATCGTTGATTAATCCGGC  
AAAAGCCAGGCTTGGCAGCACAAGGATTAACCGTCCGCTCAGCTGACTACGGCGATGAAGCCGCACTGACATCTGCACCTACAGGGAGTGGAA  
AACTACTGCTGATCTTCCAGCGAAGTGGTCAACCTGCCCAGCATCGTAATGTTAATGCGGCAAAAGGGCGCTGGTGTGAAATTTA  
TCGCTTATACCGCTGCTGATGAGTGCAGATACCTCCCGCTCGCGGATCGCAATGAGCAGCATCGAGACGAGAAATTTGGTGTATTTGGCAT  
CGTTTACACCCTGCTGCGCAACGGTGGTACAGCGAAAACCTTCGCGCAGCGCCCGGCGAGCTGGAACACGGCGTATTTATCGGTGCGGG  
GGCGATGGCAAAATCGCCTCAGCAACCGGGGAGATTAATGCGGAGCTGCGGCAACCGGTGATGAGCAAGCGGTCACGAAAGCGGATTTACG  
AACTGCGGGCGGATAGTGCCTGGACATTGACACAGTTAGCGGACAGCTGACCAACAGAGCGGCAACAGGTTACCTCAAAATCTGAGCGA  
AGCGGATTTCCCGCGCGCATGAAAAGCGTCCGACTGCCCGAGCTGCGGATATGCTGGCGGATCTGACGTTGGCGGATTTGCGGATTTGGCAGTGGT  
CTGTTTGATAACAGCAAAACCGTTAGCAAAATGATTGGCGCCCAACGACAACGTTAGCCGAAAGCGTAAGCCATCTTTTAAATGTTAATAACT  
AGTAAATTAAGTGGCATCTCCCGCATCTCTCTAATAATGACGGGATGTCGGGAGCAATCATGCTGCTTCTGAACTTTCTTCTGACAGAC  
CAATGGATGCCAGTAAATATGCGCGCTGTGACGCCCTGTAGCAGGTTGATGGGGGCTTATTTTTGTGGGCGCTGATCTGACCGG  
AATACCGGCGATGTTGGCAGTGGTGGCGCTTATCTGGCGTTAGGTTAAATGCGCTGCCATTTGCCGCTGGGAGCGGCTGCTGCTGCTCA  
GTTGGCGCTCGGACTGGTACCCTGCTGATGCTCACTATGATGGCAACCTCATTATTAATTTCTGCTTGCCTGGCAGTGCATTCAACGTA

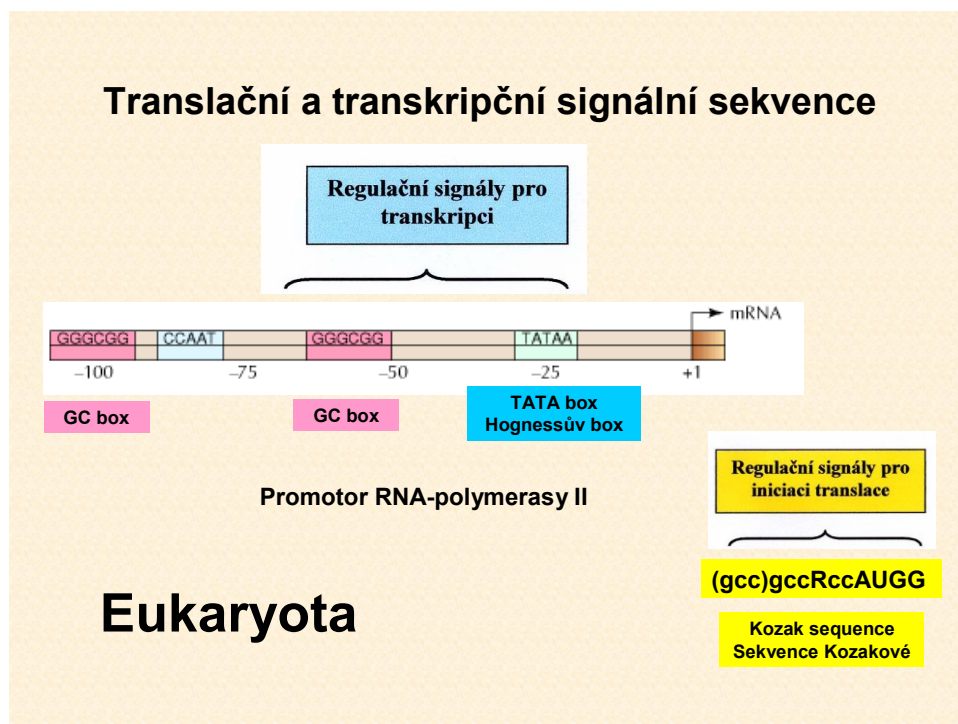


CACTATTATAACAATTCAATTGAAGAGCATTACAGTATAAAAATCCAATATCTAATAATCAAATGTGCCAATAATCGATTGATATCTCGTCAA  
 GTGTTCAATTTGTTAGTAAATGATGATGATTTCACTTTGGTATTGAATAGAATTACCAATCACCTAGCATAAAATGAATATATTTGGAAATTAT  
 TAGAAGATATTAATGATATCCATTGATTATAAAACCCTTGATTTACCTGATTATCTAATACCCAACCTGGGGGATCAATTGGGTAATAAGTT  
 GGATTCCTTTGGGTGATTACAATTAGTATATTCATCTTCTACATCTAATAGGTTGAGAAATTTTATAATTGATATCCGAAAAACGATTTACCA  
 AAATTGCTAGTGATGGGAAATATATCAAGATATTGTTAAATTCATGTATAATAATACCAAAAATAAAATTTGAAAAATTAAGAAATAGAAAAAT  
 TTATTAATAATTAATTAATATGCTAGTGATGGGAAATTTAAATTAATAATTAATGATAATCAATTAATTTGGTTTTTATTGATTGAATCTATAAA  
 ATCATCTTACAATAAGTAGTAAGTACGTGGTCAATAAAGTAATGATATATATGTGCGTGACTGAGTGATGATTTTCTATTTTATCTCGCGC  
 GTTGTATTTTTTGTGTTATTGTT  
 ATTTTTTTTTTCTCTCTCTCTCCCACTTCTACTCTTTGAACATGTGGTGAAAAAATAAATAAGACCAGATCTTTTTTATCCATATAT  
 TTCAAACCGACTTGTTCCTTTTTTCTTTTACAACCAATTAATACAAAAGAAAGACAACAATCATAATGTCTACTTCTGTTGAACCCAATGAA  
 ACAGAAGCTTTGTTGAGAAAGCAGAATGATCTTTCCACAACCTGCCTCAATTGAAGAAAAATATCCTCACCAACAGGAGAGGCTGCAGAAGCG  
 ATGACGACACTCTAAAAGAACCAATATGATGAAGCTAAAGAAACCGCTGAATCTTTAAAACAAGTTGAATCGATATTAGCACCTATTGTTTT  
 CACTGCATGTCAATTTTCGTGAGATTTTATCGTATTTTCAAGTAAATGACCATGTTGTTGGGATGAAGCTCATTTTGGTAAATTTGGATCCTAT  
 TATTTACGACACGAATTTTATCATGATGTTTCACTCTCCATTGGGTAATGTTAGTTGGTTTATCTGGTTATTTGGCTGGATACAATGGATCTT  
 GGGATTTCCCAAGTGGTAAAAATACCTGATTATATTGATTACTAAAATGAGATTGTTAATGCCACTTCTGCTGCTGTTGTTGTTGTTGTTGTTGTT  
 GGCTTATTTCACTGGGAAAGAAGTTGGATTTTCCATGTTTACTACTGGTTATTTACTTTGATGGTGGCTCTTGAATCAAGTTATGTCACTTTA  
 GGTAATTCATTTGTTGGATTCAATGTTGTTATTTCTCACCGTTGCTACTGTTTCTGTTTTTCACGTTTCAACAATTTAACAATAAATCAC  
 AAGAATTTCTAGAAAAATGGTGGAAATGGATCTTTTAACTGGTGTTCATTGGTGTACTTGTCTGTAAAAATGGTTGGATTGTTTGTGTTAC  
 CACATTGGTCGGGATTTTCACTGTTGTTGACCTTTGGAATAAATGAGTGATAAATCTATTTTCATGGACAAAATACATTCAACATTTGGTTGCT  
 AGAATTTGTTGCTTTGATTCTTGTCCCAATTTTCAATTTTTCATGCTTTTCAATTTAAAGTTCATTTTGTATTTGTTGATAAATCGGGTACTGGTGATG  
 CCAATATGTCATCACTTTTCCAAGTAATTTGGCTGGTTCGATGTTGGTGGTGGCCACGTAAGTATCCATGTTCCACTCGGTTATCACTTT  
 AAAGAATCAAGGTTAAGTGGTGGCCTTTTACTCTCCACGTTCAACATTTCCAGAAGGTTCAAAACAACAACAAAGTTACTACTTATGGTCAC  
 AAAGATTCAAAACAACAATTTGATTTTCCAAAGAGCTAGAGGACAACCTTATTATGATACTTCTGGTAACACCACGACATTGAATATATTTTTG  
 ACGGTATGCATGTAAGATTGATGCATCCACAACTGGTAGAACTTACATACTCATGATATCCAGCTCCAGTGTCTAAATCTGAATATGAAGT  
 TGCATGTTATGGTAATTTGACTATTGGTGATCCTAAAGATAATTTGGACTGTTGAAATTTGGAACAAGCAAGTATGAAGATAAATAGATTA  
 CATCTTTGACTTCGTCATTTAGATTGAAGAATGAAGTATGATGATTTGTTATTTGGGGTCACTGGTACTACATTACCTCAATGGGGTTCCAGC  
 AAGGTGAAGTTGTTGTTTACAAGAACCATTAAAAAGACAAGAGAAGT

### ÚKOL 3

Charakterizujte část genomu *Candida albicans* (kvasinka) rovněž pomocí heuristického modelu.

### 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 setřihu (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 5**

Analyzujte část genomu mnohobuněčného eukaryotického organismu také pomocí programu **GENSCAN** (<http://genes.mit.edu/GENSCAN.html>). Porovnejte s výsledky z úkolu 4.