

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*

???

PŘÍKLAD 1

???

PŘÍKLAD 2

???

PREDIKCE GENŮ U EUKARYOT

Genomy **jednobuněčných** eukaryot se navzájem výrazně liší (frekvence intronů, jak velká část genomu je tvořena geny kódujícími proteiny). Pro některá jednobuněčná eukaryota (s nízkou frekvencí intronů) je možné použít stejné postupy jako pro prokaryota.

ÚKOL 2

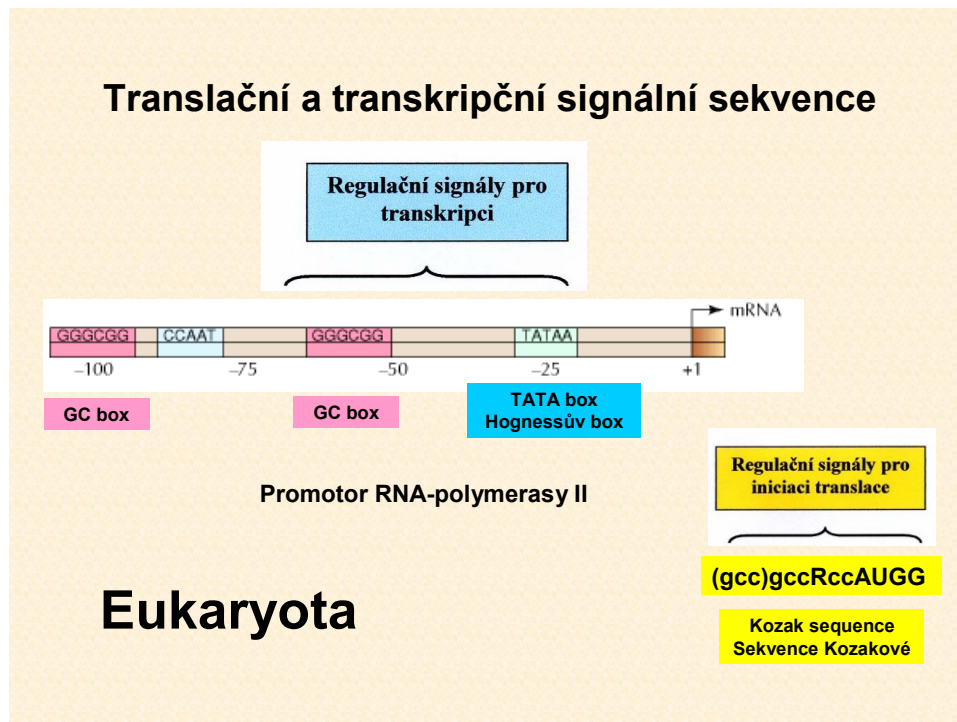
Charakterizujte část genomu *Candida albicans* (kvasinka) pomocí programu **GeneMark** (<http://exon.gatech.edu/GeneMark>), verze pro nižší eukaryota. U identifikovaných genů určete možnou funkci.

???

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 4

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

???

ÚKOL 5

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