

Softwarový projekt rPredictor

Genová translace je biologický molekulární proces implementace genetické informace, jehož výsledkem je živý organismus. Centrální jednotka genové translace je ribozom.

Ribozom, kromě jiných dosud známých molekul, obsahuje i molekulu ribonukleové kyseliny (rRNA), která tvoří molekulární kostru ribozomu, a proto zásadním způsobem přispívá k jeho funkci. Protože funkce biologických molekul je především určena jejich prostorovou strukturou, pochopení funkce rRNA v rámci ribozomu při genové translaci závisí na znalosti a pochopení funkce struktury rRNA.

Zatímco sekvence (posloupnost nukleotidů) rRNA je v dnešní době relativně snadno zjistitelná, určení prostorové struktury rRNA je stále technicky extrémně obtížné; výsledkem jsou známé sekvence rRNA pro přinejmenším stovky eukaryotních organismů, zatímco prostorová struktura je známá pro pouhé 4 eukaryotní organismy.

Mezičlánek mezi sekvencemi a prostorovými strukturami jsou sekundární struktury, jejichž znalost sice nenahrazuje znalost prostorových struktur, nicméně umožňuje v omezené míře studovat chování rRNA v rámci ribozomu. Sekundární struktury však – na rozdíl od prostorových – je možné předpovědět ze sekvencí. Na tuto situaci se zaměřuje tento projekt, jehož cílem je vývoj nástrojů, databáze a webového rozhraní schopného predikce sekundárních struktur rRNA na základě jejich sekvence a zpřístupnění těchto predikcí formou webového rozhraní široké vědecké komunitě. Pro predikci struktury RNA bude využita vhodná kombinace stávajících nástrojů (pravděpodobně z balíku Vienna RNA Package - <http://www.tbi.univie.ac.at/RNA/>) a jako zdroj sekvencí rRNA bude využita databáze SILVA - <http://www.arb-silva.de/>.

Projekt se tedy skládá ze tří částí - vývoj algoritmů, vývoj databáze a vývoj webového rozhraní. Následuje seznam požadavků na jednotlivé části:

- Predikce a vývoj bioinformatických nástrojů
 - vývoj algoritmické pipeline pro predikci sekundární rRNA struktury z nukleotidové rRNA sekvence
 - implementace podpůrných bioinformatických nástrojů (např. podobnostní hledání v RNA sekvencích přístupné z webové části)
- Databáze
 - definice struktury databáze pro uchování sekvencí rRNA a příslušných metadat
 - identifikace dat nutných k importu a jejich průběžná aktualizace
 - specifikace datových toků mezi jednotlivými částmi frameworku
- Webové UI
 - modul search
 - fulltext search
 - vyhledávání podle sekvence
 - vyhledávání podle metadat
 - možnost exportu množiny struktur podle specifikované vyhledávací podmínky
 - modul detail (náhled na jednotlivé rRNA molekuly)
 - náhled na RNA sekvence

- náhled na predikci
- náhled na metadata
- možnost exportů
- ostatní
 - statistiky
 - popis workflow pro predikci
 - RSS kanál s aktualizacemi (např. informace o automatických synchronizacích ze SILVy apod.)
 - ...

Projekt vyžaduje alespoň 5 studentů:

- Web - 2 studenti
- Databáze a datové nástroje - 1 student
- Predikce struktur rRNA a implementace bioinformatických nástrojů - 2 studenti

Projekt bude vznikat ve spolupráci s laboratoří bioinformatiky mikrobiologického ústavu AVČR.