



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Michal Hercík

Webový plugin pro vizualizaci sady sekundárních struktur RNA

Katedra softwarového inženýrství

Vedoucí bakalářské práce: doc. RNDr. David Hoksza, Ph.D.

Studijní program: Informatika

Studijní obor: Programování a vývoj software

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

TODO: podekovani

Název práce: Webový plugin pro vizualizaci sady sekundárních struktur RNA

Autor: Michal Hercík

Katedra: Katedra softwarového inženýrství

Vedoucí bakalářské práce: doc. RNDr. David Hoksza, Ph.D., Katedra softwarového inženýrství

Abstrakt: TODO

Klíčová slova: bioinformatika RNA sekundární struktura web plugin

Title: Web plugin for multiple RNA secondary structure visualization

Author: Michal Hercík

Department: Department of Software Engineering

Supervisor: doc. RNDr. David Hoksza, Ph.D., Department of Software Engineering

Abstract: TODO

Keywords: bioinformatics RNA secondary structure web plugin

Obsah

Úvod	2
1 Úvod do problematiky	3
1.1 Seznámení s biologickými pojmy	3
1.1.1 Nukleotid	3
1.1.2 Sekundární struktura RNA	3
1.2 Používané datové zdroje	3
1.2.1 Datové formáty	3
1.3 Seznámení s typy vizualizace sekundárních RNA struktur . . .	4
1.3.1 Standardní layout	4
1.3.2 Arc diagram	4
1.3.3 radial layout	4
1.4 Podobné projekty	4
1.4.1 VARNA	4
1.4.2 RNAstructViz	4
1.4.3 Forna	5
1.4.4 R-chie	5
1.5 Příbuzné projekty	5
1.5.1 TRAVeLer	5
1.5.2 R2DT	6
2 Metody vizualizace a porovnání	7
3 Programátorská dokumentace	8
3.1 Vstupní data	8
4 Uživatelská dokumentace	9
Závěr	10
Seznam použité literatury	11
Seznam obrázků	12
Seznam tabulek	13
Seznam použitých zkratk	14
A Přílohy	15
A.1 První příloha	15

Úvod

TODO: uvod

1. Úvod do problematiky

1.1 Seznámení s biologickými pojmy

1.1.1 Nukleotid

1.1.2 Sekundární struktura RNA

RNA neboli kyselina ribonukleová je nukleová kyselina, skládající se z řetězce tvořeného vzájemně propojenými nukleotidy. Ty jsou tvořeny z některé z bází adenin, uracil, guanin a cytosin. Je obvykle jednovláknová, někdy i dvouvláknová. Její přesná funkce se liší v závislosti na typu, ale slouží k přenosu genetické informace z jádra buňky na místo syntézy bílkovin, a také se procesu syntézy přímo účastní.

Primární struktura RNA je dána přesným pořadím nukleotidů v řetězci. Sekundární struktura RNA má podobu jednoduchého stočeného řetězce. V některých druzích RNA lze nalézt i zdvojené úseky řetězce. Terciální struktura RNA pak popisuje prostorové rozložení.

1.2 Používané datové zdroje

1.2.1 Datové formáty

JSON

JSON (javascriptový objektový zápis) je datový formát sloužící k ukládání dat organizovaných v polích nebo objektech. Navzdory názvu je na programovacím jazyce nezávislý. Skládá se z dvojice klíč – hodnota. Hodnota je libovolný podporovaný datový typ (např.: boolean, číslo, string, pole, objekt). Níže je ukázka JSON formátu.

```
{
  "basePairs": [
    {
      "basePairType": "canonical",
      "classes": [
        "bp-line"
      ],
      "residueIndex1": 2,
      "residueIndex2": 118
    }
  ]
}
```

1.3 Seznámení s typama vizualizace sekundárních RNA struktur

1.3.1 Standardní layout

1.3.2 Arc diagram

1.3.3 radial layout

1.4 Podobné projekty

Rádi bychom čtenáře seznámili s některými nástroji, které jsou používány pro vizualizaci sekundárních RNA struktur. Většina z nich je program a mohlo by se proto zdát zbytečné je zmiňovat nebo porovnávat s naší knihovnou. Nicméně u níže zmíněných programů není důležité řešení samotného grafického rozhraní, jako především druh zvolených metod pro vizualizaci a následné porovnávání.

Nejsou zde zmíněné všechny existující nástroje, ale byla snaha vybrat takové, které mají rozdílné přístupy a nabízí nejširší paletu funkcí.

1.4.1 VARNA

VARNA[2] (Visualization Applet for RNA) je nástroj pro automatické kreslení, vizualizaci a anotaci sekundárních RNA struktur, navržený jako doprovodný software pro webové servery a databáze.

VARNA implementuje čtyři kreslicí algoritmy, podporuje formáty dbn, ct, bpseq a RNAML pro vstup i výstup a je schopné exportovat kresbu do rastrových nebo vektorových formátů. Umožňuje ruční úpravy a strukturální anotace výsledku kresby a je považována za standard pro práci se sekundárníma RNA strukturama.

Kromě samotného faktu, že VARNA je program se liší od naší knihovny i v dalších bodech. Zatímco VARNA je rozumný nástroj pro anotaci a vykreslení výše zmíněných textových formátů různými způsoby, naše knihovna se snaží poskytnout pouze metody pro porovnávání sekundárních RNA struktur s již daným rozložením nukletidů v prostoru. VARNA navíc pro implementaci na web používá Java Applets ¹, které jsou od roku 2017 považované za zastaralé ², tím pádem je nelze považovat za vhodnou variantu pro použití na webu.

1.4.2 RNAStructViz

RNAStructViz[1] je grafický nástroj pro analýzu sekundárních RNA struktur. Jeho předností je vizuální porovnání tří konfigurací v kompaktním a standartizovaném **TODO: přeložit???** circular arc diagramu. Doplněné zabudovaným prohlížečem **TODO: přeložit???** CT-style souboru a prohlížeče **TODO: přeložit???** radial layout podstruktury, která je přímo propojená s **TODO: přeložit???** arc diagram oknem skrze nástroj pro výběr zoom. Mezi další funkce patří vypočítání číselných informací a možnost exportu obrázků a dat pro pozdější použití.

¹<https://docs.oracle.com/javase/tutorial/deployment/applet/index.html>

²<https://www.oracle.com/java/technologies/javase/9-deprecated-features.html>

Hlavním rozdílem v RNAStructViz kromě samotného faktu, že se jedná o program a ne knihovnu, vidíme zvolené metody porovnání. Zatímco naše knihovna používá standartní rozložení bez omezení na počet struktur, tak RNAStructViz používá **TODO: přeložit?** radial layout s maximální podporou zobrazení tří struktur najednou.

1.4.3 Forna

Forna[4] (force-directed rna) nabízí webové rozhraní a server, který umožňuje uživateli vložit sekundární RNA strukturu ve formátu dot-bracket a zobrazí ji jako force-directed graf. Uživatel může následně upravit pozice přetažením myši a lze i upravovat přímo strukturu. Forna umožňuje zobrazení libovolného množství struktur vedle sebe, ale nenabízí metody pro porovnávání mezi sebou.

1.4.4 R-chie

R-chie [5] je web server, který umí vygenerovat šest různých typů **TODO: přeložit?** arc diagramu. Vývoj tohoto nástroje byl se zaměřením především na složitější struktury, které nelze hezky nakreslit v rovinném diagramu. R-chie umí vygenerovat diagram pro porovnávání dvou sekundárních RNA struktur. Důležitým cílem bylo možnost generovat diagramy pro velké množství dat, proto také nenabízí grafické rozhraní a s ním spojenou interakci se strukturama.

Kromě metody vizualizace se liší od naší knihovny právě ve zmíněné snaze zpracovat velké množství dat bez interakce, zatímco naše knihovna se soustředí na interakci se sekundárníma RNA strukturama. Projekt také nabízí balíček napsaný v jazyce R³ zvaný R4RNA, který umožňuje spuštění programu lokálně a napříč operačním systémem.

1.5 Příbuzné projekty

Níže jsou zmíněné dva projekty, které úzce souvisí s naší knihovnou, protože produkují vstupní data ve formátu, se kterým pracuje naše knihovna a metody použité v generování takových dat jsou klíčové pro naši knihovnu.

1.5.1 TRAVeLer

Traveler[3] je nástroj pro vizualizaci cílové sekundární struktury, využívající existující rozložení dostatečně podobné RNA struktury jako vzor. Traveler je založený na algoritmu, který konvertuje cílovou a vzorovou strukturu do odpovídající stromové reprezentace a využije stromovou editační vzdálenost společně s modifikací rozložení k přetvoření vzorové struktury do cílové. Traveler přijme na vstupu sekundární strukturu a vzor rozložení a na výstupu dá rozložení cílové struktury. Je to tedy command-line open source nástroj schopný rychle generovat rozložení i pro největší RNA struktury za poskytnutí dostatečně podobného rozložení.

³<https://www.r-project.org/>

1.5.2 R2DT

R2DT[6] je metoda pro predikci a vizualizaci široké škály sekundárních RNA struktur ve standardním rozložení. R2DT je postaveno na knihovně se 3 647 vzory reprezentujícími většinu známých RNA struktur. R2DT se používá na ncRNA sekvence z RNACentral⁴ databáze a vytvořil více než 13 miliónů diagramů, čímž tvoří největší světovou sadu dat s 2D RNA strukturami. Pro vizualizaci neboli 2D rozložení používá právě výše zmíněný nástroj Traveler.

⁴<https://rnacentral.org/>

2. Metody vizualizace a porovnání

3. Programátorská dokumentace

3.1 Vstupní data

4. Uživatelská dokumentace

Závěr

Seznam použité literatury

- [1] CHENNEY, S., HEITSCH, C., MIZE, C., SWENSON, S., SCHMIDT, M. D., KIRKPATRICK, A. a YOON, I. (2019). Rnastructviz. URL <https://github.com/gtDMMB/RNAStructViz/wiki>. Accessed on March 29, 2023.
- [2] DARTY, K., DENISE, A. a PONTY, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**(15), 1974–1975.
- [3] ELIAS, R. a HOKSZA, D. (2017). Traveler: a tool for template-based rna secondary structure visualization. *BMC Bioinformatics*, **18**(1), 487. ISSN 1471-2105. doi: 10.1186/s12859-017-1885-4. URL <https://doi.org/10.1186/s12859-017-1885-4>.
- [4] KERPEDJIEV, P., HAMMER, S. a HOFACKER, I. L. (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**(20), 3377–3379. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv372. URL <https://doi.org/10.1093/bioinformatics/btv372>.
- [5] LAI, D., PROCTOR, J. R., ZHU, J. Y. A. a MEYER, I. M. (2012). R-chie : a web server and R package for visualizing RNA secondary structures . *Nucleic Acids Research*, **40**(12), e95–e95. ISSN 0305-1048. doi: 10.1093/nar/gks241. URL <https://doi.org/10.1093/nar/gks241>.
- [6] SWEENEY, B. A., HOKSZA, D., NAWROCKI, E. P., RIBAS, C. E., MADEIRA, F., CANNONE, J. J., GUTELL, R., MADDALA, A., MEADE, C. D., WILLIAMS, L. D., PETROV, A. S., CHAN, P. P., LOWE, T. M., FINN, R. D. a PETROV, A. I. (2021). R2dt is a framework for predicting and visualising rna secondary structure using templates. *Nature Communications*, **12**(1), 3494. ISSN 2041-1723. doi: 10.1038/s41467-021-23555-5. URL <https://doi.org/10.1038/s41467-021-23555-5>.

Seznam obrázků

Seznam tabulek

Seznam použitých zkratek

A. Přílohy

A.1 První příloha