



Um programa para análise da anotação de RNAs não-codificantes obtida via INFERNAL

A program for analyzing the annotation of non-coding RNAs obtained via INFERNAL

Daniel Amaral Augusto (orientado)*, Alexandre Rossi Paschoal (orientador)†

RESUMO

Os RNAs não-codificantes (ncRNAs) são ácidos ribonucleico (RNAs), o qual são transcritos, mas não traduzidos em proteínas. No caso, o programa INFERNAL é uma ferramenta considerado estado-da-arte para identificar ncRNAs. Entretanto, apenas uma saída texto é fornecida como resultado, tendo a necessidade de programas que permitem realizar um pós-análise. Nesse cenário, o desenvolvimento de uma aplicação web para análise de dados do INFERNAL é um hiato a ser preenchido. Por exemplo o usuário passa entrada de dados em um ou mais arquivos para o sistema web que, de forma clara e organizada, possibilitará ao usuário uma análise da utilização dessa massa de dados fornecida. Assim, este projeto teve como objetivo de desenvolver uma ferramenta web, fácil e amigável ao usuário final, de modo a contribuir para análise dos dados de saída do o programa INFERNAL. Do ponto de vista técnico, o objetivo foi buscar usar o que tinha de mais atual, baseadas em tecnologias de web, como *JavaScript* da parte voltado no *back-end* e *front-end* da aplicação de modo que os dados fossem organizados em uma tabela dinâmica. No final, como contribuição o usuário tem: sistema que permite analisar, de forma tabular, os dados de localização dos ncRNAs identificados via INFERNAL.

Palavras-chave: dados biológicos, ferramentas *web*, RNA, estruturação, banco de dados.

ABSTRACT

Non-coding RNAs (ncRNAs) are ribonucleic acids (RNAs) which are transcribed but not translated into proteins. In this case, the INFERNAL program is a state-of-the-art tool to identify ncRNAs. However, only text output is provided as a result, necessitating programs that allow you to perform a post-analysis. In this scenario, the development of a web application for analyzing INFERNAL data is a gap to be filled. For example, the user passes data entry in one or more files to the web system which, in a clear and organized way, will enable the user to analyze the use of this mass of data provided. Thus, this project aimed to develop a web tool, easy and friendly to the end user, in order to contribute to the analysis of the output data of the INFERNAL program. From a technical point of view, the objective was to seek to use the most current, based on web technologies, such as JavaScript from the back-end and front-end part of the application so that the data was organized in a pivot table. In the end, as a contribution, the user has: a system that allows analyzing, in a tabular form, the location data of the ncRNAs identified via INFERNAL.

Keywords: biological data, *web* tools, RNA, structuring, database.

* Engenharia da Computação, Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil; daugusto@alunos.utfpr.edu.br

† Universidade Tecnológica Federal do Paraná, Campus Cornélio Procópio; paschoal@utfpr.edu.br



1 INTRODUÇÃO

O número de pessoas conectadas a Internet e utilizando aplicações *web* presentes no mercado, estão em uma crescente ao longo dos anos, segundo o projeto *Internet Live Stats* (2021), cerca de 5 bilhões de pessoas são usuárias da Internet e existem quase 2 bilhões de *websites* online.

As aplicações *web* modernas utilizam, principalmente, a linguagem *JavaScript* como base, juntamente com tecnologias de bibliotecas e *frameworks*, que têm como objetivo auxiliar na criação de aplicações melhores, com mais funcionalidades e intuitivas aos usuários.

Segundo Arora (2021), dentre as ferramentas existentes atualmente, as mais conhecidas e utilizadas são, o *Angular*, *ReactJS* e *VueJS*. Com cada ferramenta possuindo uma característica específica, que as difere das demais, cabe ao desenvolvedor responsável escolher, em meio as tecnologias disponíveis, a que melhor se adequa ao que será desenvolvido.

Dentre as ferramentas tidas como as mais utilizadas, citadas anteriormente, de acordo com Buna (2017), a principal entre os desenvolvedores por conta de seu desempenho, além da sua baixa curva de aprendizado, em relação as outras, é o *ReactJS*. Essa biblioteca *JavaScript* é utilizada para criação de *User Interfaces (UI)*, trabalhando diretamente com o *Document Object Model (DOM)*.

Além disso, a ferramenta possui componentes estruturados de forma independente, onde cada componente possui suas características específicas e pode ser reutilizado em diversas partes da aplicação, sem que seja necessário recriá-lo. Assim, com essa ferramenta cada componente renderizado faz com que o DOM seja atualizado com as mudanças realizadas, de maneira eficiente.

A visualização de todo conteúdo compartilhado, através de *websites* e aplicações disponíveis, deve ser feita de maneira organizada e centralizada, mesmo com uma grande quantidade de dados, através de técnicas e ferramentas que facilitem essa organização, desde a coleta ao processamento desses dados, a serem exibidos ao usuário. De acordo com Freitas e Wagner (1993) a representação em tabelas é utilizada para demonstrar em um espaço a dependência e correlação entre os dados apresentados.

A massiva quantidade de informações obtidas através de bancos de dados dedicados, deve facilitar a identificação de sequências de *RNAs* estruturais, assim como, genes de *RNAs* não codificantes (*ncRNAs*). Como base de dados para coleções abrangentes de *ncRNAs*, tem-se o *Rfam*, para reconhecimento e classificação de famílias de sequência de RNAs (GRIFFITHS-JONES et. al, 2003). Em conjunto, o *software Inference of RNA Alignment* (NAWROCKI e EDDY, 2013), permite a pesquisa de qualquer família de RNA em uma sequência ou genoma, utilizando o conceito de busca por estruturas secundárias de RNA, via modelos de covariância (NAWROCKI e EDDY, 2013).

Com isso, utilizando a base de dados *Rfam* e o *software* INFERNAL tem-se o agrupamento e a possibilidade de pesquisa de qualquer família de RNA em um genoma, ou sequência, específica. Sendo assim, existe a possibilidade de agrupamento de dados biológicos, apresentadas na aplicação através do *upload* de um arquivo contendo esses dados, por parte do usuário, no intuito de se realizar uma análise da anotação de *ncRNAs*.

Como pode ser desenvolvida uma aplicação *web* para a organização e análise de uma grande quantidade de dados biológicos?

No desenvolvimento de um sistema *web* com a finalidade de organizar e proporcionar a análise desse tipo de dado, com tantas especificidades, deve-se levar em consideração a funcionalidade para os usuários, haja vista que bancos de dados com informações diversas e numerosas sobre um mesmo aspecto podem gerar análises equivocadas, caso não estejam organizados de maneira a facilitar tanto a visualização, quanto o estudo e inferência dos dados obtidos.



Para Gershon et. al. (1998) o desenvolvimento de aplicações *web*, com o intuito de transmitir informações e grandes quantidades de dados ao usuário, deve possuir uma interface intuitiva e de fácil uso, levando em conta os usuários, e se esses se adaptam a solução obtida.

O sistema desenvolvido possui como objetivo facilitar a visualização dos dados biológicos, particularmente, dedicado a anotação de ncRNAs, em larga escala. Com isso, por meio de tabelas, o usuário é capaz de realizar a devida análise do que é apresentado.

Desse modo, é desenvolvido o presente sistema *web*, levando consigo o nome de *BioData*, e com finalidade de atender, principalmente, a área de Bioinformática.

2 MÉTODO

De acordo com Munari (2008), um mesmo projeto pode ser desenvolvido de diversas formas, podendo ser ajustado e realizado de acordo com as necessidades apresentadas. No projeto, a necessidade se relaciona com o conhecimento das técnicas e tecnologias web e como emprega-las em uma arquitetura funcional.

Portanto, essa seção é dedicada na apresentação da arquitetura do projeto e seu desenvolvimento.

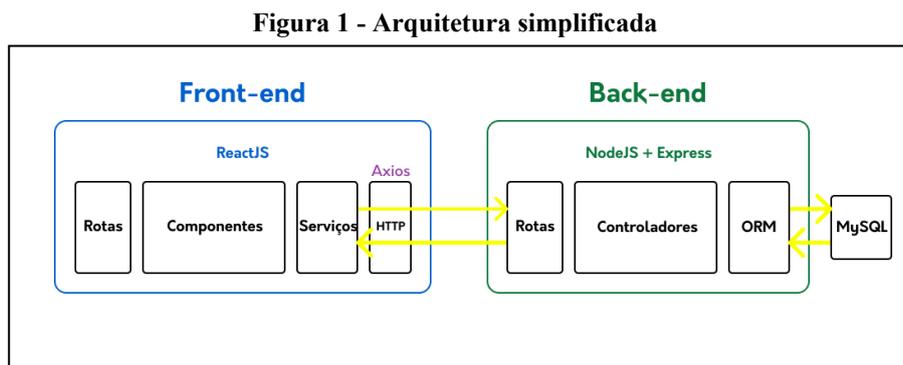
2.1 Arquitetura

A arquitetura proposta no projeto está dividida entre o back e o front-end. Cada tecnologia e ferramenta empregada no desenvolvimento da aplicação possuem grande importância em seu funcionamento.

Em uma breve explicação, o uso do NodeJS com Express no back-end exporta as APIs REST que interagem com o MySQL usando o Sequelize ORM.

No front-end o cliente ReactJS envia requisições HTTP e recebe respostas utilizando o Axios, consumindo-as nos componentes. As rotas são utilizadas para a navegação das páginas web.

A Figura 1 demonstra a arquitetura de forma simplificada.



Fonte: Autoria própria (2021).

2.1.1 Back-end

Para se iniciar o desenvolvimento do projeto, foram utilizadas, no *back-end*, as seguintes tecnologias e ferramentas:

1. NodeJS: Tecnologia capaz de executar código *JavaScript* fora do navegador.
2. ExpressJS: *Framework* que fornece os recursos mínimos para a construção de servidores web.
3. MySQL: Banco de dados relacional utilizado para o armazenamento dos dados.

Além das ferramentas apresentadas, foi utilizado um controlador de pacotes, ou *Package Managers*. O controlador de pacotes utilizado no projeto foi o NPM (NPM, 2009). Um arquivo JSON descreve as dependências necessárias para que a aplicação funcione de forma correta.

A Figura 2 demonstra a configuração do módulo *package.json*, com as dependências e versões utilizadas.

Figura 2 - Configuração do módulo *package.json*

Fonte: Autoria própria (2021).

2.1.2 Front-end

Na realização do desenvolvimento do *front-end* foram escolhidas as seguintes tecnologias:

1. ReactJS: Biblioteca JavaScript para desenvolvimento *front-end* na criação de aplicações *web* modernas.
2. Bootstrap: Biblioteca contendo componentes *web* pré-definidos.
3. Axios: Ferramenta para a realização de requisições HTTP.

Assim como o *bootstrap*, o *ReactJS* tem a função de organizar o conteúdo que é mostrado ao usuário, ou seja, na criação de interfaces visuais.

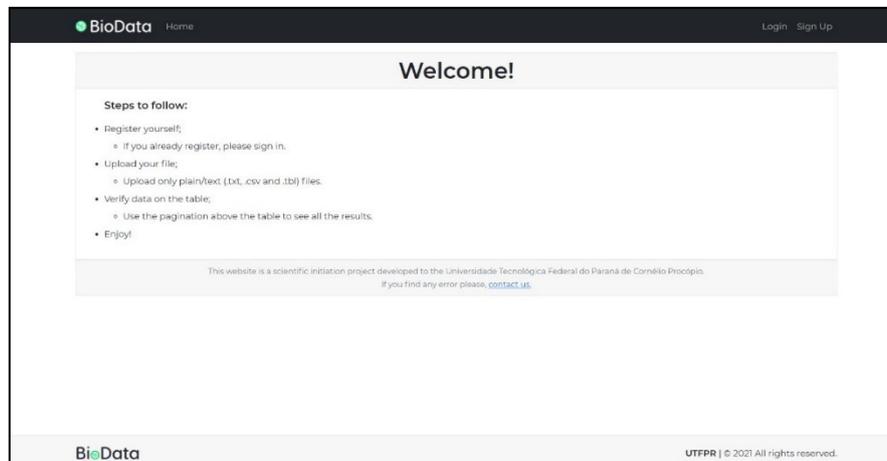
A criação e desenvolvimento do *front-end* foi realizada a partir do momento em que o *back-end* estava finalizado.

3 RESULTADOS

Essa seção apresenta os resultados mais relevantes obtidos ao longo do desenvolvimento do projeto. Somente após a finalização do back e do front-end foi possível obter os resultados esperados. Para isso, pode-se visualizar na Fig. 3, a tela principal da aplicação, nela há elementos explicativos que auxiliam na utilização do sistema.



Figura 3 - Tela de início



Fonte: Autoria própria (2021).

Após o registro do usuário, a tela de *upload* é apresentada, com isso, o usuário registrado é capaz de fazer *upload*, de um arquivo por vez, contendo os dados biológicos a serem analisados.

Após o *upload* do arquivo, os dados são apresentados em uma tabela, como pode ser visto na Fig. 4.

Por conta da grande quantidade de registros, é necessário que cada paginação gere uma nova requisição dos dados, assim, o carregamento desses dados para dentro da tabela se torna mais eficiente.

Figura 4 - Tela de dados

#	Target Name	Accession Name	Query	Accession	Mdl	Mdl From	Mdl To	Seq From	Seq To	Strand	Trunc	Pass	GC	Bias	Score	eValue	Inc	Description of Target
1	PfaliknerVC_selvet31_2686	-	tRNA	RF00005	cm	1	71	354	282	-	no	1	0.38	0	38.6	0.000047	1	-
2	PfaliknerVC_trinity_5754	-	tRNA	RF00005	cm	1	71	1165	1236	+	no	1	0.46	0	38.3	0.000058	1	-
3	PfaliknerVC_spades_kmer75_13335	-	tRNA	RF00005	cm	1	71	814	885	+	no	1	0.46	0	38.3	0.000058	1	-
4	PfaliknerVC_spades_default_5855	-	tRNA	RF00005	cm	1	71	1546	1617	+	no	1	0.46	0	38.3	0.000058	1	-
5	PfaliknerVC_trinity_36338	-	tRNA	RF00005	cm	1	71	2656	2725	+	no	1	0.4	0	35.5	0.00037	1	-

Fonte: Autoria própria (2021).

A tabela fica disposta abaixo de um sistema de busca, assim, o usuário pode realizar uma pesquisa pelo *Target Name*, do *ncRNA*, podendo, se necessário, fazer a devida análise sobre esses dados.

4 CONCLUSÃO

Este artigo apresentou um programa web para contribuir na análise da anotação de *ncRNAs* a partir da saída do programar INFERNAL. O programa INFERNAL busca identificar *ncRNAs* via modelos de



covariância por meio da estrutura secundária de RNA. Após a sua execução, este ainda fornece um arquivo texto com os resultados, o que o usuário final terá que manipular este arquivo para fazer sua análise. Nesse sentido que este projeto contribuiu e foi desenvolvido, o de criar um programa web que recebe uma ou mais entradas obtidas do programa INFERNAL, e permite a tabular e analisar os resultados.

Deste modo, este projeto apresenta como conclusões finais: (i), ao usuário final, um sistema simples e prático que permite analisar os dados de maneira clara e objetiva na web; (ii) o programa permite ainda a exportação dos dados e a verificação dos tipos de RNAs contido nele.

Por fim, este projeto é um esboço inicial que, a partir desta versão 1, quer-se aprimorar e contribuir futuramente para criar relatório de visualização gráfica destes dados, e análise comparativa contra outras espécies. Isso já está programa a ser feito na continuidade deste projeto.

AGRADECIMENTOS

Agradeço ao CNPq e ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC). Ao professor Alexandre Rossi Paschoal que me orientou durante esse projeto e a Universidade Tecnológica Federal do Paraná de Cornélio Procópio.

REFERÊNCIAS

ARORA, Simran. **10 Best JavaScript frameworks**. Disponível em: <<https://hackr.io/blog/best-javascript-frameworks>>. Acesso em: 5 de jun. 2021.

BUNA, S. **Yes, React is taking over front-end development**. Medium, 2017. Disponível em: <https://medium.freecodecamp.org/yes-react-is-taking-over-front-end-development-the-question-is-why-40837af8ab76>. Acesso em: 10 jun. 2021.

FREITAS, Carla Maria dal Sasso; WAGNER, Flavio Rech. **A Methodology for Selecting Visual Representation in Scientific and Simulation Applications**. In: SIBGRAPI, 6., 1993, Porto Alegre. Anais Porto Alegre: Sibgrapi, 1993. p. 89 - 97. Disponível em: <<http://sibgrapi.sid.inpe.br/rep/sid.inpe.br/sibgrapi/2012/12.11.15.18>>. Acesso em: 5 jun. 2021.

GERSHON, Nahum; EICK, Stephen G.; CARD, Stuart. **Design: Information Visualization**. Interactions, New York, v. 2, n. 5, p.9-15, abr. 1998.

_____. **INTERNETLIVESTATS**. Disponível em: <<http://www.internetlivestats.com/>>. Acesso em: 09 jun. 2021.

MUNARI, Bruno. **Das coisas nascem coisas**. 2 ed. São Paulo: Martins Fontes, 2008. 378p.
NAWROCKI, Eric; EDDY, Sean. **Infernal 1.1: 100-fold faster RNA homology searches**. Bioinformatics 29:2933-2935, 2013.

NPM. **NPM - Node Package Manager**. 2009. Disponível em: <<https://docs.npmjs.com/>>. Acesso em: 11 jun. 2021.

GRIFFITHS-JONES, Sam; BATEMAN, Alex; MARSHALL, et al; **Rfam: an RNA family database, Nucleic Acids Research**, Volume 31, Issue 1, 1 January 2003, Pages 439–441.