
ANÁLISES DE FERRAMENTAS DA BIOINFORMÁTICA APLICADAS EM ESTUDOS GENÉTICOS

NASCIMENTO, Rafael Franco
FILHO, Rodrigo Santiago Claro
PERUCCI, Camilo Cesar

Centro Universitário Hermínio Ometto – UNIARARAS, Araras – SP, Brasil

Resumo

O avanço tecnológico, sempre ininterrupto, vem gerando amplas mudanças na sociedade científica, criando novas áreas de conhecimento e mesclando outras já existentes. Com o intuito de utilizar as novas tecnologias e integra-las à biologia em 1990 pesquisadores de diversos países iniciaram o “Projeto Genoma”. Esse projeto mapeou os dados do genoma humano e diversos dados foram obtidos e armazenados. Surgiu-se então a necessidade de obter esses dados e transformá-los em informações que possam ser utilizadas pelos pesquisadores. Sendo este o papel do bioinformata. Com o progresso da tecnologia os bioinformatas realizaram a obtenção de uma grande quantidade de dados e ferramentas de análises. Por ser uma área de conhecimentos compartilhados, pode-se haver dificuldades no entendimento da utilização destas ferramentas quando não intuitivas. O objetivo desta pesquisa é analisar e desenvolver um guia de plataformas de análise, exemplificando e detalhando os processos para a obtenção dos resultados de sequenciamento genético, utilizando tecnologias da informática, buscando facilitar a integração das áreas envolvidas. Para isso foi desenvolvido uma plataforma web local, onde é possível realizar análises iniciais de Bioinformática, durante o desenvolvimento da plataforma o foco foi realizar análises precisas e criar uma interface simples para o usuário. Juntamente com a plataforma foi desenvolvido um guia informativo de todos os passos realizados neste trabalho de conclusão de curso. Com a conclusão do projeto é esperado deixar um trabalho que possa ser continuado por outros alunos da instituição.

Palavras chave: DNA, RNA, análise gênica, Biopython

1 Introdução

A bioinformática é um tema emergente que visa a utilização de ferramentas desenvolvidas pela área da computação para analisar e apresentar dados biológicos, surgiu de forma gradual com o avanço tecnológico. Integrado com conhecimento da Biologia e as tecnologias da informática, pesquisadores tiveram suas análises facilitadas por ferramentas que conseguem realizar análises de forma rápida, assim gerando o momento atual da Bioinformática, onde existem muitos dados analisados, mas poucos foram estudados e interpretados.

A utilização de ferramentas agilizou o processo de análise, porém ainda existe uma dificuldade no entendimento das ferramentas e como elas funcionam, por serem muito novas e necessitarem que pesquisadores de áreas de conhecimento distintas da tecnologia as utilizem, dificultando a criação de ferramentas úteis para novas linhas de pesquisa. Este projeto apresenta o estudo de ferramentas, revisões de literaturas e estudo de casos, para o desenvolvimento de um ambiente que possibilite a utilização destas ferramentas aplicadas à Bioinformática.

A obtenção de conhecimento sempre foi o objetivo principal de pesquisadores e desenvolvedores, no entanto, obter informações em grande quantidade é trabalhoso. Com avanço da tecnologia e o surgimento de ferramentas mais avançadas no estudo de dados, tornou-se possível o manuseio de informação em grande escala. Na área da Bioinformática atualmente é disponibilizado uma grande quantidade de base de dados genéticos para serem analisados, porém existe uma dificuldade dos pesquisadores de biologia em entender o funcionamento de certas ferramentas. O objetivo deste projeto é integrar o curso de Sistemas de Informação a área de biológicas, buscando unir estes conhecimentos para gerar análises concisas e objetivas, facilitando os estudos dos resultados destas análises.

Como objetivo deste projeto é apresentado a utilização de ferramentas e métodos de sequenciamento genético aplicados à Bioinformática no desenvolvimento de um sistema web local.

Objetivos Específicos

- Realizar pesquisas na área de Bioinformática.
- Utilizar bases de dados de Bioinformática.
- Implementar as ferramentas de Bioinformática ao ambiente.
- Validar os resultados obtidos.
- Criar guia de utilização das ferramentas utilizadas.

2 Revisão Bibliográfica

A revisão bibliográfica deste trabalho apresenta conceitos importantes da Biologia, criando uma base que facilite a integração de conhecimentos.

DNA e RNA:

DNA e RNA são moléculas extremamente importantes por estarem entre as mais pesquisadas, são denominadas pelos pesquisadores como dogma central. Estudando essas moléculas é possível reconhecer problemas que ocorrem no corpo humano. Para essa pesquisa é importante saber que o DNA é transcrito para um RNA, nessa transcrição acontecem mudanças que necessitam de adaptações para análises do DNA ou RNA. Após o RNA também existe a tradução para proteínas, as proteínas são utilizadas pela célula para diversas finalidades como, composição de anticorpos e transporte de oxigênio. Esses três elementos e as duas fases, transcrição e tradução, são importantes como bases para saber onde os pesquisadores irão utilizar as análises.

Na Bioinformática são utilizados sequenciamentos para estudar o DNA e RNA. Um sequenciamento é uma série de processos bioquímicos, tendo como finalidade determinar a ordem dos nucleotídeos (adenina, guanina, citosina e timina) de determinada amostra. Esse sequenciamento pode ser realizado com o DNA e também com o RNA, substituindo a timina por uracila quando o RNA está sendo sequenciado (TORTORA 2012).

Microarray:

Uma das técnicas utilizadas para gerar dados a partir da amostra gênica é a técnica microarray, que consiste em organizar os dados em um chip de DNA deixando os dados prontos para a análise, tem uma facilidade de lidar com uma grande quantidade de dados simultaneamente e em várias condições diferentes, utiliza vários métodos matemáticos para gerar expressões gênicas, realizando hibridização entre as amostras para com marcadores gerar uma matriz com os genes, possibilitando mostrar quais são os genes identificados com as características desejadas, essa matriz é utilizada para gerar os dados finais em heatmaps, volcano plots, e outros gráficos para visualização de grande quantidade de dados. Para armazenar a grande quantidade de dados sequenciais foram utilizadas bases de dados públicas, que tem extrema importância no meio científico, já que a mesma armazena dados de diversos organismos e linhas de pesquisas. Dentre as mais conhecidas, destacam-se neste trabalho: PDB, NCBI e UNIPROT (REIS 2015).

Ferramentas da Bioinformática:

Durante a realização desse trabalho de conclusão de curso serão utilizadas algumas ferramentas que são empregadas na área da Bioinformática, as principais são Biopython e Blast e pipelines.

O Biopython é um projeto de associações que disponibiliza ferramentas na linguagem de programação python, essa é uma linguagem orientada a objetos de alto nível, gratuita e portátil. Existem funcionalidades úteis para facilitar o estudo e análises de bioinformática, uma função útil é a tradução de DNA para RNA e a transcrição para proteínas, também são disponibilizadas classes que trabalham com alinhamento. Como o Biopython tem uma gama de funcionalidades e formas de utilização o projeto possui um tutorial on-line para facilitar o entendimento das ferramentas disponíveis (SANTOS 2010).

O Blast é um programa open source, o que o torna um programa de fácil acesso para diferentes máquinas e sistemas operacionais. A função principal do Blast é buscar sequências semelhantes em diversas bases de dados. Essa ferramenta utiliza uma rotina de entrada da sequência estudada, após escolher algumas especificações como base de dados e tipo de arquivo de entrada, o Blast realiza a pesquisa de sequências semelhantes a inserida inicialmente, então apresenta a lista de sequências encontradas e dados de análise como taxa de aparição da sequência base na sequência encontrada (SANTOS 2010).

Pipeline é um conjunto de ferramentas que se relacionam entre si. Simplificando o seu contexto, o pipeline neste projeto é o processo seguido para se obter a modelagem de uma proteína 3D. Neste sentido, o código trabalha em conjunto com softwares integrados da bioinformática, executando uma “linha de tarefas” em sequência para obter o resultado (LOPES 2018).

Trabalhos Relacionados

Yuri Ferreti (2015) apresenta a implementação de uma ferramenta de análise de dados, IPTrans A2Tool, com o objetivo de testar a eficiência dessa ferramenta, os dados utilizados nos testes de eficiência foram retirados de bases de dados públicos como o TCGA. Após os testes foi constatado que a ferramenta é eficiente em suas análises, foi utilizado o método de comparação com outras ferramentas que proporcionam funcionalidades semelhantes ao IPTrans A2Tool, que também foi testada em usabilidade e apresentou grande utilidade nesse recurso, permitindo que uma proporção maior de pessoas consigam utilizá-la para suas análises e uma facilidade na integração das áreas de pesquisas da Bioinformática.

Katia Cristiane Brumatti Gonçalves (2015) apresenta um trabalho onde é utilizado várias ferramentas de Bioinformática para demonstrar suas eficiências e utilidades para os pesquisadores da Bioinformática, apresenta comparações com utilização de diferentes pipelines afirmando melhor apresentação dos dados pela ferramenta MISO comparado com o Cuffdiff, esse trabalho exemplifica as melhores formas de apresentação de resultados apresentando uma partição de controle de qualidade nas análises apresentadas. Também confirma a utilidade das ferramentas no auxílio dos pesquisadores e o excesso de dados que existem para serem analisados.

Rogério F. de Souza (2015) apresenta um guia para Análise de expressão diferencial em transcriptomas. Neste guia são apresentadas diversas ferramentas como Galaxy para coleta de dados e programas de análise online, Ferramenta de Busca por alinhamento local (BLAST), programação em Shell entre outros.

3 Metodologia

3.1 Materiais:

Durante esse trabalho foi realizada a implementação das ferramentas de Bioinformática em um ambiente web local, utilizando a linguagem de programação Python, juntamente com o Biopython. Essa biblioteca fornece funções pré-definidas para funções de análises, como exemplo a utilização em alinhamentos, comparando duas ou mais sequências de RNAs e apresentando seus respectivos dados estatísticos.

A coleta de dados foi realizada por meio de bancos de dados públicos como UniProt, PDB e NCBI, que disponibilizam dados públicos em seus próprios sites.

3.2 Métodos:

O projeto foi subdividido em etapas, sendo realizadas na seguinte ordem:

A primeira etapa foi constituída de realizar uma revisão de literatura buscando compreender as análises geradas pelas ferramentas de Bioinformática, foi essencial para entender a utilização dos dados obtidos através das análises, assim otimizando o processo de utilização das ferramentas, podendo descartar dados desnecessários para o resultado final. A revisão de literatura também é necessária para entender os cálculos aplicados, processamentos e saídas de dados que correspondam às entradas, sendo essencial para realização das análises.

Na segunda etapa foi necessário definir um foco de quais moléculas utilizar para realizar as análises. Por existirem especificidades em cada molécula é importante definir qual delas é o foco da pesquisa, nesse trabalho o foco é analisar RNA e proteínas, por serem moléculas que já existem estudos realizados e bem documentados.

Para realizar as análises foi utilizada a proteína beta-glucosidase, relacionada a funções como resposta imune, enovelamento e reconhecimento intercelular. Outra proteína utilizada foi a beta-glucanase, relacionada a glicose e produção de energia da célula.

Na terceira etapa do processo com a linha de pesquisa definida o próximo passo a se realizar é a definição de quais base de dados utilizar e quais moléculas analisar. Neste trabalho foram utilizados dados na extensão de arquivo “.fasta” retirados das base PDB(Protein Data Bank).

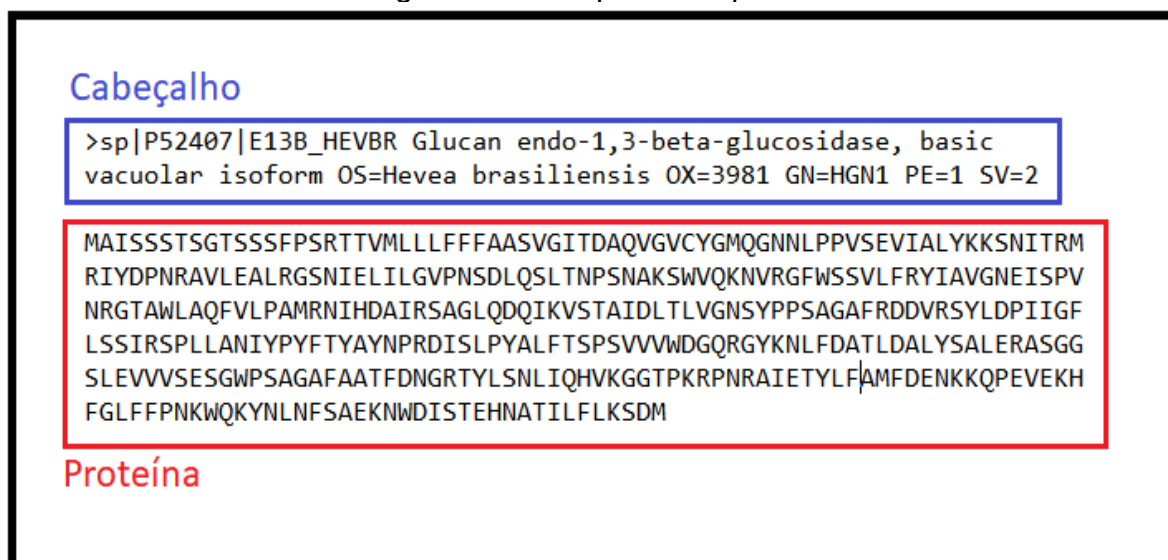
Arquivos “.fasta” são arquivos utilizados para guardar dados científicos de ácidos nucleicos ou proteínas, um “.fasta” é composto por um cabeçalho, que nomeia e identifica a proteína, seguido de uma sequência de letras que representam a proteína.

Na quarta etapa se inicia a fase de utilização das ferramentas da Bioinformática, onde foram extraídas as informações dos arquivos formato “.fasta”. Nesta etapa foram utilizadas práticas, conceitos e ferramentas da bioinformática para transformar dados em informações. As informações a serem geradas estão relacionadas às características de determinada molécula. A partir do seu arquivo “.fasta” respectivo, foram abstraídas informações como a porcentagem de aparição de códons, por meio da leitura de todo o arquivo. Os resultados das aparições foram apresentados em gráfico de cores. Também é apresentada a possibilidade de Tradução, ou seja, converter o RNA para Proteína e Transcrição converter da Proteína para o RNA.

Há possibilidade do usuário gerar uma comparação entre 2 moléculas de proteína. A comparação é feita a nível de aparição por códons, identificando a taxa de aparição de todos códons das proteínas inseridas.

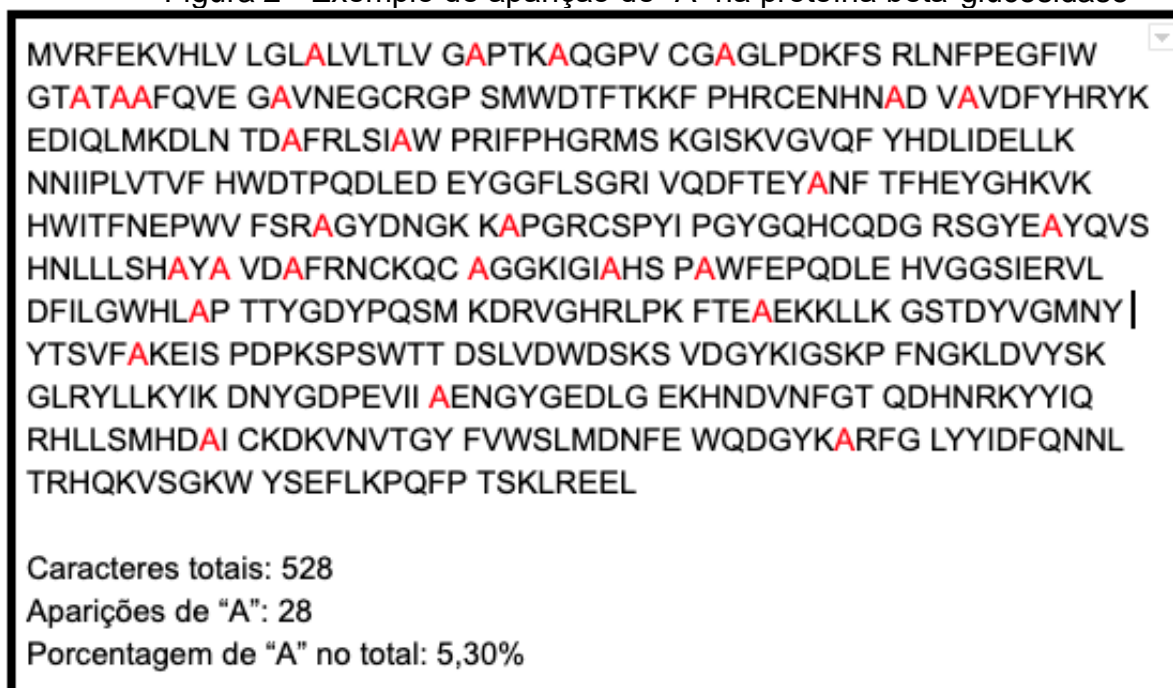
Considerando esse exemplo de comparação de 1 códon “A” em duas proteínas beta-glucosidase e acetate-kinase fornecidas pela base do Uniprot. A figura 1 e a figura 2 apresentam exemplos de aparição do códon “A” nas respectivas proteínas.

Figura 1 - Exemplo de arquivos ".fasta"



fonte: autoria própria

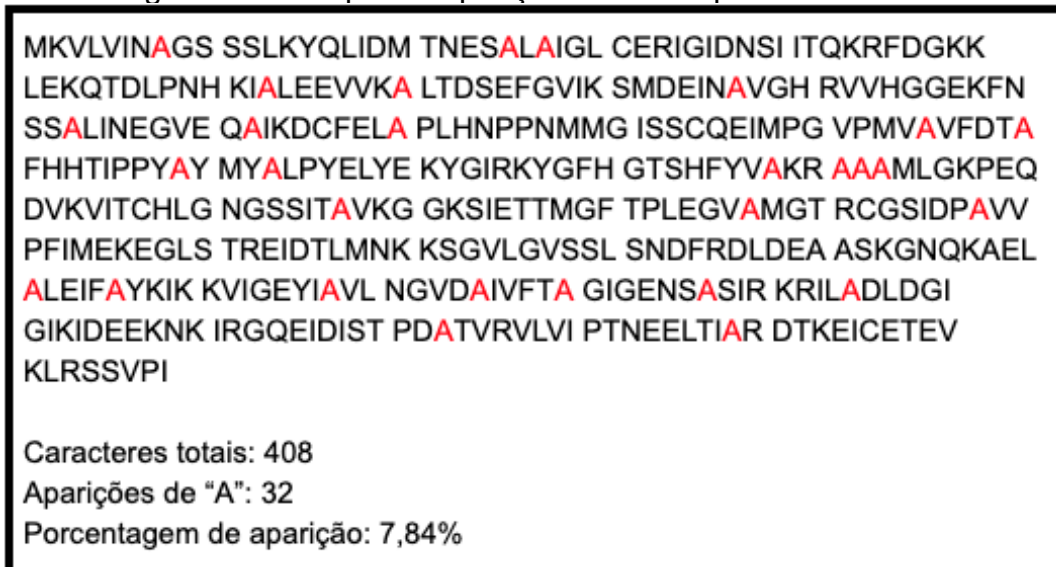
Figura 2 - Exemplo de aparição de "A" na proteína beta-glucosidase



fonte: Autoria própria

A figura 3 apresenta um exemplo de porcentagem de aparição de 1 códon "A" individual em cada sequência de proteína, neste trabalho será realizado esta comparação para todos os códons existentes.

Figura 3 - Exemplo de aparição de "A" na proteína acetate kinase



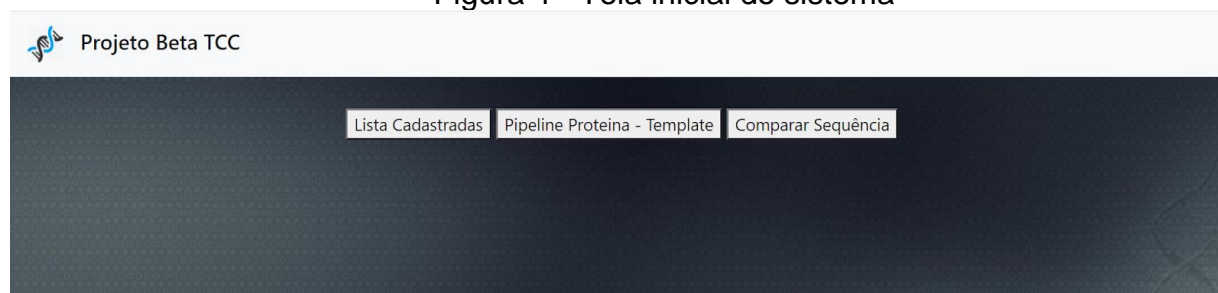
fonte: Autoria própria

A quinta etapa com resultados obtidos foi necessário implementar elementos gráfico, para visualizar os resultados. Para facilitar a integração das áreas de pesquisas foi criada uma aplicação web, utilizando o framework Django, onde o usuário poderá visualizar as sequências analisadas, também poderá escolher duas sequências para serem comparadas, o que irá retornar a semelhança entre elas, também poderá inserir sequências novas para serem visualizadas e comparadas. Juntamente com essas apresentações o sistema web implementa uma modelagem 3D para visualização de proteínas, para realizar a apresentação de um sistema 3D foram utilizadas as ferramentas Modeller e ClustalW. O ClustalW realiza o alinhamento múltiplo de sequências. O Modeller utiliza uma sequência de proteínas e um modelo 3D do qual servirá de base para a criação do modelo referente a sequência inserida.

4 Resultados

Como resultado deste projeto é apresentado um sistema de web local, onde o usuário inicialmente pode selecionar três botões "Listas cadastradas", "Pipeline Proteína - Template", "Comparar Sequência". Como é apresentado na figura 4.

Figura 4 - Tela inicial do sistema



fonte: autoria própria

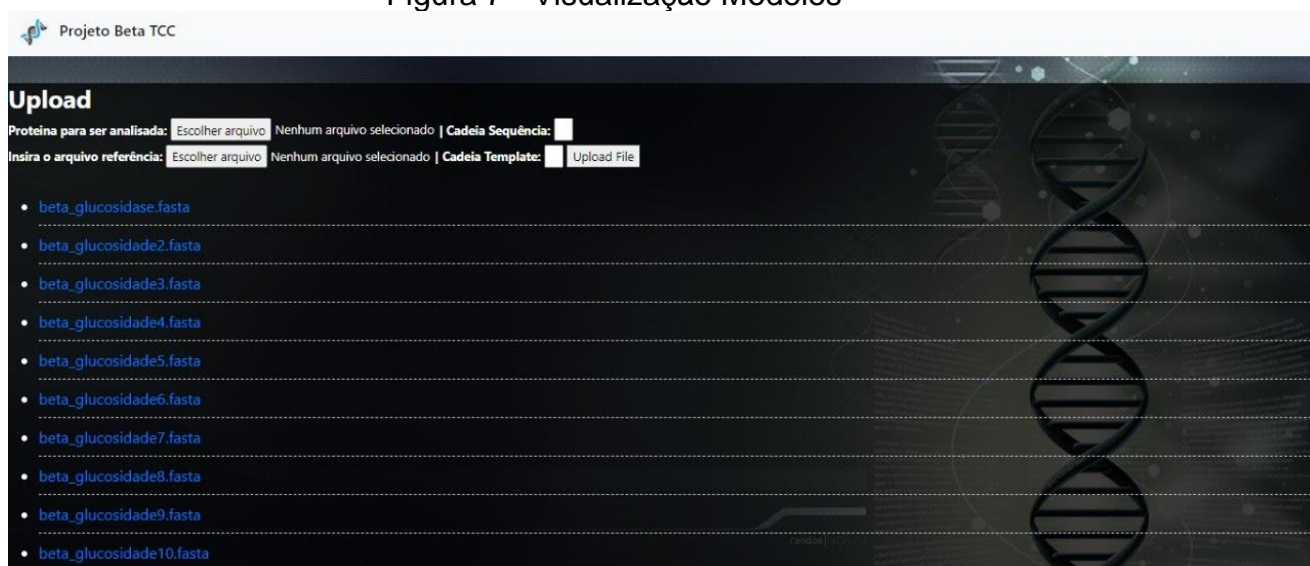
4.2 Pipeline Proteína - Template

Este módulo disponibiliza uma rotina onde é possível o desenvolvimento do modelo tridimensional a partir da sequência de uma proteína. Como entrada de dados são necessárias a sequência da proteína em questão (arquivo “.fasta”), seu respectivo template válido (arquivo “.pdb”) e as cadeias que serão analisadas, tanto da proteína, quanto ao template.

Com a inserção dos dados, será iniciado uma cadeia de scripts (pipeline) onde serão utilizadas ferramentas da bioinformática para o desenvolvimento dos modelos tridimensionais.

A primeira tela vista pelo usuário após selecionar o Pipeline Proteína, demonstrada na figura 7, apresenta os modelos 3D já realizados no sistema, nessa tela também é apresentada opções de inserção de arquivos para criar uma modelagem de proteína, os arquivos solicitados são um arquivo FASTA da proteína que será representada pelo modelo e um arquivo PDB que servirá de template para a criação do modelo. Caso os arquivos inseridos sejam cadeias de sequências, o sistema possibilita que seja escolhido qual cadeia deve ser analisada

Figura 7 - Visualização Modelos



fonte: autoria própria

Após a inserção dos arquivos serão criados 3 modelos, essa quantia pode ser alterada no código fonte, os modelos gerados são apresentados com dados específicos de nome da sequência, quantidade de cadeias e etc. Na Figura 8 é retratado como os modelos gerados são apresentados.

Figura 8 - Visualização dos modelos criados a partir de upload

Carregado com Sucesso!

Modelo: 4hpg.pdb Score: Template - Sem Score

HEADER HYDROLASE, ALLERGEN 23-OCT-12 4HPG
TITLE CRYSTAL STRUCTURE OF A GLYCOSYLATED BETA-1,3-GLUCANASE (HEV B 2), AN
TITLE 2 ALLERGEN FROM HEVEA BRASILIENSIS
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: BETA-1,3-GLUCANASE;
COMPND 3 CHAIN: A, B, C, D;
COMPND 4 EC: 3.2.1.39

Bloquear Scroll Download Arquivo

Modelo: beta_glucosidade10.fasta.B99990001.pdb Score: 2072.6064

EXPDTA THEORETICAL MODEL, MODELLER 9.24 2020/10/26 17:25:58
REMARK 6 MODELLER OBJECTIVE FUNCTION: 2072.6064
REMARK 6 MODELLER BEST TEMPLATE % SEQ ID: 97.778
REMARK 6 SEQUENCE: beta_glucosidade10.fasta
REMARK 6 ALIGNMENT: new_alinha.pir
REMARK 6 SCRIPT: run.py
REMARK 6 TEMPLATE: 4hpg.pdb 2A - 316A MODELS 38: - 352: AT 97.8%

Bloquear Scroll Download Arquivo

Modelo: beta_glucosidade10.fasta.B99990002.pdb Score: 2115.0159

ATOM	11	CB	ALA	2	8.742	34.557	116.085	1.00188.88	C
ATOM	12	C	ALA	2	7.951	36.077	117.831	1.00188.88	C

fonte: autoria própria

4.3 Comparar Sequência

Disponibiliza uma tela onde é possível inserir duas proteínas para serem comparadas, foram realizadas comparações manuais de taxa de aparição como descrito na metodologia. Também foram efetuadas comparações utilizando a ferramenta Blast, permitindo demonstrar o alinhamento global entre duas sequências.

Figura 9 - Comparação entre Sequências

Alinhar

Resultado do Alinhamento

Cabeçalho >sp|P52407|E13B_HEVBR Glucan endo-1,3-beta-glucosidase, basic vacuolar isc Cabeçalho >4hpg.pdb

Sequência MAISSSTSGTSSSFRTTVMLLFFFAASVGITDAQVGCYGMQGNLPPVSEVIALYKSNITRMRIYDPNRAVLEALRGSNIELILGVPNSDLQSLTNPSNAKSWVQKNVR Sequência QVGCYGMQGNLPPVSEVIALYKSNITRMRIYDPNRAVLEALRGSNIELILGVPNSDLQSLTNPSNAKSWVQKNVRGFWSSVRFRIYAGNEISPVNRTAWLAQFVLPA

Legenda: M: Match D: Diferente

Funções

Alinhamento Detalhado das Sequências a partir do Back End Django | Comparação 1:

MAISSSTSGTSSSFRTTVMLLFFFAASVGITDAQVGCYGMQGNLPPVSEVIALYK
K:ILTRPRIYDPNRAVLEALRG:NIETILG:PIISDLQSLTNPSNAKSW:QKNV:GFWSSV
LFRYIAGNEIS:PVNRG:AKLAQFVLP:MRNIH:AIRSAGLQDQIKV:TAIDL:TLVGNS
PPSAGAFR:DVRSY:LDPIIGFLSSIRSPLLANIYPYFTAYNPRDISLPALFTSPS:VV
WDGQR:YKMLFDATLDALY:AL:ERASGGSLVVVSSEGPSAG:FAATFDNGRTYLSNL
IQHYGGTP:RPNRAIETYLFAFMDENKKQPEVEKHGGLFFPNKQKYNLNFSAEKIMDI
STEHWATILFLKSDM

Quantidade de Aparições para cada Amionocido

A: 32 C: 1 D: 17 E: 13 F: 21 G: 24 H: 4 I: 22 K: 15 L: 35 M: 7 N: 25 P: 21 Q: 11 R: 19 S: 40 T: 19
V: 25 W: 7 Y: 16

Porcentagem de Aparições para cada Amionocido

A: 8.56% C: 0.27% D: 4.55% E: 3.48% F: 5.61% G: 6.42% H: 1.07% I: 5.88% K: 4.01% L: 9.36%
M: 1.87% N: 6.68% P: 5.61% Q: 2.94% R: 5.08% S: 10.7% T: 5.08% V: 6.68% W: 1.87% Y: 4.28%

Tamanho da Sequência: 374

Alinhamento Detalhado das Sequências a partir do Back End Django | Comparação 2:

QVGCYGMQGNLPPVSEVIALYKSNITRMRIYDPNRAVLEALRGSNIELILGVPNSDLQSLTNPSNAKSWVQKNVRGFWSSVRFRIYAGNEISPVNRTAWLAQFVLPA
Q:LTNPSNAKSWVQKNVRGFWSSVRFRIYAGNEISPVNRTAWLAQFVLPA:NIHDAI
RSAGLQDQIKV:TAIDL:LVGNSYPPS:GAFRD:VRSYLNPIIRFLS:IRSP:LANIYP
FTY:GNPRDISLP:ALFTSPSVVMDGGRGYKNLFDALDALYSALERASGGSLVV:SE
SQMPSAG:AFATFDNGRTYLSNL:IQHYGGTP:RPNRAIETYLFAFMDENKKQPEVEKHGGLFFPNKQKYNLNFSAEKIMDI
GLFFPNKQKYNLNFSAEKIMDI

Quantidade de Aparições para cada Amionocido

A: 26 C: 1 D: 13 E: 11 F: 16 G: 21 H: 3 I: 18 K: 14 L: 29 M: 4 N: 23 P: 20 Q: 12 R: 20 S: 29 T: 12
V: 23 W: 6 Y: 15

Porcentagem de Aparições para cada Amionocido

A: 8.25% C: 0.32% D: 4.13% E: 3.49% F: 5.08% G: 6.67% H: 0.95% I: 5.71% K: 4.44% L: 9.21%
M: 1.27% N: 7.3% P: 6.35% Q: 3.81% R: 6.35% S: 9.21% T: 3.81% V: 7.3% W: 1.9% Y: 4.76%

Tamanho da Sequência: 315

fonte: autoria própria

Durante a realização das análises foram efetuados testes nas plataformas Blast e Rosalind que já estão consolidadas na área de Bioinformática, assim podendo observar possíveis diferenças e falhas nas análises realizadas pelo sistema web. Todas análises que demonstram significativa diferença para com as plataformas de teste foram solucionadas ou retiradas do código.

Utilizando a plataforma Rosalind foi possível confirmar a eficiência dos códigos de análise de DNA referentes a transcrição e tradução e apresentação da fita completa do DNA. Também foi possível comprovar a análise de número de aparições de cada aminoácido em sequências de proteínas.

Com a ferramenta Blast foi iniciada a validação das comparações entre proteínas, apresentando semelhanças entre comparações feitas pelo código e comparações utilizando a ferramenta de forma online. Blast também foi utilizado para apresentar os dados na tela de comparações entre proteínas, usando a ferramenta de forma local.

Paralelamente com a realização do trabalho de conclusão de curso e desenvolvimento do sistema web local foi criado um guia de como utilizar o sistema. Esse guia apresenta a forma de baixar o sistema e utilizá-lo em sua máquina, durante o guia é apontado quais ferramentas é necessário baixar para conseguir utilizar o sistema, também apresenta passos de como realizar as instalações e verificar se tudo ocorreu como deveria. Após as instalações o guia segue com uma apresentação de cada função do sistema, nesta o foco é mostrar a função de cada botão.

5 Considerações Finais

Com a conclusão do projeto foi possível entender a importância de cada fase do projeto, realizar uma revisão de literatura na área da Bioinformática foi de extrema importância para construir uma base teórica na criação do sistema web. Podendo realizar um desenvolvimento focado em como utilizar as ferramentas que realizam as análises vistas na revisão.

Uma das dificuldades mais notáveis está na utilização das ferramentas de forma unificada e em entender como utilizar cada ferramenta para complementar outra.

O framework Django demonstrou eficiência em apresentar os dados das análises de forma gráfica, possibilitando a utilização de CSS para formatar as saídas gráficas da melhor maneira possível para o entendimento de pesquisadores que utilizam o sistema.

As saídas das análises ao serem comparadas em ferramentas online de Bioinformática demonstraram que o Sistema Web do projeto realiza com eficiência suas análises requisitadas. Com a constatação da eficiência da ferramenta web local foi demonstrado que o projeto teve eficiência em iniciar pesquisas sobre Bioinformática.

É esperado que trabalhos futuros da instituição possam utilizar esse trabalho de conclusão de curso como base para implementar análises mais específicas e utilizar a ferramenta para realizar projetos que integrem o curso Sistemas de Informação com cursos de Biologia existentes na Fundação Hermínio Ometto.

Referências Bibliográficas

BORGES, Murilo Guimarães. Aplicação de protocolos e métodos em Bioinformática para análise de sequenciamento de exomas humanos. 2015. Dissertação (Mestrado em Ciências). Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, 2015.

BOVOLenta, Luiz Gustavo. Análise exploratória em larga escala de micrornas expressos em tilápia do nilo utilizando ferramentas de Bioinformática. 2016. Tese(Doutorado em Biologia Geral e Aplicada) - Programa de Pós-Graduação em Biologia Geral e Aplicada, Universidade Estadual Paulista, Botucatu, 2016.

CARVALHO, Gesiele Almeida Barros de. Análise computacional dos genomas de duas estirpes brasileiras de *Bradyrhizobium* de importância econômica. 2016. Tese(Doutorado em Ciências) - Interunidades de Pós-Graduação em Bioinformática, Universidade de São Paulo, São Paulo, 2016.

CASANOVA, Dalcimar. Redes complexas em visão computacional com aplicações em Bioinformática. 2013. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

ESTEVES, Gustavo Henrique. Métodos estatísticos para análise de dados de cDNA microarray em um ambiente computacional integrado. 2007. Tese(Doutorado em Bioinformática) - Programa de Doutorado em Bioinformática, Universidade de São Paulo, São Paulo, 2007.

FERRETTI, Yuri. Ferramenta Computacional para Análise Integrada de Dados Clínicos e Biomoleculares. 2015. Dissertação (Mestrado em Bioinformática) - Programa de Interunidades de Pós-Graduação em Bioinformática, Universidade de São Paulo, Ribeirão Preto, 2015.

GONÇALVES, Katia Cristiane Brumatti. Estudo de Bioinformática aplicado à análise de expressão gênica utilizando dados oriundos de sequenciamento por tecnologia de "NEXT-GENERATION" em animais controle e em modelos de epilepsia do lobo temporal mesial. 2015. Dissertação (Mestrado em Ciências). Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, 2015.

LOPES, Geannine Marchezan do Nascimento. **Investigando Paradigmas de Programação Paralela sobre MPSoc HeMPS**. 2018. Trabalho de Conclusão de Curso (Bacharelado Em Ciência da Computação) - Universidade Federal do Pampa, Alegrete. 2018.

MORAIS, Diego Arthur de Azevedo. Transcriptogramer: pacote em R para análise transcricional. 2018. Dissertação (Mestrado em Bioinformática) Programa de Pós-Graduação em Bioinformática, Universidade Federal do Rio Grande do Norte, Natal, 2018.

REIS, Bianca Fujita. Análise cromossômica por microarray em pacientes com deficiência intelectual associada à obesidade. 2015. Dissertação (Mestrado em Ciências da Saúde) Programa de Pós-Graduação em Ciências da Saúde, Universidade de Brasília, Brasília, 2015.

SANTOS, Sérgio Miguel Cardoso dos. **Estudo Computacional Das Interações Proteína-Proteína**. 2010. Dissertação (Mestrado em Bioinformática e Biologia Computacional Especialização em Bioinformática) Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, 2010.

SOUZA, Rogério F. Curso de introdução à bioinformática aplicada a genômica. 2015.

Disponível em:

http://www.uel.br/laboratorios/lbi/pages/arquivos/curso/01_curso_transcriptoma.pdf

Tian, Yuan. Morris, Tiffany J. Webster, Amy P. Yang, Zhen. Beck, Stephan. Feber, Andrew. Teschendorff, Andrew E. The Chip Analysis Methylation Pipeline. 2019.

Disponível em:

<https://www.bioconductor.org/packages/devel/bioc/vignettes/ChAMP/inst/doc/ChAMP>.

html acesso em: 27 set 2019.

TORTORA, Gerard J. FUNKE, Berdell R. CASE, Christine L. MICROBIOLOGIA. 10ªed. Artmed 2012.

VERLI, Hugo. Bioinformática da Biologia à Flexibilidade Molecular. 1. ed. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2014.