



UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
SISTEMAS DE INFORMÇÃO

ROBSON SOARES SILVA
GILDO MATHEUS NASCIMENTO JÚNIOR

**FERRAMENTA DE BIOINFORMÁTICA PARA COMPARAÇÃO
ENTRE SEQUÊNCIAS DE AMINOÁCIDOS**

Ponta Porã
2024

RESUMO

A bioinformática é um campo interdisciplinar que combina biologia e informática para analisar e interpretar grandes conjuntos de dados biológicos, como sequências de DNA, proteínas, nucleotídeos e aminoácidos. No contexto deste estudo, o problema abordado é a necessidade de desenvolver ferramentas computacionais eficazes para comparar sequências biológicas e identificar relações entre diferentes organismos. O objetivo deste trabalho é desenvolver um sistema web que permita aos usuários comparar sequências de proteínas ou aminoácidos, utilizando a ferramenta Basic Local Alignment Search Tool (BLAST) e o banco de dados Protein Data Bank (PDB). O sistema visa facilitar a análise de similaridades entre sequências e a identificação de organismos relacionados com base nessas similaridades. Para atingir os objetivos propostos, foram utilizadas técnicas de desenvolvimento web e integração com o NCBI, o sistema foi implementado utilizando as linguagens HTML, CSS, JavaScript e PHP, além de frameworks de estilização e interação com o BLAST e o banco de dados PDB. O sistema desenvolvido permitiu aos usuários inserir sequências de proteínas e realizar buscas utilizando o BLASTp. Os resultados das buscas foram apresentados de forma clara e intuitiva, em formato de sequências primárias e estruturas 3D, destacando os organismos mais similares às sequências inseridas, em ordem decrescente, no formato de porcentagem. Isso proporcionou aos usuários uma maneira fácil e eficiente de explorar as relações entre suas sequências pesquisadas e as sequências encontradas nos organismos analisados, facilitando a compreensão da diversidade genética.

Palavras-chave: Bioinformática. BLAST. Sequências biológicas.

1. INTRODUÇÃO

A bioinformática é uma disciplina que combina biologia, ciência da computação e estatística para analisar e interpretar dados biológicos. Com o avanço das tecnologias de identificação e sequenciamento de organismos, a quantidade de dados biológicos disponíveis cresceu exponencialmente, criando a necessidade de ferramentas computacionais eficientes para sua análise. Este campo tem desempenhado um papel crucial na pesquisa e análise genômica, facilitando a compreensão de processos biológicos complexos, no diagnóstico de doenças e no desenvolvimento de novos medicamentos.

Um dos desafios fundamentais na bioinformática é a análise, o processamento e a interpretação dos dados biológicos, este processo é essencial para várias aplicações, incluindo a identificação de genes, a análise de relações evolutivas, pesquisas biomédicas, proteômica, desenvolvimento de medicações, saúde pública, epidemiologia e agricultura, por exemplo, dentre diversas outras possíveis aplicações importantes. No entanto, a análise manual de grandes volumes de dados biológicos é impraticável devido à sua complexidade e tamanho, ferramentas automatizadas, como o BLAST, são amplamente utilizadas para resolver este problema. Contudo, a integração dessas ferramentas em plataformas acessíveis, intuitivas e fáceis de usar ainda é limitada, especialmente para pesquisadores sem experiência em bioinformática e pessoas comuns, que não possuem muito conhecimento na área.

A necessidade de uma ferramenta acessível, de fácil utilização e eficiente para a comparação de sequências biológicas norteia este trabalho, a criação de uma aplicação web que utilize o BLAST e as bases de dados do PDB oferece uma solução prática para este problema. Este sistema permite que pesquisadores e estudantes insiram sequências de proteínas e obtenham rapidamente resultados sobre as similaridades com outras sequências conhecidas, através do sistema desenvolvido chamado BLASTExplorer, que faz buscas por meio do BLASTp, que realiza busca e comparação entre sequências de proteínas.

A importância dessa ferramenta reside em sua capacidade de democratizar o acesso a análises bioinformáticas avançadas, ao disponibilizar uma interface web intuitiva, de fácil uso e acesso, pesquisadores de diferentes áreas, como biologia, matemática, medicina e agricultura, podem realizar análises complexas sem a necessidade de um conhecimento aprofundado em bioinformática. Isso acelera o processo de descoberta científica e permite uma utilização mais ampla e eficaz dos dados biológicos disponíveis.

Além disso, a aplicação desenvolvida neste trabalho pode servir como uma plataforma educacional, facilitando o ensino de bioinformática e genética, estudantes podem aprender

sobre a análise de sequências biológicas de maneira interativa, compreendendo melhor os conceitos teóricos através da prática.

O objetivo deste trabalho é desenvolver uma aplicação web que permita a comparação de sequências de proteínas utilizando o BLASTp, a aplicação é capaz de realizar buscas eficientes, retornando resultados relevantes e apresentando as informações de forma clara e acessível aos usuários. Este projeto inclui a implementação de uma interface amigável que aceita sequências no formato FASTA, executa buscas com o BLASTp e exibe os resultados com detalhes sobre as similaridades encontradas.

O artigo está organizado da seguinte forma: inicialmente, são apresentados o referencial teórico e os procedimentos metodológicos usados, em seguida são apresentados conceitos fundamentais de bioinformática, tais como DNA, nucleotídeos, aminoácidos proteínas, sequências FASTA, o NCBI (National Center for Biotechnology), e a ferramenta BLAST. Posteriormente, aborda-se a importância da bioinformática e suas principais áreas de aplicação. A seguir, descreve-se o desenvolvimento do sistema web, incluindo suas funcionalidades e tecnologias utilizadas. Por fim, são apresentados os resultados e as considerações finais.

2. REFERENCIAL TEÓRICO

Este capítulo tem por objetivo apresentar o referencial teórico do trabalho e fundamentar os conceitos utilizados para o desenvolvimento da ferramenta web de bioinformática. Assim, buscaram-se no campo da bioinformática as principais referências na definição dos conceitos, técnicas e ferramentas relacionadas à análise e comparação de sequências biológicas.

A comparação de sequências biológicas é uma das operações fundamentais na bioinformática, permitindo identificar similaridades, diferenças evolutivas e relações estruturais entre as sequências. Algoritmos de alinhamento, como o Smith-Waterman e o BLAST, são amplamente utilizados para comparar sequências de DNA, RNA e proteínas.

Para o desenvolvimento da ferramenta proposta neste trabalho, foram estudadas as principais técnicas de comparação de sequências biológicas, com foco em proteínas e aminoácidos. Foram analisados os conceitos de alinhamento global e local, sequências FASTA, arquivos em formato PDB, bancos de dados biológicos, como o NCBI e o PDB, e a utilização de ferramentas como o BLAST para busca por similaridades em organismos nesses bancos de dados.

Além disso, buscou-se embasamento teórico sobre a importância da bioinformática na análise de dados biológicos e suas aplicações em diversas áreas, como genética, genômica, agricultura, diagnóstico de doenças e desenvolvimento de medicamentos. A revisão da literatura também abordou o potencial da bioinformática como ferramenta educacional para o ensino de conceitos relacionados à biologia molecular e genética (Cassoti, 2022).

Este referencial teórico apresenta os principais conceitos e técnicas da bioinformática aplicados ao desenvolvimento da ferramenta de comparação de sequências de proteínas e aminoácidos, fundamentando as escolhas metodológicas e embasando a relevância deste trabalho para a área da bioinformática.

2.1 IMPORTÂNCIA DA ANÁLISE DE DADOS BIOLÓGICOS

A análise de dados biológicos é uma prática essencial na bioinformática, desempenhando um papel crucial na interpretação e compreensão de processos biológicos complexos. Com a evolução das tecnologias de sequenciamento e a explosão da quantidade de dados biológicos disponíveis, a necessidade de métodos eficientes para analisar e interpretar esses dados se tornou mais evidente do que nunca. Este tópico explora a

importância da análise de dados biológicos e como ela contribui para diversas áreas da ciência e da medicina.

2.1.1 AVANÇOS EM GENÔMICA E PROTEÔMICA

Os avanços em genômica e proteômica têm permitido uma compreensão mais profunda da estrutura e função dos genomas e proteomas. A análise de sequências de DNA, RNA e proteínas fornece insights sobre a organização genômica, a regulação dos genes, a e a interação entre proteínas. Ferramentas como o BLAST são fundamentais para comparar essas sequências e identificar similaridades que podem indicar relações evolutivas entre organismos (Lander, 2001).

2.1.2 DIAGNÓSTICO E TRATAMENTO DE DOENÇAS

A bioinformática tem um impacto significativo na área da saúde, especialmente no diagnóstico e tratamento de doenças. A análise de dados biológicos permite a identificação de fatores genéticos associados a doenças, facilitando o desenvolvimento de diagnósticos mais precisos e novos tratamentos. Por exemplo, a identificação de mutações específicas em genes pode ajudar a prever a resposta de um paciente a determinados medicamentos, permitindo um tratamento personalizado e mais eficaz (Collins, 2003).

2.1.3 DESENVOLVIMENTO DE NOVOS MEDICAMENTOS

No campo da farmacogenômica, a análise de dados biológicos é essencial para o desenvolvimento de novos medicamentos, a comparação de sequências de proteínas pode revelar alvos terapêuticos potenciais e ajudar na criação de drogas que interajam especificamente com esses alvos. Ferramentas como o BLAST permitem a identificação de proteínas com funções similares, o que pode acelerar o processo de descoberta de novos fármacos.

2.1.4 AGRICULTURA E BIOTECNOLOGIA

Na agricultura, a análise de dados biológicos é utilizada para melhorar a resistência das plantas a doenças e aumentar a produtividade agrícola, a identificação de genes responsáveis por características desejáveis pode levar ao desenvolvimento de variedades de plantas mais resistentes e nutritivas. Na biotecnologia, a engenharia genética se beneficia da

análise de sequências para criar organismos modificados com propriedades específicas, como a produção de enzimas industriais ou a degradação de poluentes (Embrapa).

2.1.5 SAÚDE PÚBLICA E EPIDEMIOLOGIA

A análise de dados biológicos desempenha um papel crucial na saúde pública e epidemiologia, especialmente no monitoramento e controle de doenças infecciosas. Sequenciar estruturas de patógenos e compará-los pode ajudar a rastrear surtos de doenças, identificar modos de transmissão e desenvolver vacinas e tratamentos eficazes.

2.1.6 EDUCAÇÃO E PESQUISA ACADÊMICA

A análise de dados biológicos também tem um impacto significativo na educação e pesquisa acadêmica, ferramentas de bioinformática acessíveis permitem que estudantes e pesquisadores conduzam análises complexas de sequências biológicas, promovendo a aprendizagem na prática e o avanço do conhecimento científico. Plataformas como o BLASTExplorer, proporcionam interfaces intuitivas e acessíveis, democratizando o acesso a técnicas avançadas de bioinformática e contribuindo para a formação de futuros cientistas.

2.2 DEFINIÇÃO DOS TERMOS UTILIZADOS

2.2.1 DNA

O DNA (ácido desoxirribonucleico) é o material genético que contém as instruções necessárias para o desenvolvimento, funcionamento, crescimento e reprodução de todos os organismos vivos e muitos vírus. É composto por duas fitas de nucleotídeos que formam uma dupla hélice, onde cada nucleotídeo é composto por um açúcar (desoxirribose), um grupo fosfato e uma base nitrogenada (adenina, timina, citosina ou guanina) (Watson, 1953).

2.2.2 RNA

O RNA (ácido ribonucleico) é uma molécula essencial para várias funções biológicas, incluindo a codificação, decodificação, regulação e expressão de genes.

2.2.3 NUCLEOTÍDEOS

Os nucleotídeos são as unidades básicas do DNA e RNA, cada nucleotídeo é composto por três componentes: um açúcar (desoxirribose no DNA e ribose no RNA), um

grupo fosfato e uma base nitrogenada. As bases nitrogenadas no DNA são adenina (A), timina (T), citosina (C) e guanina (G); no RNA, a timina é substituída por uracila (U). A sequência de nucleotídeos forma a estrutura primária dos ácidos nucleicos e é fundamental para a codificação genética (Alberts, 2002).

2.2.4 BLAST

O BLAST (Basic Local Alignment Search Tool) é uma ferramenta amplamente utilizada para comparar uma sequência biológica (DNA, RNA ou proteína) com um banco de dados de sequências e identificar regiões de similaridade local. Desenvolvido pelo NCBI, o BLAST permite a identificação rápida de sequências, facilitando a análise de funções biológicas e relações evolutivas (Neis, 2020).

2.2.5 NCBI

O NCBI (National Center for Biotechnology Information) é uma instituição que fornece acesso a uma vasta gama de informações biológicas e bioinformáticas. Entre seus muitos recursos, o NCBI oferece o GenBank, um banco de dados de sequências genéticas, e ferramentas como o BLAST, que permitem a análise e comparação de sequências biológicas (NCBI, 2023).

2.2.6 PDB

O PDB (Protein Data Bank) é um banco de dados que armazena informações sobre estruturas tridimensionais de proteínas, ácidos nucleicos e complexos proteicos. Este recurso é crucial para pesquisadores que buscam entender a estrutura e função de biomoléculas, facilitando o estudo de interações moleculares (Berman, 2000).

2.2.7 SEQUÊNCIAS FASTA

O formato FASTA é um padrão de arquivo para representar sequências biológicas, como nucleotídeos ou aminoácidos. Cada sequência no formato FASTA começa com uma linha de descrição, iniciada por um sinal de maior ">", seguida pela sequência biológica estrutural de cada molécula (Yokomizo, 2022).

2.2.8 PROTEÔMICA

A proteômica é o estudo em larga escala das proteínas, particularmente suas estruturas e funções. Este campo envolve a identificação e quantificação de proteínas, bem como a análise de suas modificações e interações. A proteômica é fundamental para entender os processos biológicos a nível molecular e tem aplicações em diagnóstico, tratamento de doenças e desenvolvimento de medicamentos (Barbosa, 2012).

2.2.9 FARMACOGENÔMICA

A farmacogenômica é o estudo de como os genes de uma pessoa influenciam sua resposta a medicamentos, este campo combina farmacologia e genômica para desenvolver medicamentos eficazes e seguros, adaptados ao perfil genético individual de cada paciente. A farmacogenômica visa personalizar tratamentos médicos para melhorar a eficácia e reduzir os efeitos colaterais (Oncoguia, 2015).

2.2.10 GENÔMICA

A genômica é o estudo abrangente dos genomas de organismos, envolve a análise da estrutura, função, evolução e mapeamento de genomas inteiros. A genômica permite a identificação de todos os genes de um organismo, sua localização e suas funções, fornecendo insights sobre a biologia, evolução e mecanismos de doenças (Biazon, 2018).

2.2.11 PATÓGENOS

Patógenos são organismos, como vírus, bactérias, fungos e parasitas, que causam doenças em seus hospedeiros. A análise genômica e proteômica de patógenos é essencial para entender os mecanismos de infecção e resistência, desenvolver diagnósticos e criar tratamentos e vacinas eficazes (Cetesb).

2.2.12 SEQUÊNCIAS BIOLÓGICAS

Sequências biológicas são cadeias de nucleotídeos ou aminoácidos que formam as estruturas primárias de DNA, RNA e proteínas. A análise dessas sequências é fundamental para entender a informação genética e as funções biológicas dos organismos (Freitas).

2.3 FERRAMENTAS COMPUTACIONAIS

As ferramentas computacionais desempenham um papel de grande importância na bioinformática, oferecendo recursos essenciais para análise, interpretação e compreensão de dados biológicos. Elas permitem a manipulação eficiente de grandes conjuntos de dados, acelerando descobertas científicas e avanços em áreas como genética, medicina e agricultura. O uso de ferramentas computacionais na bioinformática é fundamental para impulsionar a pesquisa científica, facilitando análises complexas e ajudando os cientistas a fazerem novas descobertas.

3. PROCEDIMENTOS METODOLÓGICOS

Este trabalho teve como objetivo a criação de um sistema web para busca e comparação de sequências de proteínas, utilizando ferramentas e bancos de dados consagrados na área de bioinformática. O sistema permite aos usuários comparar sequências de proteínas e aminoácidos utilizando o Basic Local Alignment Search Tool (BLAST), através do seu executável disponível no National Center for Biotechnology Information (NCBI) e o banco de dados Protein Data Bank (PDB), também foi criado um banco de dados local com organismos encontrados no PDB para realizar a busca e visualização de estruturas 3D das moléculas.

Este estudo caracteriza-se como uma pesquisa aplicada, uma vez que visa resolver um problema prático no campo da bioinformática através do desenvolvimento de uma ferramenta específica. A natureza quantitativa da pesquisa está presente na análise das similaridades entre as sequências de proteínas, apresentadas de forma numérica (porcentagem, a pesquisa também é descritiva, pois documenta o processo de desenvolvimento do sistema.

O desenvolvimento do sistema foi baseado principalmente no trabalho de Diego Mariano, editor-chefe da revista BIOINFO e doutor em Bioinformática pela Universidade Federal de Minas Gerais. Mariano é autor do livro "Introdução à Programação para Web para Bioinformática", que serviu como guia metodológico e técnico para este trabalho. Além disso, o sistema desenvolvido por Mariano foi uma referência crucial, fornecendo a base conceitual e prática necessária para a criação do site e deste artigo.

3.1 DESENVOLVIMENTO DO SISTEMA

3.1.1 FERRAMENTAS E TECNOLOGIAS UTILIZADAS:

- **Linguagens de programação:** o sistema foi desenvolvido utilizando HTML, CSS, JavaScript e PHP.
- **Frameworks:** foi utilizado o Bootstrap para estilização e interação do sistema, garantindo uma interface amigável e responsiva.
- **Conexões e bancos de dados:** o executável do NCBI foi baixado para permitir a utilização do BLAST localmente. O banco de dados PDB foi utilizado para obter as sequências de proteínas necessárias para as comparações e para as visualizações em 3D.
- **Banco de dados local:** foi criado um banco de dados local que armazena organismos encontrados no PDB, juntamente com as sequências FASTA.

4. ORIGENS E EVOLUÇÃO DA BIOINFORMÁTICA

A bioinformática é uma área que combina disciplinas como biologia, ciência da computação, engenharias, matemática e estatística para analisar e interpretar dados biológicos, dessa forma, os estudiosos e cientistas podem aproveitar diversas ferramentas e tecnologias para facilitar e analisar suas pesquisas.



Figura 1: Composição da bioinformática.

Com a evolução da nossa sociedade nos últimos anos, conseguimos reunir uma enorme quantidade de informações, o que também se aplica à biologia, há alguns anos, as pesquisas em genômica e proteômica, por exemplo, resultaram em uma quantidade significativa de dados biológicos, esses dados precisam ser analisados pelos cientistas, pois, quando agrupados corretamente, se transformam em informações úteis.

No entanto, o ser humano é incapaz de processar e organizar enormes quantidades de dados sozinho, a complexidade e o volume desses dados são tão vastos que análises manuais seriam insuficientes para extrair informações significativas de maneira rápida e eficiente. Por isso, surge a necessidade de sistemas computacionais para auxílio dessa tarefa.

Esses sistemas computacionais são projetados para lidar com grandes volumes de dados, realizando análises complexas em uma pequena fração do tempo que seria necessário para um ser humano. Com o auxílio de softwares especializados, os estudiosos podem armazenar uma grande quantidade de dados biológicos, fazer buscas, fazer comparações, identificar padrões, realizar simulações e prever resultados com alta velocidade e precisão, ajudando a compreender a complexidade dos sistemas biológicos.

A bioinformática começou a se desenvolver a partir dos anos 1950, quando foi possível obter as primeiras sequências de proteínas usando a técnica de degradação de Edman, a leitura direta de sequências de DNA só começou na década de 1970, com a criação do método de sequenciamento Sanger, que foi empregado para sequenciar todo o conjunto de moléculas de DNA de um organismo, conhecido como genoma, até o início do século 21. Essa tecnologia foi fundamental para o Projeto Genoma Humano, uma colaboração científica internacional destinada a determinar a sequência completa do genoma do Homo sapiens, esse projeto durou 13 anos e teve um custo estimado superior a US\$500 milhões [1].

A partir de meados dos anos 2000, surgiram as primeiras tecnologias de sequenciamento de DNA de nova geração, conhecidas como NGS (Next Generation Sequencing), que reduziram significativamente o custo do sequenciamento de genomas completos [1]. Isso acarretou um crescimento exponencial na quantidade de dados biológicos gerados diariamente, trabalhadores da área, atuam tanto no desenvolvimento quanto no uso de programas para aquisição, armazenamento, organização, análise e visualização de dados de saúde, comportamentais, epidemiológicos e, principalmente, da genética e biologia molecular, sequências de DNA, RNA e proteínas.

No contexto atual, a bioinformática desempenha um papel crucial em várias frentes, desde a pesquisa genômica até o desenvolvimento de novos medicamentos e terapias, abaixo estão alguns campos de estudo de grande importância na bioinformática:

- **Genômica e Sequenciamento de DNA**
- **Medicina de Precisão**
- **Descoberta de Fármacos**
- **Estudos de Populações e Evolução**
- **Diagnóstico Clínico**
- **Agricultura**

- **Biotecnologia**
- **Fontes de energia alternativa**
- **Modelagem Molecular**

As principais tarefas da Bioinformática estão ligadas a análises de informações de sequências que envolvem a identificação de genes nas sequências de DNA de vários organismos, desempenhando um papel crucial na análise de informações dessas sequências, centrando-se na identificação de genes [2].

Utilizando sistemas especializados, os bioinformatas conseguem prever a localização e a função dos genes dentro do genoma de diversos organismos, esse processo envolve o alinhamento de sequências, a anotação genômica e a análise de padrões específicos. A identificação precisa de genes é fundamental para entender a biologia molecular, facilitando avanços na medicina, agricultura e biotecnologia, por exemplo.

4.1 O DOGMA CENTRAL DA BIOLOGIA MOLECULAR

O DNA é o material genético de todos os organismos da Terra, o dogma central da biologia molecular está diretamente ligado a ele, é um conceito fundamental que descreve o fluxo de informação genética dentro de uma célula, este conceito foi proposto por Francis Crick em 1958 e estabelece que a informação genética flui do DNA para o RNA e, finalmente, para as proteínas [3]. Este processo é essencial para a expressão gênica e a regulação das funções celulares, há duas etapas principais fundamentais nesse processo:

1. **Transcrição:** a sequência de DNA de um gene é copiada para fazer uma molécula de RNA, essa etapa leva o nome de transcrição porque envolve transcrever a sequência de DNA em uma sequência similar de RNA.
2. **Tradução:** nessa etapa o RNA é traduzido em uma sequência de aminoácidos para formar uma proteína.

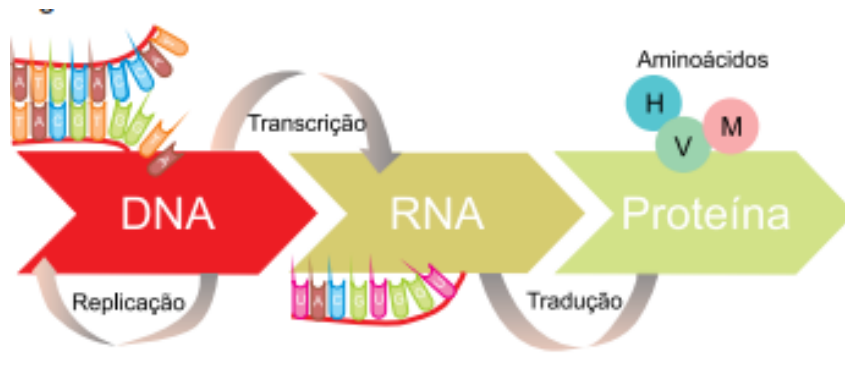


Figura 2: O dogma central da biologia molecular.

Assim sendo, nesse processo da expressão de um gene codificante de proteína, a informação flui do DNA para o RNA e dele para proteína, esse fluxo direcional de informação genética é conhecido como o dogma central da biologia molecular.

4.2 CÓDIGO GENÉTICO

A compreensão da biologia molecular começa com o conhecimento das moléculas que formam a base da vida, o DNA, nucleotídeos, aminoácidos e proteínas desempenham papéis cruciais na estrutura e função das células. Este tópico aborda cada uma dessas moléculas essenciais:

4.2.1 DNA: O ARMAZENADOR DA INFORMAÇÃO GENÉTICA

O ácido desoxirribonucleico (DNA) é uma molécula composta por nucleotídeos, que contém as instruções genéticas usadas no desenvolvimento e funcionamento de todos os organismos vivos e alguns vírus.

- **Estrutura do DNA:** tem o formato de dupla hélice composta por duas cadeias polinucleotídicas antiparalelas., cada cadeia é formada por nucleotídeos, que são as unidades básicas do DNA.
- **Função do DNA:** ele armazena e transmite a informação genética de geração em geração, é responsável pela hereditariedade e pelo controle das atividades celulares através da expressão gênica.

4.2.2 NUCLEOTÍDEOS: OS BLOCOS DE CONSTRUÇÃO DO DNA

Os nucleotídeos são as unidades que compõem o DNA e o RNA:

- **Composição:** cada nucleotídeo é composto por três componentes:
 1. **Uma base nitrogenada:** (A) adenina, (T) timina, (C) citosina ou (G) guanina no DNA; adenina, (U) uracila, citosina ou guanina no RNA.
 2. **Um açúcar de cinco carbonos:** desoxirribose no DNA e ribose no RNA.
 3. **Um grupo fosfato.**
- **Função:** os nucleotídeos são os blocos de construção do DNA e RNA, eles também desempenham papéis importantes no metabolismo celular, como transportadores de energia, como a molécula ATP, que é um nucleotídeo e transporta energia para inúmeros processos metabólicos.

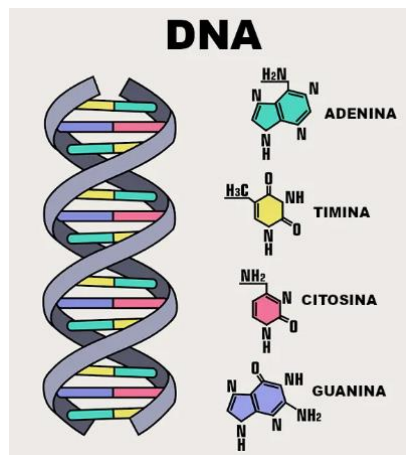


Figura 3: a composição do DNA.

4.2.3 AMINOÁCIDOS: OS BLOCOS DE CONSTRUÇÃO DAS PROTEÍNAS

Os aminoácidos são moléculas orgânicas que combinam para formar proteínas. São conhecidos 20 aminoácidos (Alanina, Arginina, Aspartato, Asparagina, Cisteína, Fenilalanina, Glicina, Glutamato, Glutamina, Histidina, Isoleucina, Leucina, Lisina, Metionina, Prolina, Serina, Tirosina, Treonina, Triptofano e Valina), que são responsáveis por formar todas as proteínas existentes, no corpo humano são cerca de cem mil proteínas diferentes formadas pelas combinações dos aminoácidos.

4.2.4 PROTEÍNAS

As proteínas são macromoléculas complexas que desempenham uma ampla variedade de funções nas células, elas estão relacionadas com praticamente todas as funções de um organismo vivo.

- **Estrutura:** as proteínas são formadas pela combinação de aminoácidos ligados por ligações peptídicas. A estrutura das proteínas pode ser dividida em quatro níveis:
 1. **Primária:** sequência linear de aminoácidos.
 2. **Secundária:** é caracterizada por padrões regulares e repetitivos que ocorrem localmente, causada pela atração entre certos átomos de aminoácidos próximos.
 3. **Terciária:** dobramento tridimensional da proteína inteira sobre si mesma.
 4. **Quaternária:** formada por múltiplas cadeias polipeptídicas.
- **Função:** As proteínas desempenham diversas funções, incluindo:
 1. **Enzimática:** catalisam reações bioquímicas.
 2. **Estrutural:** fornecem suporte estrutural às células e tecidos.
 3. **Transporte:** transportam moléculas e íons através das membranas celulares.
 4. **Regulatória:** regulam processos celulares e gênicos.
 5. **Defensiva:** participam da resposta imune.

Função	Exemplo
Estruturação	Microtúbulos
Trasporte	Albumina
Movimentação	Actina e Miosina
Alimentação	Caseína
Defesa	Imunoglobulinas
Coordenação	Insulina
Catalisador	Enzimas Proteicas

Figura 4: exemplos de proteínas e suas funções.

5. SEQUÊNCIAS BIOLÓGICAS

As sequências biológicas referem-se à ordem de nucleotídeos no DNA e RNA ou à sequência de aminoácidos em proteínas, essas sequências são fundamentais para a codificação

e organização da informação genética e a determinação da estrutura e função das moléculas biológicas. Essas sequências são compostas por cadeias de caracteres, um dos trabalhos na área da bioinformática é fazer a comparação dessas cadeias de diferentes organismos, portanto, podemos dizer que essa função seria fazer a comparação entre cadeias de caracteres.

- **Sequências de DNA e RNA:** as sequências de DNA e RNA são compostas por nucleotídeos, que são as unidades básicas dessas moléculas. No DNA, as bases nitrogenadas adenina (A), timina (T), citosina (C) e guanina (G) se organizam em uma dupla hélice. No RNA, a timina é substituída pela uracila (U).
- **Sequências de Aminoácidos:** as sequências de aminoácidos formam as proteínas.

5.1 FORMATOS DE REPRESENTAÇÃO

As sequências biológicas, como DNA, RNA e proteínas, são representadas em formato de texto padronizado para facilitar sua análise e interpretação, esses formatos de representação são essenciais para a o armazenamento e análise de dados biológicos na bioinformática.

Um dos formatos mais utilizados para representar sequências de DNA, RNA e proteínas é o formato **FASTA**, que é um padrão de armazenamento de sequências biológicas, que representa tanto sequências de nucleotídeos quanto sequências de aminoácidos. Ele é um arquivo de texto, com a extensão **.FASTA**, e é organizado de forma padronizada, da seguinte maneira:

- É representado por uma série de linhas, que não devem ter mais do que 120 caracteres e, geralmente, não excedem 80 caracteres.
- É caracterizado por uma estrutura simples, composta por duas partes principais: a linha de cabeçalho e a sequência biológica.
- A linha de cabeçalho é a primeira linha de um arquivo no formato FASTA, ela começa com o caractere > (maior) e pode ser seguida por uma descrição opcional da sequência.
- Após a linha de cabeçalho, as linhas posteriores contêm a sequência biológica, que pode ser de nucleotídeos aminoácidos.



Figura 5: representação de sequência no formato de arquivo FASTA.

5.1.2 APLICAÇÕES DO FORMATO FASTA

- **Alinhamento de sequências:** ferramentas de alinhamento de sequências, como BLAST e Clustal, aceitam arquivos FASTA como entrada para comparar e alinhar sequências biológicas.
- **Armazenamento em bancos de dados:** grandes bancos de dados genômicos, como GenBank, do NCBI, utilizam o formato FASTA para armazenar e compartilhar sequências.
- **Análise de Sequências:** softwares de bioinformática, como BioEdit e MEGA, utilizam arquivos FASTA para alinhamento e análises filogenéticas.

6. A FERRAMENTA BLAST

A ferramenta BLAST é um algoritmo de comparação de sequências que utiliza um método de alinhamento local para identificar regiões semelhantes entre sequências de nucleotídeos ou aminoácidos. A ferramenta foi desenvolvida pelo NCBI, ela permite que os pesquisadores encontrem regiões de similaridade local entre sequências, o que é fundamental para a análise de dados genômicos e proteômicos.

6.1 FUNCIONAMENTO DO BLAST

1. **Entrada da sequência:** o usuário insere uma sequência de nucleotídeos (DNA/RNA) ou de aminoácidos (proteínas) que deseja comparar com um banco de dados de sequências conhecidas.
2. **Divisão em palavras:** a sequência de entrada é dividida em fragmentos curtos, chamados de palavras.

3. **Pesquisa por palavras correspondentes:** o BLAST procura essas palavras em um banco de dados para encontrar sequências semelhantes.
4. **Busca pela maior correspondência:** quando uma correspondência é encontrada, o algoritmo tenta estender o alinhamento tanto quanto possível para obter uma correspondência mais longa.
5. **Listagem das moléculas correspondentes:** O BLAST lista as sequências correspondentes à pesquisa, avaliando e ordenando as sequências com base nas similaridades encontradas.

6.2 TIPOS DE BLAST

Existem vários tipos de programas BLAST, cada um específico para cada tipo de comparação de sequências:

1. **BLASTN:** compara uma sequência de nucleotídeos com um banco de dados de nucleotídeos.
2. **BLASTP:** compara uma sequência de proteínas com um banco de dados de proteínas.
3. **BLASTX:** compara uma sequência de nucleotídeos com um banco de dados de proteínas, traduzindo a consulta em tempo real.
4. **TBLASTN:** compara uma sequência de proteína com um banco de dados de nucleotídeos, traduzindo a consulta em tempo real.
5. **TBLASTX:** compara uma sequência de nucleotídeos traduzida com outra sequência de nucleotídeos traduzida. É a variação mais lenta do BLAST, pois a sequência de nucleotídeos é convertida em 6 sequências de aminoácidos (uma para cada fase de leitura), que depois são comparadas com as 6 possíveis fases de leitura em um banco de dados de nucleotídeos.

O objetivo deste trabalho foi desenvolver uma ferramenta web chamada BLASTExplorer, que permite aos usuários comparar sequências de aminoácidos utilizando a ferramenta BLAST e o banco de dados Protein Data Bank (PDB). O sistema visa facilitar a análise de similaridades entre sequências biológicas, a identificação e organização de organismos relacionados com base nessas similaridades. Nos próximos tópicos, será detalhado como foi o desenvolvimento do sistema BLASTExplorer, abordando seus objetivos, funcionalidades, arquitetura linguagens utilizadas e a implementação prática da ferramenta.

7. DESENVOLVIMENTO DO SISTEMA WEB

O sistema desenvolvido é uma aplicação web chamada BLASTExplorer, que permite aos usuários realizar buscas utilizando a ferramenta BLASTp, o objetivo principal do sistema é fornecer uma ferramenta eficiente e acessível para a comparação de sequências de aminoácidos, facilitando a análise de similaridades entre essas sequências e a identificação de organismos relacionados com base nessas similaridades.

Além da capacidade de realizar buscas utilizando o BLASTp, que compara as sequências inseridas de proteínas com um banco de dados de proteínas, o sistema possui um banco de dados criado localmente, composto por moléculas encontradas no banco de dados Protein Data Bank (PDB) e suas respectivas sequências em FASTA. Isso permite que os usuários não apenas realizem buscas por similaridades entre sequências, mas também naveguem e explorem o banco de dados local, podendo visualizar as estruturas 3D de moléculas disponíveis nele.

7.1 ARQUITETURA DO SISTEMA

O BLASTExplorer é uma aplicação web desenvolvida utilizando uma arquitetura cliente-servidor, o cliente, que é a interface web acessada pelos usuários, envia requisições para o servidor, que processa as solicitações e retorna os resultados.

- **O sistema é composto por três principais componentes:**

1. **Interface web:** desenvolvida em HTML, CSS e JavaScript, permite que os usuários insiram sequências de proteínas no formato FASTA e visualizem os resultados das buscas.
2. **Servidor web:** implementado em PHP, recebe as requisições do cliente, processa as sequências usando o BLASTp e retorna os resultados.
3. **Banco de dados local:** criado a partir de estruturas baixadas do banco de dados PDB juntamente com as suas sequências em FASTA, armazena informações sobre as sequências primárias e estruturas 3D das proteínas.

- **Fluxo de execução do sistema:**

1. O usuário insere uma sequência de proteína no formato FASTA através da interface web.
2. A sequência é enviada como uma requisição HTTP para o servidor web.

3. O servidor Apache recebe a requisição e a envia para o servidor PHP.
4. O servidor PHP processa a sequência usando o BLASTp.
5. O servidor PHP consulta o banco de dados local para obter informações sobre as sequências similares encontradas.
6. Os resultados, incluindo informações sobre os organismos com mais similaridade são enviados de volta para a interface web.
7. A interface web exibe os resultados de forma clara e intuitiva, em formato de sequências primárias, destacando os organismos mais similares às sequências inseridas, em ordem decrescente, no formato de porcentagem.

- **Integração com o NCBI e bancos de dados:**

Para desenvolver o sistema, foi baixado o executável do BLAST+, disponibilizado no site do NCBI [24], que permite ao usuário fazer buscas locais em seu próprio servidor e banco de dados, utilizando, nesse caso, o BLASTp. Esse banco de dados local é criado a partir de moléculas encontradas no banco de dados PDB, que tiveram seus arquivos em formato .PDB e as sequências no formato .FASTA baixados e organizadas em uma pasta chamada 'data', onde o BLAST faz a busca para comparação com as sequências inseridas pelo usuário, além dos resultados encontrados mostrarem as moléculas com maior grau de similaridade, é possível a visualização das estruturas 3D das proteínas, através da aba 'navegar' do site.

7.2 EXECUÇÃO DO SISTEMA

Neste tópico, vamos explorar a implementação e funcionalidades do sistema na prática, a seguir será mostrada a interface da aplicação, sua homepage, seguida de cada seção do site, depois a página ‘navegar’, que mostra as visualizações 3D das moléculas disponíveis no banco de dados e, por último, a página onde o usuário insere as sequências para realizar as buscas desejadas.

1. **Homepage:** a página inicial do sistema.

BLASTExplorer Sobre Navegar BLAST Ajuda

Proteínas

Existem 100.000 tipos de proteínas compostas por apenas 20 aminoácidos

Ferramenta Blast

O programa **BLAST** é um ferramenta que foi desenvolvida para a realização de buscas, através da comparação de sequências biológicas primárias, como as sequências de aminoácidos de proteínas ou os nucleotídeos das sequências de DNA e/ou RNA em um banco de dados que contém uma grande quantidade de informações. Nessa busca, são retomadas as sequências com maior grau de similaridade e de maior significância estatística em relação a sequência submetida.

Saiba mais

Sobre

As proteínas são constituídas por sequências de aminoácidos, são cruciais para a estrutura e função celular, desempenhando papéis fundamentais no organismo.

Saiba mais

Navegar

Nesta base de dados listamos cadeias de proteínas disponíveis no banco de dados PDB. Descubra todas as sequências com estruturas tridimensionais disponíveis aqui.

Navegar

Buscar

Realize buscas por sequências de aminoácidos usando a ferramenta BLAST. BLASTExplorer permite o uso da ferramenta blastp (proteína -> proteína) no banco de dados PDB.

Buscar

BLASTExplorer Sobre Navegar Buscar Ajuda

© 2024 BLASTExplorer | Alguns direitos reservados. Construído com base no sistema criado por Diego Mariano.

Figura 6: página inicial do BLASTExplorer.

2. **Página sobre:** nessa seção é apresentada uma descrição sobre o sistema e seu funcionamento.

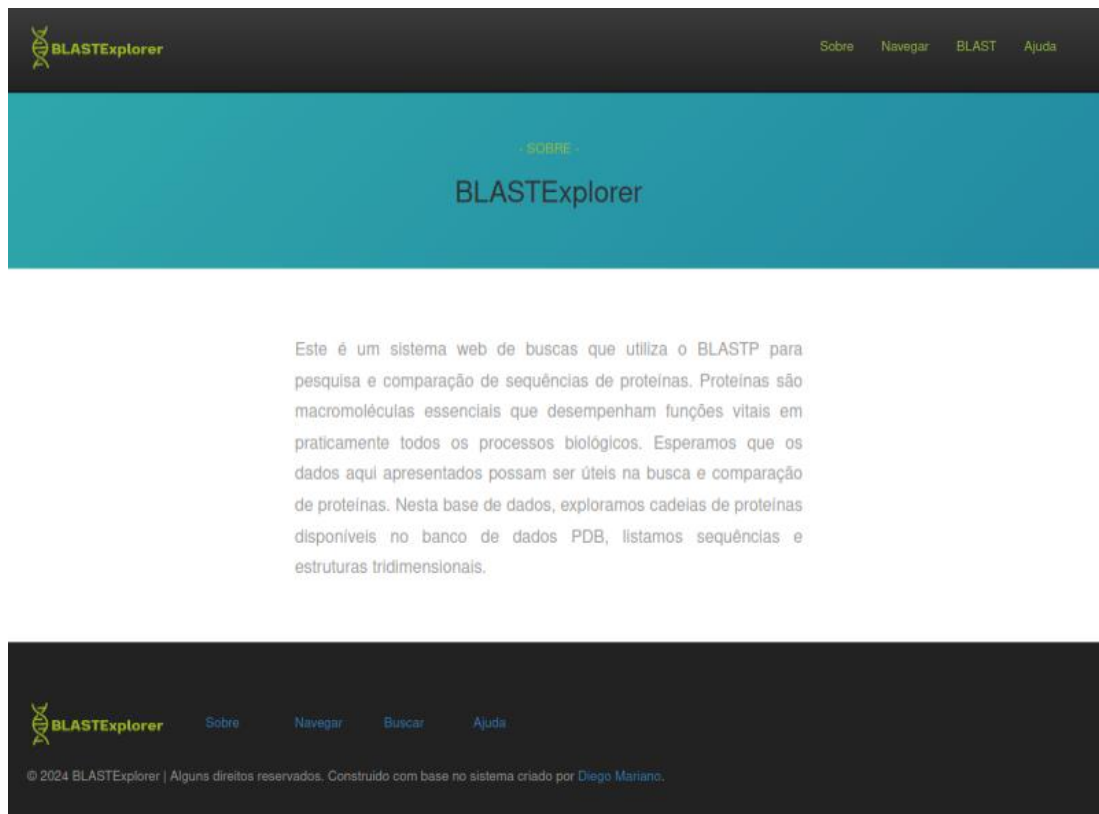
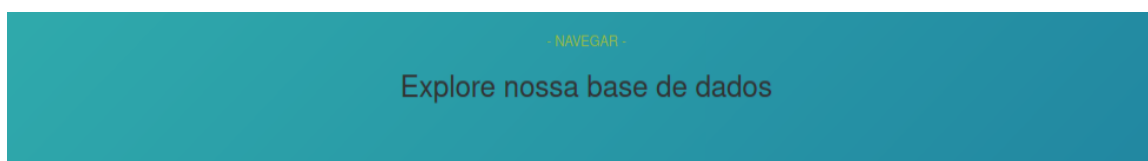


Figura 7: página sobre do BLASTExplorer.

3. **Página navegar:** nessa página o usuário encontra uma lista com as moléculas disponíveis no banco de dados local do sistema. Clicando no ID de cada molécula, na coluna 'PDB ID', o usuário pode fazer o download do arquivo pdb referente. Clicando na coluna 'Sequências' o usuário consegue ver a estrutura primária das moléculas. Clicando no ícone da coluna 'Show PDB', é mostrada a visualização 3D da molécula escolhida.



#	PDB ID	Show PDB	Organismo de origem	Sequências
1	1BGA		<i>Paenibacillus polymyxa</i>	A B C D
2	1BGG		<i>Paenibacillus polymyxa</i>	A B C D
3	1CBG		<i>Trifolium repens</i>	A
4	1E4I		<i>Bacillus polymyxa</i>	A
5	1GNX		<i>Streptomyces sp.</i>	A B
6	1GON		<i>Streptomyces sp.</i>	A B
7	1H4L		<i>Homo sapiens</i>	A B D E
8	1NP2		<i>Thermus nonproteolyticus</i>	A B
9	1OD0		<i>Thermotoga maritima</i>	A B
10	1OIF		<i>Thermotoga maritima</i>	A B

Figura 7: página navegar do BLASTExplorer.

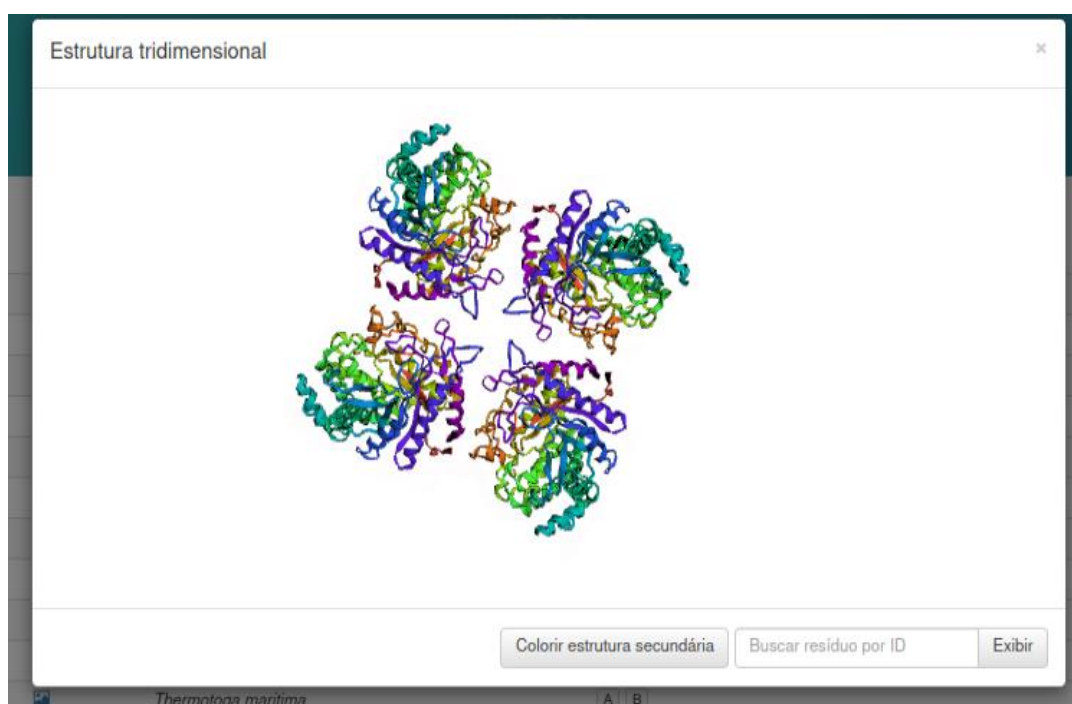


Figura 8: visualização 3D da bactéria *Paenibacillus polymyxa*.

4. **Página BLAST:** nessa página o usuário insere a sequência de proteína desejada para que o sistema faça a busca, ao inserir a sequência, basta clicar em ‘Executar BLAST’ que o sistema buscará no banco de dados e mostrará uma página com os resultados.

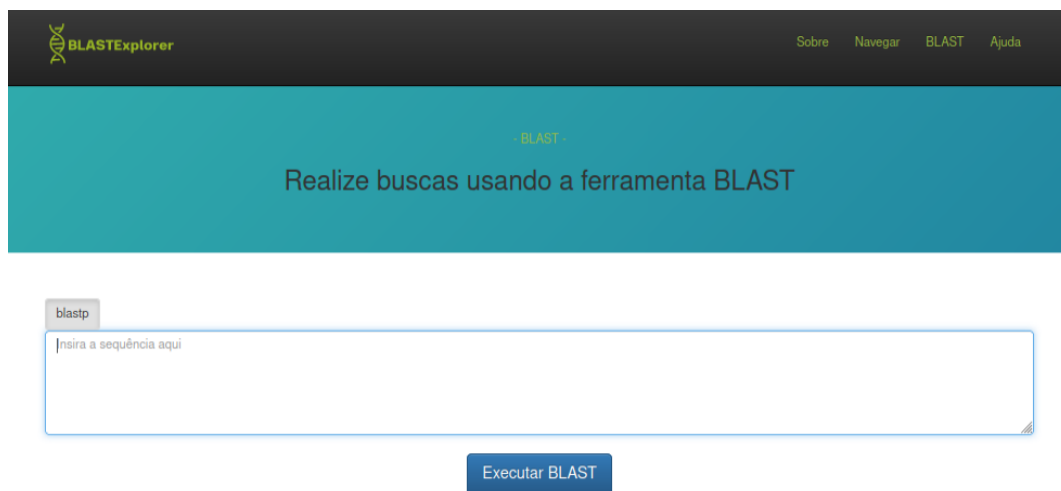


Figura 9: página de busca do BLASTExplorer.

5. **Página de resultados:** página onde é mostrado o resultado da pesquisa feita pelo usuário, nessa página é possível visualizar a sequência buscada, o arquivo pdb de cada molécula encontrada, o nome do organismo, a identidade (similaridade) e o alinhamento local das sequências, os organismos aparecem ordenados com base em sua identidade, em ordem decrescente.

Resultados (blastp)

Identidade	Sequência query	PDB Referência	Organismo referência	Identidade %	Positivos %	Gaps %	Alinhamento
██████████	☰	1H4L:B	<i>Homo sapiens</i>	292/292 (100%)	292/292 (100%)	0/292 (0%)	🔍
██████████	☰	1H4L:A	<i>Homo sapiens</i>	292/292 (100%)	292/292 (100%)	0/292 (0%)	🔍
██████████	☰	4AU8:B	<i>Homo sapiens</i>	290/291 (99%)	290/291 (99%)	0/291 (0%)	🔍
██████████	☰	4AU8:A	<i>Homo sapiens</i>	290/291 (99%)	290/291 (99%)	0/291 (0%)	🔍
██████████	☰	3OOG:B	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	3OOG:A	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNL:B	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNL:A	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNH:B	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNH:A	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNG:B	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
██████████	☰	1UNG:A	<i>Homo sapiens</i>	290/292 (99%)	291/292 (99%)	0/292 (0%)	🔍
███	☰	5BX5:A	<i>Thermoanaerobacterium xylanolyticum Ix-11</i>	28/148 (18%)	54/148 (36%)	27/148 (18%)	🔍
███	☰	5BX4:A	<i>Thermoanaerobacterium xylanolyticum Ix-11</i>	28/148 (18%)	54/148 (36%)	27/148 (18%)	🔍
███	☰	5BX3:A	<i>Thermoanaerobacterium xylanolyticum Ix-11</i>	28/148 (18%)	54/148 (36%)	27/148 (18%)	🔍
███	☰	5BX2:A	<i>Thermoanaerobacterium xylanolyticum Ix-11</i>	28/148 (18%)	54/148 (36%)	27/148 (18%)	🔍
███	☰	5BVU:A	<i>Thermoanaerobacterium xylanolyticum Ix-11</i>	28/148 (18%)	54/148 (36%)	27/148 (18%)	🔍

6. CONSIDERAÇÕES FINAIS

A bioinformática é uma área interdisciplinar que combina disciplinas como biologia, ciência da computação, engenharias, matemática e estatística para analisar dados biológicos, com o aumento exponencial da quantidade de dados biológicos disponíveis, a bioinformática tem se tornado cada vez mais importante, visto que ferramentas como o BLAST facilitam a análise de grandes volumes de dados.

O desenvolvimento do sistema BLASTExplorer proporcionou uma ferramenta grátis, acessível e eficiente para a comparação de sequências de aminoácidos, o sistema permitiu aos usuários realizar buscas utilizando o BLASTp, com resultados apresentados de forma clara e intuitiva, facilitando a análise de similaridades entre sequências biológicas e a identificação de organismos relacionados.

Os objetivos propostos foram alcançados com sucesso, desenvolvendo um sistema com uma interface amigável que permite a inserção de sequências no formato FASTA, a execução de buscas eficientes e a visualização dos resultados em detalhes. A integração com o banco de dados PDB e a criação de um banco de dados local foram elementos cruciais para o funcionamento do sistema, permitindo uma exploração das relações entre as sequências analisadas, além da visualização 3D das moléculas.

A ferramenta desenvolvida pode ser usada não só como plataforma de pesquisa, mas também como um recurso educacional valioso, o BLASTExplorer pode auxiliar estudantes e pesquisadores de diversas áreas a realizarem análises complexas sem a necessidade de um conhecimento muito profundo em bioinformática.

Para trabalhos futuros, pode-se considerar a expansão do banco de dados local para incluir uma maior variedade de sequências e a implementação de novas funcionalidades, como a comparação de sequências de DNA e RNA, com a inclusão do BLASTn. Além disso, também pode ser implementada uma função que permitisse ao usuário elaborar e analisar árvores filogenéticas das sequências desejadas.

7. REFERÊNCIAS

1. LORENZONI, Rodrigo. Bioinformática – parte 2: fundamentos e aplicações. Laborgene, 16 de out. de 2019. Disponível em: <https://www.laborgene.com.br/fundamentos-da-bioinformatica/>. Acesso em 06 de jun. de 2024.
2. Introdução à expressão gênica (dogma central). Khan Academy, 2020. Disponível em: <https://pt.khanacademy.org/science/ap-biology/gene-expression-and-regulation/translation/a/intro-to-gene-expression-central-dogma>. Acesso em 06 de jun. de 2024.
3. Bioinformática: o estudo da informação biológica. PUC PR. Disponível em: <https://blogs.pucpr.br/emcv/2023/06/07/bioinformatica/>. Acesso em: 06/06/2024.
4. AGUIAR, Carla. et al. Bioinformática como ferramenta de aprendizagem no ensino de biologia celular e molecular pela percepção do estudante. Research Gate, Vitória, 01 de fev. de 2022.
5. JUNIOR, Robson. Revisão e aplicação de técnicas de análise de dados para o processamento de dados rna-seq. Centro universitário Sagrado Coração, Bauru, 2022.
6. PROSDOCIMI, Francisco. Introdução à bioinformática. Biotecnologia, ciência e desenvolvimento, 2007.
7. OLIVEIRA, Samuel. Introdução à bioinformática estrutural: a construção de uma ementa. Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas, Inconfidentes, 2016.
8. COLLINS, F. S., et al. A vision for the future of genomics research. Nature, 2000.
9. DREWS, J. Drug discovery: a historical perspective. Science, 2013.
10. GARDNER, S. N., et al. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. Plos One.
11. Transgênicos: perguntas e respostas. Embrapa. Disponível em: <https://www.embrapa.br/tema-transgenicos/perguntas-e-respostas>. Acesso em 06 de jun. de 2024.
12. TAJARA, Eloiza, et al. Proteômica: metodologias e aplicações no estudo de doenças humanas. Scielo, jun. de 2012. Disponível em:

<https://www.scielo.br/j/ramb/a/rbzzwDkg4gWNqcPSLNMK6wD#>. Acesso em 06 de jun. de 2024.

13. BIAZON, Tássia. Genômica: a ciência que rompe fronteiras e desafia os cientistas. *Jornal da Usp*, 15 de jun. de 2018. Disponível em: <https://jornal.usp.br/ciencias/genomica-a-ciencia-que-rompe-fronteiras-e-desafia-os-cientistas/>. Acesso em 06 de jun. de 2024.

14. FREITAS, Ana. Genômica funciona e bioinformática. *GBInfo*. Disponível em: <https://web.tecnico.ulisboa.pt/ana.freitas/bioinformatics.ath.cx/bioinformatics.ath.cx/index1e35.html?id=39>. Acesso em 06 de jun. de 2024.

15. CASSANEGO, Samuel. O que é a Bioinformática. *PET Sistemas de Informação, UFSM*, 03 de dez. de 2020. Disponível em: <https://www.ufsm.br/pet/sistemas-de-informacao/2020/12/03/o-que-e-a-bioinformatica>. Acesso em 06 de jun. de 2024.

16. SANTOS, Vanessa. Dogma Central da Biologia Molecular. *Mundo Educação*. Disponível em: <https://mundoeducacao.uol.com.br/biologia/dogma-central-biologia-molecular.htm>. Acesso em: 06/06/2024.

17. SANTOS, Vanessa. DNA. *Brasil Escola*. Disponível em: <https://brasilecola.uol.com.br/biologia/dna.htm>. Acesso em 06 de jun. de 2024.

18. ATP – Adenosina trifosfato. *Biologia Net*. Disponível em: <https://www.biologianet.com/biologia-celular/atp-adenosina-trifosfato.htm>. Acesso em 06 de jun. de 2024.

19. SANTOS, Vanessa. Proteínas. *Brasil Escola*. Disponível em: <https://brasilecola.uol.com.br/biologia/proteinas.htm>. Acesso em 06 de jun. de 2024.

20. MAGALHÃES, Lana. Estrutura das Proteínas. *Toda Matéria*, [s.d.]. Disponível em: <https://www.todamateria.com.br/estrutura-das-proteinas/>. Acesso em: 6 jun. 2024.

21. HOSSEIN, Morteza. A Survey on Data Compression Methods for Biological Sequences. *Research Gate*, jun. de 2016. Disponível em: https://www.researchgate.net/figure/A-sample-of-the-Multi-FASTA-file_fig1_309134977. Acesso em: 6 jun. 2024.

22. NEIS, Alessandra. Entendendo o BLAST (parte I): conceitos principais. *Medium*, 18 de nov. de 2020. Disponível em: <https://medium.com/omixdata/entendendo-blast-parte-i-conceitos-principais-4711e34cc2b6>. Acesso em: 6 jun. 2024.

23. BLAST Software and Databases. NCBI. Disponível em:

<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html>. Acesso em: 6 jun. 2024.

24. MARIANO, Diego. Introdução à programação para bioinformática com biopython. Diego

Mariano, 15 de ago. de 2020. Disponível em: <https://diegomariano.com/blast/>. Acesso em: 6

jun. 2024.

25. MARIANO, Diego, MINARDI, Raquel de Melo. Introdução à programação web para

bioinformática. Laboratório de Bioinformática e Sistemas. 4ed. Belo Horizonte:

Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, 2017.