

UNIVERSIDADE FEDERAL DE PERNAMBUCO - UFPE
CIÊNCIAS BIOLÓGICAS - BACHARELADO

RAFAEL LUCAS DA SILVA

Plataforma DBAMPRecord: um banco de dados de sequências, estruturas e assinaturas de peptídeos antimicrobianos

RECIFE
2024

RAFAEL LUCAS DA SILVA

Plataforma DBAMPRecord: um banco de dados de sequências, estruturas e assinaturas de peptídeos antimicrobianos

Trabalho de conclusão de curso apresentado à Coordenação do curso de Bacharelado em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título em Bacharel em Ciências Biológicas.

Orientador (a): Carlos André Santos Silva
Coorientador (a): Ana Maria Benko Iseppon

RECIFE
2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

da Silva, Rafael Lucas .

Plataforma DBAMPRecord: um banco de dados de sequências, estruturas e assinaturas de peptídeos antimicrobianos / Rafael Lucas da Silva. - Recife, 2024.

47 p. : il., tab.

Orientador(a): Carlos André dos Santos Silva

Cooorientador(a): Ana Maria Benko Iseppon

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Ciências Biológicas - Bacharelado, 2024.

Inclui referências.

1. Bioinformática. 2. Proteômica. 3. Desenvolvimento de software. 4. Banco de dados. I. Silva, Carlos André dos Santos . (Orientação). II. Iseppon, Ana Maria Benko . (Coorientação). IV. Título.

600 CDD (22.ed.)

RAFAEL LUCAS DA SILVA

Plataforma DBAMPRecord: um banco de dados de sequências, estruturas e assinaturas de peptídeos antimicrobianos

Trabalho de conclusão de curso apresentado à Coordenação do curso de Bacharelado em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título em Bacharel em Ciências Biológicas.

Aprovada em: 26 /09/ 2024

BANCA EXAMINADORA

Dra. Ana Maria Benko Iseppon

Universidade Federal de Pernambuco

Dr. Jose Ribamar Costa Ferreira Neto

Universidade Federal de Pernambuco

Dr. João Pacífico Bezerra Neto

Universidade de Pernambuco

RESUMO

A crescente quantidade de dados biológicos, especialmente de peptídeos antimicrobianos (AMPs), exige ferramentas eficazes para armazenamento e recuperação dessas sequências. A falta de plataformas especializadas pode dificultar a organização e análise desses dados, essenciais para avanços em bioinformática. Para solucionar esse problema, apresentamos a DBAMPRecord, uma plataforma web desenvolvida para o armazenamento, consulta e análise de sequências de AMPs. Utilizando o PostgreSQL, um sistema de gerenciamento de banco de dados relacional, a ferramenta organiza os dados de forma escalável, permitindo consultas rápidas e precisas. O formato FASTA, amplamente utilizado em bioinformática por sua simplicidade e padronização, foi escolhido como padronização das sequências, otimizando o fluxo de inserção e recuperação de informações. A interface do sistema, desenvolvida com Flask e Bootstrap, oferece uma experiência de uso intuitiva, com funcionalidades de cadastro, busca e visualização de sequências.

Palavras-chave: Bioinformática; Interface web; FASTA.

ABSTRACT

The increasing volume of biological data, particularly antimicrobial peptides (AMPs), demands effective tools for storing and retrieving these sequences. The lack of specialized platforms can hinder the organization and analysis of such data, which is crucial for advancements in bioinformatics. To address this issue, we present DBAMPRecord, a web-based platform designed for storing, querying, and analyzing AMP sequences. Using PostgreSQL, a relational database management system, the tool organizes data in a scalable manner, enabling fast and accurate queries. The FASTA format, widely adopted in bioinformatics for its simplicity and standardization, was chosen to standardize the sequences, optimizing the flow of data insertion and retrieval. The system interface, developed with Flask and Bootstrap, provides an intuitive user experience, featuring functionalities for sequence registration, search, and visualization.

Keywords: Bioinformatics; Web interface; FASTA.

LISTA DE ILUSTRAÇÕES

Figura 1 - Arquitetura de um banco de dados biológico.....	14
Figura 2 - Exemplo de sequência em formato FASTA. A) Identificador da sequência. B) Descrição ou comentário da sequência. C) Sequência.....	22
Figura 3 - Função <code>parse_fasta</code>	24
Figura 4 - Exemplo do formato JSON gerado ao processar um arquivo multifasta com a função <code>'parse_fasta'</code>	24
Figura 5 - Representação das tabelas em formato de classes.....	25
Figura 6 - Diagrama das tabelas do banco de dados. A) Representação da tabela 'sequence_info'. B) Representação da tabela 'sequence'.....	26
Figura 7 - Relação one-to-many.....	27
Figura 8 - Exemplo de relação entre tabelas "one-to-many".....	28
Figura 9 - Demonstração em HTML da estrutura da interface do DBAMPRecord....	29
Figura 10 - Diagrama do design de interface da plataforma DBAMPRecord.....	30
Figura 11 - Tela principal. A) Área de descrição do DBAMPRecord. B) Botão "Sign in". C) Botão "Sign up".....	31
Figura 12 - Tela de acesso. A) Campo de inserção "User ID". B) Campo de inserção "Password". C) Botão "Login".....	32
Figura 13 - Tela de solicitação. A) Campo de inserção "e-mail". B) Botão "Sign up"....	33
Figura 14 - Tela do menu principal. A) Painel com o botão "Register". B) Painel com o botão "Search".....	34
Figura 15 - Tela de cadastro de sequências. A) Botão "Submit".....	35
Figura 16 - Tela de busca de sequências. A) Campo "Search the sequences". B) Menu de filtragem. C) Botão "Search".....	37
Figura 17 - Tela de visualização de sequências. A) Menu suspenso. B) Botão "Download".....	38

LISTA DE TABELAS

Tabela 1 - Lista de banco de dados de DNA.....	14
Tabela 2 - Lista de banco de dados RNA.....	16
Tabela 3 - Lista de banco de dados de proteína.....	17
Tabela 4 - Descrição do elementos da tela de cadastro de sequências.....	35
Tabela 5 - Descrição do elementos da tela de busca de sequências.....	37
Tabela 6 - Descrição do elementos da tela de visualização de sequências.....	38

LISTA DE ABREVIATURAS E SIGLAS

AMP	Peptídeos Antimicrobianos (<i>Antimicrobial Peptides</i>)
API	Application Programming Interface
CSS	Cascading Style Sheets
HTML	Hypertext Markup Language
LGBV	Laboratório de Genética e Biotecnologia Vegetal
ORM	Object-Relational Mapping

SUMÁRIO

1. INTRODUÇÃO.....	11
1.1 Objetivos.....	12
2.1.1 Objetivos geral.....	12
2.1.2 Objetivos específicos.....	12
2. REFERENCIAL TEÓRICO.....	13
2.1 Banco de dados biológicos.....	13
2.2 Diversidade dos bancos de dados biológicos.....	14
2.3.2 Banco de dados de AMPs.....	18
3. MATERIAS E MÉTODOS.....	20
3.1 Ferramentas.....	20
3.1.1 Visual Studio Code.....	20
3.1.2 pgAdmin 4.....	20
3.1.3 Python.....	20
3.1.3.1 Flask.....	21
3.1.3.2 SQLAlchemy.....	21
3.2.4 Bootstrap.....	21
3.2.4 HTML.....	21
3.3 Desenvolvimento.....	22
3.3.1 Reconhecimento de sequências em formato FASTA.....	22
3.3.2 Construção e design do banco de dados.....	25
3.3.3 Design de interface do DBAMPRecord.....	28
4. RESULTADOS.....	31
4.1 Tela principal.....	31
4.2 Tela de acesso.....	31
4.3 Tela de solicitação de acesso.....	32
4.4 Menu principal.....	33
4.5 Tela de busca de sequências.....	36

4.6 Tela de visualização de sequências.....	38
5. CONCLUSÃO.....	40
6. REFERÊNCIAS.....	41

1. INTRODUÇÃO

Os seres vivos apresentam mecanismos eficazes no processo de defesa contra vírus, bactérias e fungos (BENKO-ISEPPON *et al.*, 2010 ;LI, J. *et al.*, 2021). Dentre esses, um grupo de moléculas denominadas Peptídeos Antimicrobianos (AMPs, do inglês, *Antimicrobial Peptides*) são uma classe de moléculas bioativas do sistema imune inato, com atividades antimicrobianas, que fornecem uma linha de defesa, contra bactérias, fungos, nematóides e vírus (SANTOS-SILVA, C. A. DOS *et al.*, 2020; ZUPIN *et al.*, 2022). Os AMPs são caracterizados por seu tamanho reduzido (entre 10 e 100 resíduos de aminoácidos), e por possuírem carga superficial catiônica (HUAN *et al.*, 2020). Além disso, essas moléculas figuram entre uma das mais promissoras abordagens vislumbradas no combate à resistência aos compostos antimicrobianos (RIMA *et al.*, 2021; ZHANG *et al.*, 2021).

Junto a isso, o surgimento das abordagens multiômicas, que permitem o acesso a diferentes níveis de dados moleculares, como genômicos, transcriptômicos, proteômicos e metabolômicos, e com o crescente número de genomas sequenciados disponíveis em domínio público, o uso de bancos de dados biológicos, se mostrou como uma importante ferramenta para análises prospectivas e gerenciamento eficaz desses dados (MA, L. *et al.*, 2022). Atualmente, existem diversos bancos de dados, como o *National Center for Biotechnology Information* (NCBI) (SCHOCH *et al.*, 2020), Peptaibol (WHITMORE *et al.*, 2003), Defensins Knowledgebase (SEEBACH *et al.*, 2007) e Uniprot (UNIPROT CONSORTIUM, 2023), que são representantes de bancos especializados no armazenamento de informações biológicas de diversas naturezas biológicas, incluindo os AMPs.

Essa crescente demanda por infraestrutura robusta de dados e com o avanço ao acesso a dados multi-ômicos, como o sequenciamento de genomas, transcriptomas e metagenomas, tem gerado uma quantidade massiva de dados, incluindo sequências de AMPs (CHEN, C. *et al.*, 2020; MA, L. *et al.*, 2022). Diante disso, a criação de bancos de dados biológicos específicos para o gerenciamento dessas informações tornou-se crucial para empresas, hospitais e laboratórios de pesquisa.

Dessa forma, o Laboratório de Genética e Biotecnologia Vegetal (LGBV) tem se deparado com um grande volume de sequências de AMPs obtidas a partir do sequenciamento de plantas do Nordeste, seus patógenos e amostras ambientais. A

necessidade de armazenar, categorizar e descrever essas sequências de forma eficiente impulsionou a iniciativa de desenvolver um banco de dados dedicado, capaz de auxiliar na análise e no aproveitamento do potencial biotecnológico desses peptídeos.

1.1 Objetivos

2.1.1 Objetivos geral

Desenvolver a plataforma DBAMPRecord, uma ferramenta para o armazenamento e manuseio de sequências de AMPs identificados em plantas, seus patógenos e microbiomas, além de peptídeos bioinspirados racionalmente desenhados a partir desses AMPs

2.1.2 Objetivos específicos

- Desenvolver um algoritmo de reconhecimento de sequências em formato FASTA;
- Desenvolver um algoritmo capaz de reconhecer informações em arquivos com extensão fasta e separar as informações e sequências;
- Elaborar um banco de dados para armazenamento das sequências;
- Criar uma interface amigável com a função de cadastramento, visualização e obtenção de sequências de AMPs, disponível para navegadores de internet.

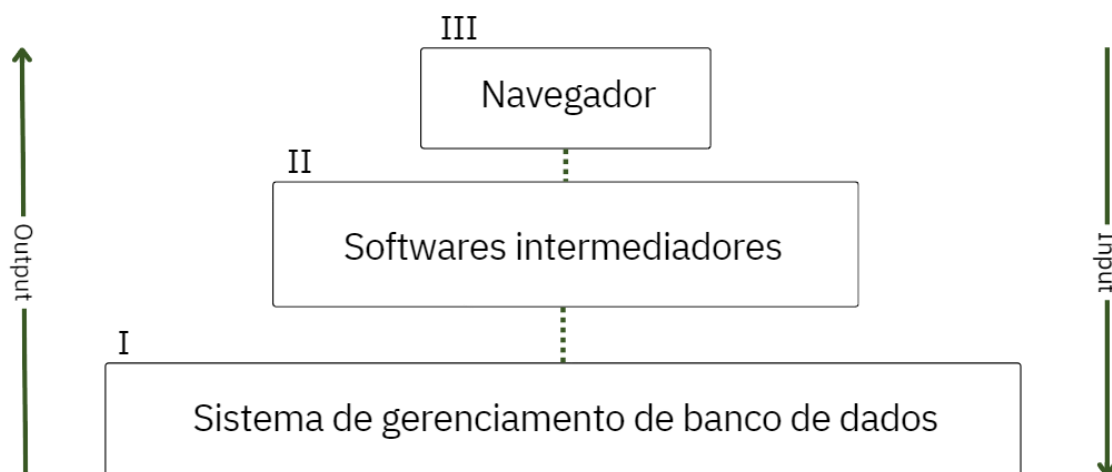
2. REFERENCIAL TEÓRICO

2.1 Banco de dados biológicos

Um banco de dados é uma coleção de dados organizados de forma a permitir o armazenamento, recuperação e manipulação eficientes desses dados. Os dados são normalmente armazenados em tabelas, com cada tabela contendo um conjunto de dados relacionados (GRAY, 1978). Em 1960, Charles W. Bachman criou o primeiro sistema de gerenciamento de banco de dados (DBMS, do inglês, *database management system*) , sistema que fornece ferramentas para criar, modificar e consultar os dados em um banco de dados (“The architecture of an active database management system”, 1989).

Os primeiros bancos de dados biológicos surgiram nas décadas de 1980 e 1990 para armazenar sequências de DNA e proteínas determinadas experimentalmente(SAPUNDZHI; DZIMBOVA, 2022). Esses bancos de dados contém informações sobre diversos aspectos da biologia, como genes, proteínas, vias metabólicas e organismos. Eles são projetados para serem pesquisáveis e geralmente incluem ferramentas que permitem a análise e visualização dos dados (STEIN, 2003).

De acordo com Stain (2003), cada banco de dados biológico é composto por três camadas de software (Figura 1) (STEIN, 2003). Na camada inferior, encontra-se o Sistema de Gerenciamento de Banco de Dados (DBMS), responsável por controlar e organizar a coleção de dados. Na camada superior, há o navegador da web, que transmite solicitações ao banco de dados e apresenta as respostas na forma de páginas da web. Entre essas duas camadas, há uma camada intermediária de *software*, que atua como mediadora entre o DBMS e o navegador. Essa camada transforma as solicitações de dados em consultas ao banco de dados e converte as respostas dessas consultas em linguagem de marcação de hipertexto (HTML), permitindo que as informações sejam exibidas de maneira adequada no navegador.

Figura 1 - Arquitetura de um banco de dados biológico

Fonte : O autor (2024)

2.2 Diversidade dos bancos de dados biológicos

De acordo com o tipo de dados armazenados e gerenciados, os bancos de dados biológicos podem ser classificados em diferentes categorias, como bancos de DNA, RNA e proteína (ZOU *et al.*, 2015).

Um banco de dados de DNA tem o objetivo de armazenar e gerenciar dados e sequências de DNA, sendo de diversas espécies ou de uma espécie específica. O NCBI RefSeq (O'LEARY *et al.*, 2016), dbSNP (SHERRY *et al.*, 2001) e Phytozome (GOODSTEIN *et al.*, 2012) são alguns exemplos de bancos de dados de DNA, mais exemplos na tabela 1. Outro exemplo representativo de banco de dados de DNA é o GenBank (BENSON, D. A. *et al.*, 2013), que desde a sua criação em 1982, cresceu extraordinariamente e em agosto de 2024, contém mais de 3 trilhões de bases de nucleotídeos em mais de 250 milhões de sequências (<http://www.ncbi.nlm.nih.gov/genbank/statistics>).

Tabela 1 - Lista de banco de dados de DNA

Nome do Banco	Descrição	Referência
NCBI RefSeq	Coleção de sequências de referência de	(O'LEARY <i>et al.</i> , 2016)

	DNA, RNA e proteínas de organismos diversos.	
dbSNP	Banco de dados de variações genéticas (SNPs) em diferentes populações e espécies.	(SHERRY <i>et al.</i> , 2001)
GenBank	Repositório de sequências de DNA de diversos organismos, com contribuições de pesquisadores de todo o mundo.	(BENSON, D. A. <i>et al.</i> , 2013)
Ensembl	Banco de dados de genomas anotados de vertebrados e outros organismos eucarióticos.	(CUNNINGHAM <i>et al.</i> , 2022)
UCSC Genome Browser	Ferramenta de visualização de genomas com acesso a dados de sequências e anotações genômicas.	(LEE, B. T. <i>et al.</i> , 2022)
GISAID	Banco de dados de sequências de genomas virais, como o SARS-CoV-2, para o estudo de pandemias.	(SHU, Yuelong; MCCAULEY, 2017)
1000 Genomes Project	Projeto que oferece um catálogo de variações genômicas de diferentes populações humanas.	(1000 GENOMES PROJECT CONSORTIUM <i>et al.</i> , 2015)
ClinVar	Banco de dados de variantes genéticas relacionadas a condições de saúde e significância clínica.	(LANDRUM <i>et al.</i> , 2018)
DGV (Database of Genomic Variants)	Repositório de variações estruturais no genoma humano.	(MACDONALD <i>et al.</i> , 2014)
EMBL-EBI	Banco europeu de dados biológicos que inclui sequências de DNA e proteínas, entre outros recursos.	(THAKUR, M. <i>et al.</i> , 2024)

Fonte: O autor (2024)

Os bancos de dados responsáveis pelo armazenamento de sequências de RNA enfrentam desafios consideráveis em termos de organização dos dados. Isso se deve ao fato de que apenas uma pequena fração do genoma das espécies é transcrita em mRNA (RNA mensageiro), que desempenha um papel crucial na síntese de proteínas, a maior parte do genoma, entretanto, é transcrita em RNAs não codificantes (ncRNAs), os quais não codificam proteínas (MA, L.; BAJIC;

ZHANG, Z., 2013). Dentro da categoria dos RNAs não codificantes, incluem-se os microRNAs (miRNAs), RNAs longos não codificantes (lncRNAs) e RNAs nucleolares pequenos (snoRNAs). Alguns exemplos notáveis de bancos de dados que armazenam tanto RNAs codificantes quanto não codificantes incluem o GENCODE (FRANKISH *et al.*, 2021), RNA-seq (YU, Y. *et al.*, 2022) e RNACentral (THE RNACENTRAL CONSORTIUM, 2019), mais exemplos na tabela 2.

Tabela 2 - Lista de banco de dados RNA

Nome do Banco	Descrição	Referência
NCBI RefSeq (RNA)	Parte do NCBI RefSeq que contém sequências de RNA de alta qualidade para uma ampla gama de organismos.	(O'LEARY <i>et al.</i> , 2016)
RNACentral	Banco de dados de sequências de RNA não codificantes, incluindo microRNAs e lncRNAs de diferentes espécies.	(THE RNACENTRAL CONSORTIUM, 2019)
miRBase	Repositório de sequências e anotações de microRNAs de várias espécies, incluindo humanos.	(KOZOMARA; BIRGAOANU; GRIFFITHS-JONES, 2019)
Rfam	Banco de dados de famílias de RNA não codificantes baseadas em similaridade de sequência e estrutura.	(KALVARI <i>et al.</i> , 2021)
NONCODE	Base de dados especializada em RNA não codificante (ncRNA), incluindo lncRNA, snoRNA, e siRNA.	(ZHAO, Y. <i>et al.</i> , 2016)
LncBook	Coleta e anota informações de lncRNAs em humanos, abrangendo expressão, função e regulação.	(LI, Zhao <i>et al.</i> , 2023)
RNAdb	Banco de dados de RNAs não codificantes com foco em transcritos funcionais em humanos e outros organismos.	(PANG, K. C. <i>et al.</i> , 2007)
GENCODE	Projeto que identifica e classifica características genéticas nos genomas humano e de camundongo com alta precisão.	(FRANKISH <i>et al.</i> , 2021)

Fonte: O autor (2024)

Na construção de bancos de dados de proteínas, diversos objetivos podem ser perseguidos, dependendo da necessidade do projeto. Entre os principais propósitos estão: a coleta de proteínas universais (por exemplo, UniProt (UNIPROT CONSORTIUM, 2023), que visa identificar e catalogar proteínas presentes em diferentes organismos; a identificação de famílias e domínios conservados (por exemplo, Pfam (MISTRY *et al.*, 2020)), essencial para o estudo de evolução e função proteica; e a criação de perfis estruturais (por exemplo, PDB (KARUPPASAMY; VENKATESWARAN; SUBBIAH, 2020)), que busca mapear as características tridimensionais de proteínas, contribuindo para a previsão de sua função e para o desenvolvimento de modelos estruturais.

Tabela 3 - Lista de banco de dados de proteína

Nome do Banco	Descrição	Referência
Pfam	Banco de dados de famílias de proteínas e domínios baseados em múltiplas sequências alinhadas e estruturas.	(MISTRY <i>et al.</i> , 2020)
PDB (Protein Data Bank)	Repositório de estruturas tridimensionais de proteínas e ácidos nucleicos resolvidas por cristalografia e NMR.	(KARUPPASAMY; VENKATESWARAN; SUBBIAH, 2020; MISTRY <i>et al.</i> , 2020)
UniProt	Banco de dados de sequências de proteínas e anotações funcionais abrangentes para uma vasta gama de organismos.	(UNIPROT CONSORTIUM, 2023)
InterPro	Repositório de domínios de proteínas, famílias e anotações funcionais, integrando dados de vários bancos.	(PAYSAN-LAFOSSE <i>et al.</i> , 2023)
PROSITE	Repositório de padrões e perfis para a identificação de domínios e motivos conservados em proteínas.	(SIGRIST <i>et al.</i> , 2010)
CATH	Banco de dados de classificação hierárquica de domínios de proteínas baseado em estrutura.	(SILLITOE <i>et al.</i> , 2021)
SCOP	Banco de dados que classifica estruturas de proteínas com base em suas semelhanças	

	estruturais e evolutivas.	
BMRB (Biological Magnetic Resonance Data Bank)	Repositório de dados de NMR para proteínas e ácidos nucleicos.	(HOCH <i>et al.</i> , 2023)
Swiss-Prot	Parte do UniProt, fornece anotações de alta qualidade para sequências de proteínas e informações funcionais.	(BOUTET <i>et al.</i> , 2007)

Fonte: O autor (2024)

2.3.2 Banco de dados de AMPs

A descoberta dos peptídeos antimicrobianos (AMPs) remonta a 1939, quando Dubos isolou um agente antimicrobiano de uma cepa de *Bacillus* encontrada no solo, eficaz na defesa de camundongos contra infecções por pneumococos (Dubos, 1939). Posteriormente Hotchkiss e Dubos (DUBOS; HOTCHKISS, 1940) identificaram nesse extrato um AMP que recebeu o nome de gramicidina, que apesar de possuir toxicidade, foi considerado eficaz no tratamento de feridas e úlceras (VAN EPPS, 2006). No ano seguinte a tirocidina foi descoberta e se mostrou efetiva contra bactérias Gram-negativas e Gram-positivas (DUBOS; HOTCHKISS, 1941).

Desde essas descobertas iniciais, o interesse e a pesquisa sobre AMPs cresceram significativamente, o com banco de dados de AMPs Data Repository of Antimicrobial Peptides (DRAMP) relatou, em sua versão de 2009, um total de 3.791 AMPs, provenientes de seis reinos biológicos, incluindo 431 de bactérias, 4 de archaea, 7 de protozoários, 6 de fungos, 824 de plantas e 2.519 de animais (KANG, X. *et al.*, 2019). A versão mais recente do DRAMP, publicada em 2022, indica um aumento significativo no número de AMPs catalogados, passando de 3.791 para 6.341 (SHI, G. *et al.*, 2022). Contudo, essa atualização não fornece detalhes adicionais sobre a origem dos novos AMPs.

Geralmente os AMPs são eficazes contra uma classe específica de microrganismos (por exemplo, bactérias ou vírus) (HANCOCK; SCOTT, 2000). Sendo assim classificados como peptídeos antibacterianos, antivirais, antifúngicos e antiparasitários.

Os representantes antibacterianos dos AMPs são os mais estudados no presente momento e sua grande maioria são peptídeos catiônicos que causam a desintegração estrutural da bicamada lipídica das membranas celulares bacterianas (SHAI, 2002; ZHANG, L.; ROZEK; HANCOCK, 2001). Esses AMPs têm entre 18 a 20 resíduos de aminoácidos com um sítio ativo para seu alvo intracelular (KRAGOL *et al.*, 2001; OTVOS *et al.*, 2000). Com foco nessas moléculas, o banco de dados YADAMP possui informações detalhadas de 2133 peptídeos ativos contra bactérias (“YADAMP: yet another database of antimicrobial peptides”, 2012)

Já os AMPs antivirais podem se juntar aos envelopes virais e causar instabilidade da membrana, modificando a estrutura do vírus e o tornando incapaz de se ligar à célula (AGAMENNONE *et al.*, 2022). Esses peptídeos também podem reduzir a ligação entre o vírus e as células hospedeiras (YASIN *et al.*, 2004) Um exemplo de banco de dados voltado para AMPs antivirais é o AVPdb, que reúne informações detalhadas sobre 2.683 peptídeos, incluindo 624 peptídeos modificados testados experimentalmente quanto à atividade antiviral (QURESHI, A. *et al.*, 2013). Outro exemplo é o DRAVP, banco de dados recente que inclui 64 peptídeos antivirais obtidos de dados clínicos (LIU, Y. *et al.*, 2023).

Os AMPs antifúngicos exercem sua ação letal sobre fungos ao comprometer a estrutura da parede celular ou estruturas intracelulares (FERNÁNDEZ DE ULLIVARRI *et al.*, 2020). A quitina, um dos principais constituintes das paredes celulares fúngicas, é uma das estruturas-alvo desses peptídeos. Diversos estudos demonstram que os AMPs antifúngicos têm a capacidade de se ligar à quitina (PUSHPANATHAN *et al.*, 2012; YOKOYAMA *et al.*, 2009), o que lhes permite atuar de maneira eficiente contra as células fúngicas. Embora, até o momento, não existam bancos de dados específicos dedicados exclusivamente a peptídeos antifúngicos, o DRAMP (Database of Antimicrobial Peptides) contém atualmente no mínimo 1.459 AMPs com ação antifúngica (AGRAWAL *et al.*, 2018; SHI, G. *et al.*, 2022).

Em relação com as outras três classes de AMP, os peptídeos antiparasitários se caracterizam como um grupo menor (BAHAR; REN, 2013). Esses peptídeos matam as células interagindo diretamente com a membrana celular (PARK *et al.*, 2004). ParaPep é um repositório de peptídeos antiparasitários, que fornece informações abrangentes relacionadas a sequências de peptídeos antiparasitários validadas experimentalmente e suas estruturas. Os dados foram coletados e

compilados de artigos de pesquisa publicados, patentes e de vários bancos de dados. A versão atual do ParaPep contém 863 entradas, entre as quais 519 são peptídeos não redundantes (MEHTA *et al.*, 2014).

3. MATERIAS E MÉTODOS

Este trabalho consiste no desenvolvimento de um aplicativo web acessível através de navegadores de internet, visando atender às necessidades de armazenamento e recuperação de sequências de peptídeos antimicrobianos (AMPs). A plataforma foi projetada para facilitar a interação dos usuários com o sistema, permitindo um fluxo de trabalho intuitivo e eficiente.

3.1 Ferramentas

Nesta seção, apresentamos as ferramentas e tecnologias utilizadas no desenvolvimento da plataforma DBAMPRecord, detalhando suas funções e como contribuem para o alcance dos objetivos propostos.

3.1.1 Visual Studio Code

O Visual Studio Code (VS Code) é um editor de código-fonte desenvolvido pela Microsoft, amplamente utilizado para o desenvolvimento de aplicações em diversas linguagens de programação. Optamos pela versão 1.80.2 devido ao seu suporte robusto para Python, HTML, CSS e JavaScript, além de extensões que facilitam o gerenciamento de pacotes e a depuração do código.

3.1.2 pgAdmin 4

O pgAdmin 4, versão 6.12, é uma ferramenta de administração web para PostgreSQL. Ele é essencial para a gestão e manutenção do banco de dados relacional utilizado na plataforma. Com o pgAdmin, é possível criar, editar e consultar tabelas e registros de forma visual, além de realizar testes em consultas SQL.

3.1.3 Python

Python é uma linguagem de programação de alto nível, interpretada e versátil, que foi utilizada para desenvolver a lógica do servidor da aplicação. A escolha do

Python deve-se à sua simplicidade e riqueza de bibliotecas, que facilitam tarefas como manipulação de dados e integração com bancos de dados. Além disso, a linguagem é amplamente utilizada na bioinformática, o que a torna uma opção natural para o presente projeto.

3.1.3.1 *Flask*

O Flask, versão 2.2.3, é um *framework* para Python que permite a criação rápida de aplicações web. Flask foi escolhido por sua flexibilidade e simplicidade, permitindo que os desenvolvedores criem rotas, gerenciem sessões e integrem funcionalidades com facilidade.

3.1.3.2 *SQLAlchemy*

SQLAlchemy é uma biblioteca de mapeamento objeto-relacional (ORM) para Python que facilita a interação com bancos de dados. Utilizamos o SQLAlchemy para simplificar a criação e manipulação de tabelas e registros no PostgreSQL, eliminando a necessidade de escrever consultas SQL complexas. Através do SQLAlchemy, o desenvolvimento do banco de dados foi realizado de forma mais eficiente e organizada.

3.2.4 Bootstrap

O Bootstrap, versão 5.1.3, é um framework front-end que fornece componentes prontos e responsivos para o desenvolvimento de interfaces web. Utilizamos o Bootstrap para garantir que a interface do DBAMPRecord seja esteticamente agradável e funcional em diversos dispositivos. A biblioteca permite a criação de layouts fluidos e interativos, melhorando a experiência do usuário.

3.2.4 HTML

HTML (*Hypertext Markup Language*) na versão 5, é a linguagem padrão para a criação de páginas web. O HTML foi usado para estruturar o conteúdo da interface do DBAMPRecord, permitindo que os usuários interajam com a aplicação de

maneira intuitiva. A combinação desta tecnologia em conjunto com o Bootstrap, assegura que a interface seja responsiva e acessível em diferentes dispositivos, facilitando a visualização e manipulação de dados.

3.3 Desenvolvimento

3.3.1 Reconhecimento de sequências em formato FASTA

Na bioinformática, a formatação FASTA é um formato baseado em texto usado para representar tanto sequências de nucleotídeos quanto sequências de peptídeos. A primeira linha de uma entrada de sequência consiste em um ">", seguido por um identificador, que não contém espaços em branco (Figura 2, A), que pode ser seguido por um espaço em branco e um comentário ou descrição (Figura 2, B). Em seguida uma ou mais linhas de sequências (Figura 2, C) (MARKEL; LEÓN, 2003).

Figura 2 - Exemplo de sequência em formato FASTA. A) Identificador da sequência. B) Descrição ou comentário da sequência. C) Sequência.

A >NC_001224.1:24156-24870,24906-25255	B Saccharomyces cerevisiae S288c mitochondrion.
--	---

C

```

ATATTAATATTATTAATAATAATTTTAACTAATAATAAGTTTTATAGAACTTTATATTATTATTAA
TATTTAATTTTCAATTAATATCTCTTTTGGGGTTCCGGTCCCTGGTCCGCCCCGAACTAAAGATAT
TAAGAATTTATATGAATCAATTATAAATAATTATATTAATTTTTAAATAAATATCTTATTAATATTAAT
AAAGATAATATTAATAAAATTTTAAATTTTATAGATAATTATACTGAAGAAGAAAAAGGTTATTATTTATCTG
GATTATTTGAAGGAGATGGTAATATTTATACTAGATGTTTTTCAATTACTTTTTCTTTAGAAGATGTTTT
ATTAGCTAATTATTTATGTCTTTATTTTAAAAATTGGTCATATTACAGCTAAATATAATTTTAATAAGAA
TTAACAGCTGTAAATGAAATATTATAAAAAAAGAAACAAGAAGTATTTATAAATTATATTAATGGTA
AATTATTAACATATAAAAGATATGATCAATATTTTAAATATAATTTTAATAATCGTTTAAATATTAATTT
ATTAAACCTAAAGAATTTGATTTACTATTAAATCCTTGATTAACAGGTTTTAATGATGCTGATGGTTAT
TTTTATCTAGGTTTTCAAAAACATAAAAAATAGTCAATGATTAATTTTCATTTAGAATTATCACAAAAAG
ATAGTTATATTTTAGATATTATTAATAAATATTTTAACTTGGTGGTATTTTAAAAAGAGATTATAAATC
TGGTGCTACAGCTTATATTTATAAGCTCAATCATCAAAAGCTATAAACCTTTTATTGAATATTTTAAT
AATTATCAACCATTAAGTCTTAGAAGATATAACAATATTTATTATTAAATATTGCTTACTTATTAATTT
TAAATAAATTACATATATTACTTAATTCCTTTATTAATATTAAGAATTAATATTATTACAAAGTGTTAA
AAATATATCTTTAGAAATAAAAAATGAATTAATAATAGAGTTAAAATTATTATTAATAAACTTCATTAT
ACAATATCGAATAA
        
```

Fonte: O autor (2024).

A simplicidade do formato FASTA torna mais fácil manipular e analisar sequências usando ferramentas de processamento de texto e linguagens de programação como o Python.

Para o reconhecimento e processamento de sequências fornecidas pelos usuários na plataforma DBAMPRecord, com o objetivo de preparar os dados para armazenamento em um banco de dados, um algoritmo foi desenvolvido em Python encapsulado na função `'parse_fasta'` (Figura 3). Este algoritmo transforma sequências do formato FASTA em um formato de dicionário, semelhante ao JSON, que é um formato leve de intercâmbio de dados amplamente usado para armazenamento e transmissão de dados.

A função `'parse_fasta'` começa verificando se a entrada é um caminho para um arquivo (Figura 3, A). Caso a condição seja verdadeira, o algoritmo abre e lê o conteúdo do arquivo linha por linha. Quando encontra uma linha começando com o caractere `'>'`, ele identifica essa linha como o início de uma nova sequência. O identificador e a descrição da sequência são extraídos, e os dados da sequência subsequente são acumulados até que uma nova sequência seja encontrada ou o fim do arquivo seja alcançado. As informações de cada sequência são então armazenadas em um dicionário, que é convertido para JSON e então recarregado em uma estrutura de dados Python para garantir a consistência. O resultado é uma lista de dicionários, cada um representando uma sequência com seu identificador, descrição e dados de sequência, representado na Figura 4.

Se a entrada for uma cadeia de caracteres (Figura 3, B), o processo de transformação é semelhante, mas adaptado para manipular o texto diretamente. O texto é primeiro dividido em blocos com base em quebras de linha que precedem identificadores de sequência (indicados pelo caractere `'>'`). Cada bloco é então processado para extrair o identificador, a descrição e a sequência, que são limpos e organizados em dicionários. Esses dicionários são então convertidos para JSON e recarregados em uma estrutura de dados Python. O resultado final também é uma lista de dicionários, estruturados para facilitar o armazenamento e a manipulação das informações.

Figura 3 - Função parse_fasta. A) Porção do código responsável pelo processamento de arquivos FASTA. B) Porção do código responsável pelo processamento de cadeia de caracteres em formato FASTA.

A

```

65 def parse_fasta(dir:str) -> Any:
66     data:list = []
67
68     if os.path.isfile(dir):
69         with open(dir, 'r') as f:
70             seq_id:str = ''
71             seq_desc:str = ''
72             seq:str = ''
73
74             for line in f:
75                 if line.startswith('>'):
76                     if seq_id != '':
77                         data.append({'id': seq_id[1:], 'description': seq_desc, 'sequence':
seq})
78
79                         seq = ''
80                         seq_id, seq_desc = line.strip().split(maxsplit=1)
81                         seq_id = seq_id[1:]
82                     else:
83                         seq += line.strip()
84
85                 data.append({'id': seq_id, 'description': seq_desc, 'sequence': seq})
86                 json_data:str = json.dumps(data)
87                 data_list:Any = json.loads(json_data)
88                 return data_list

```

B

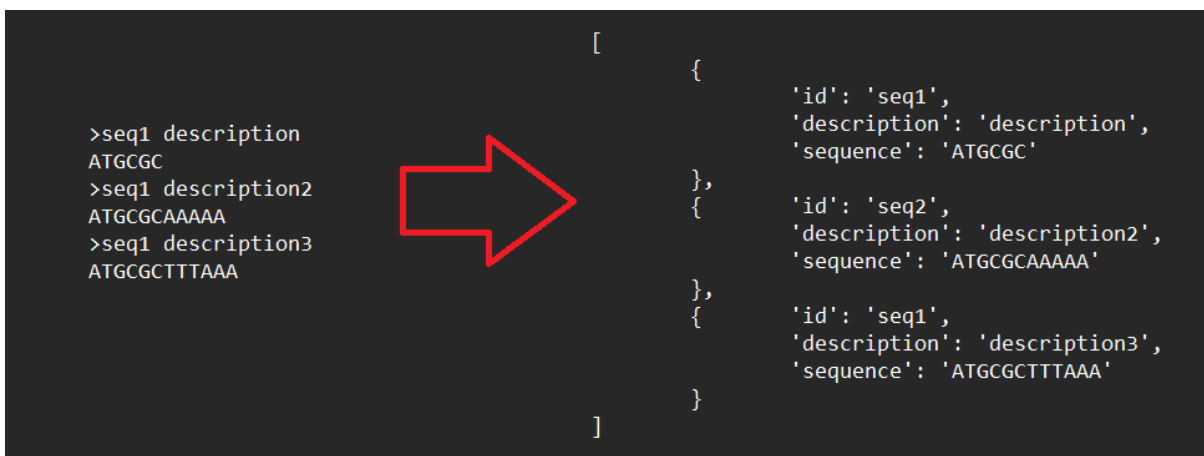
```

89     else:
90         text_area_data:str = dir
91         sequences = text_area_data.strip().split('\n>')
92
93         seq_dicts = []
94         for seq in sequences:
95
96             seq_id, *seq_desc = seq.split('\n')[0].lstrip('>').split(' ', maxsplit=1)
97             seq_desc = seq_desc[0] if seq_desc else ''
98
99
100             seq_data = ''.join(seq.split('\n')[1:]).upper().replace(' ', '')
101             seq_data = ''.join(filter(str.isalpha, seq_data))
102
103
104             seq_dict = {'id': seq_id, 'description': seq_desc, 'sequence': seq_data}
105             seq_dicts.append(seq_dict)
106
107             json_data:str = json.dumps(seq_dicts, indent=4)
108             data_list:Any = json.loads(json_data)
109             return data_list

```

Fonte: O autor (2024)

Figura 4 - Exemplo do formato JSON gerado ao processar um arquivo multifasta com a função 'parse_fasta'.



Fonte: O autor (2024).

3.3.2 Construção e *design* do banco de dados.

Para a construção do banco de dados, que armazena as sequências de AMPs, foi usado o sistema de gerenciamento de banco de dados (DBMS) PostgreSQL versão 14.5, um DBMS que usa a linguagem padrão SQL para manipulação de registros em bancos de dados relacionais. Neste projeto, as tabelas do banco de dados foram construídas utilizando o ORM (*Object-Relational Mapping*) SQLAlchemy (biblioteca Python), que faz o mapeamento entre objetos de uma linguagem orientada a objetos e tabelas relacionais no banco de dados, simplificando o trabalho eliminando a necessidade de escrever consultas SQL complexas.

O design do banco de dados foi estruturado em duas tabelas, construídas usando a linguagem Python com auxílio do ORM SQLAlchemy (Figura 5), que foram desenvolvidas para armazenar as informações fornecidas pelo usuário. Essa estrutura é ilustrada no diagrama da Figura 6.

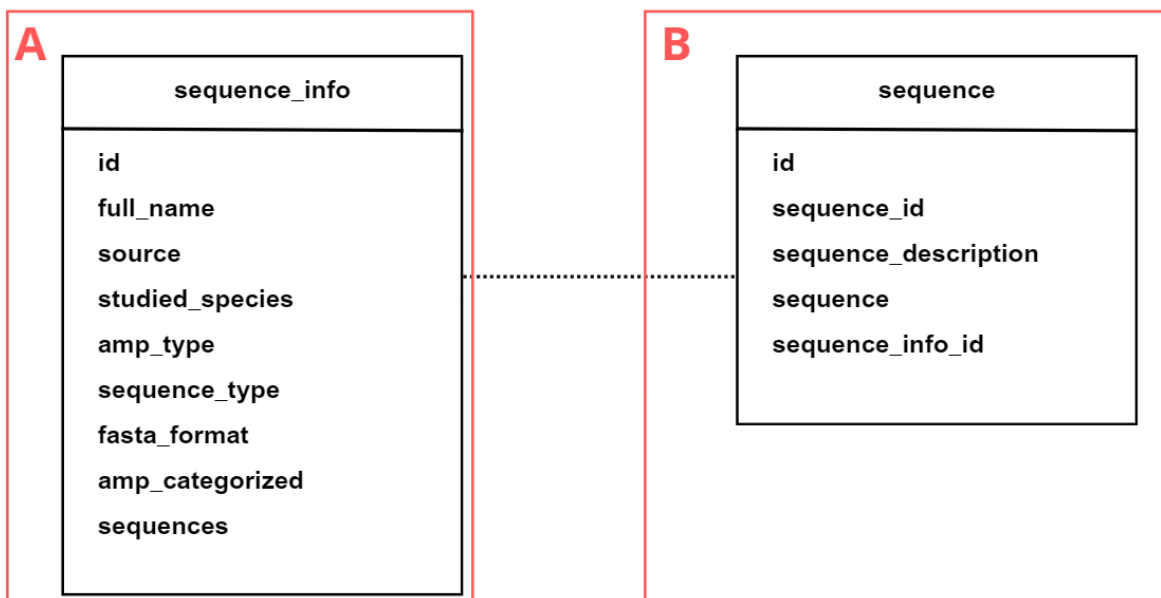
Figura 5 - Representação das tabelas em formato de classes

```

9 class Sequence(db.Model):
10     id = db.Column(db.Integer, primary_key=True)
11     sequence_id = db.Column(db.String(150))
12     sequence_description = db.Column(db.String(150))
13     sequence = db.Column(db.String(10000))
14     sequence_info_id = db.Column(db.Integer, db.ForeignKey('sequence_info.id'))
15
16 class SequenceInfo(db.Model):
17     id = db.Column(db.Integer, primary_key=True)
18     full_name = db.Column(db.String(150))
19     source = db.Column(db.String(150))
20     studied_species = db.Column(db.String(150))
21     amp_type = db.Column(db.String(150))
22     sequence_type = db.Column(db.String(150))
23     fasta_format = db.Column(db.String(150))
24     amp_categorized = db.Column(db.String(150))
25     sequences = db.relationship('Sequence')

```

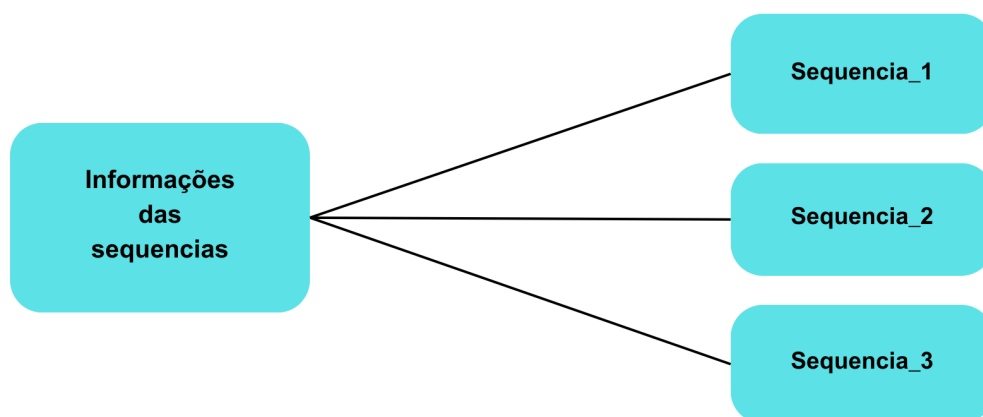
Fonte: O autor (2024)

Figura 6 - Diagrama das tabelas do banco de dados. A) Representação da tabela 'sequence_info'. B) Representação da tabela 'sequence'.

Fonte: O autor (2024).

No modelo de banco de dados relacional criado com SQL, é possível estabelecer relacionamentos entre as tabelas, permitindo que informações de uma tabela possam ser acessadas a partir de outra. O modelo de relacionamento escolhido para este trabalho foi o “*one-to-many*” (um-para-muitos), onde é determinado que um item de uma tabela pode se relacionar com vários itens de uma outra tabela (Figura 7). Na tabela nomeada ‘sequence_info’ (Figura 6, A), foram definidas oito colunas onde todas, com exceção da coluna ‘id’, são preenchidas com as informações inseridas pelo usuário sobre as características das sequências a serem armazenadas no banco de dados. A partir dessas características, é possível acessar suas sequências. Por exemplo, ao filtrar a busca pelo nome de uma espécie, todas as sequências de AMPs associadas a essa espécie serão retornadas como resultado.

Figura 7 - Relação *one-to-many*



Fonte: O autor (2024)

Por fim, na tabela nomeada ‘sequence’ (Figura 6, B) foram criadas cinco colunas para armazenar os dados das sequências. Dentre elas, destaca-se a coluna ‘sequence_info_id’, que armazena os identificadores associados à tabela ‘sequence_info’ (Figura 6, A) (especificamente os valores da coluna id). Isso

estabelece a relação entre os itens das duas tabelas, conforme demonstrado na (Figura 8).

Figura 8 - Exemplo de relação entre tabelas “one-to-many”

id	full_name	source	studied_species	amp_type	sequence_type	fasta_format	amp_categorized
Filtro	Filtro	Filtro	Calo	Filtro	Filtro	Filtro	Filtro
11	Wilson Dias de Oliveira	LGBV	Calotropis procera	TLP	aminoacid sequence	multifasta	yes

id	sequence_id ▼	sequence_description	sequence	sequence_info_id
Filtro	Filtro	Filtro	Filtro	11
23	ptlpcp	1	VGPGILPFGWRGRFWGRFGC	11
24	ptlpcp	2	WPGILCPGGSRLRFFGRTGWC	11
25	ptlpcp	3	FILGPGWRGRFFGRTGCGSC	11
26	ptlpcp	4	WLPPGSFGRFCGRTGCGSCL	11

Fonte: O autor (2024).

3.3.3 Design de interface do DBAMPRecord

No presente trabalho foi proposta a criação de uma interface de usuário para navegadores de internet. Diante dessa demanda foi utilizado o *framework web* de linguagem Python, Flask versão 2.2.3, que permite a criação de sites, aplicativos *web* e APIs (*Application Programming Interface*) de forma rápida e eficiente. Utilizando HTML em conjunto com o *framework web* Bootstrap versão 5.1.3 para o desenvolvimento dos componentes da interface da aplicação, conforme demonstrado na Figura 9.

O design da interface foi organizado em sete telas principais, como ilustrado no diagrama de uso na figura Figura 10. A partir da tela inicial, o usuário pode acessar a tela de acesso, onde são solicitados nome de usuário e senha para autenticação no banco de dados. Caso o usuário não possua credenciais, há uma opção para solicitá-las por meio de uma tela específica.

Após a autenticação, o usuário é direcionado ao menu principal, que oferece duas funcionalidades centrais: o cadastramento de novas sequências de AMPs e a

busca por sequências já cadastradas. Após realizar uma busca, os resultados são exibidos em uma tela dedicada, onde as sequências correspondentes à pesquisa são apresentadas com seus detalhes.

Figura 9 - Demonstração em HTML da estrutura da interface do DBAMPRecord

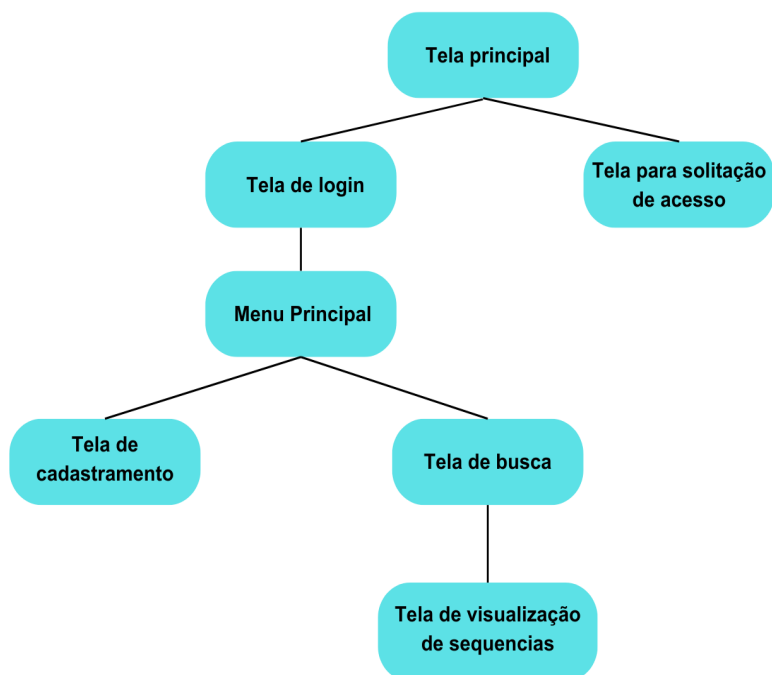
```

1  {% extends 'base.html'%}{% block title%}Home{% endblock%}
2  {% block welcome %}
3      <div class="px-4 py-5 my-5 text-center">
4          <h1 class="display-5 fw-bold text-body-emphasis"> Welcome to DBAMP Record</h1>
5          <div class="col-lg-6 mx-auto">
6              <p class="lead mb-4">
7
8                  DBAMP Record Platform is a database of sequences, structures and
signatures of antimicrobial
9                  peptides obtained at the Laboratory of Plant Genetics and Biotechnology
at the Federal University of Pernambuco.
10
11              </p>
12  {% endblock%}
13
14  {% block content%}
15  <div class="accordion" id="accordionFlushExample">
16      <div class="accordion-item">
17          <h2 class="accordion-header">
18              <button class="accordion-button collapsed" type="button" data-bs-
toggle="collapse" data-bs-target="#flush-collapseTwo" aria-expanded="false" aria-
controls="flush-collapseTwo">
19                  About us
20              </button>
21          </h2>
22          <div id="flush-collapseTwo" class="accordion-collapse collapse" data-bs-
parent="#accordionFlushExample">
23              <div class="accordion-body">
24                  The Laboratory of Plant Genetics and Biotechnology (LGBV) belongs to the
Department of Genetics at
25                  the Federal University of Pernambuco (UFPE) in Recife, Brazil. The group
is involved with research, teaching
26                  and extension involving various aspects of plant biology, with emphasis
on genetics, evolution and omics, as
27                  well as with possible uses and applications of molecules of plant origin.
28              </div>
29          </div>
30      </div>
31  </div>
32
33  {% endblock%}

```

Fonte: O autor (2024)

Figura 10 - Diagrama do uso de interface da plataforma DBAMPRecord



Fonte: O autor (2024)

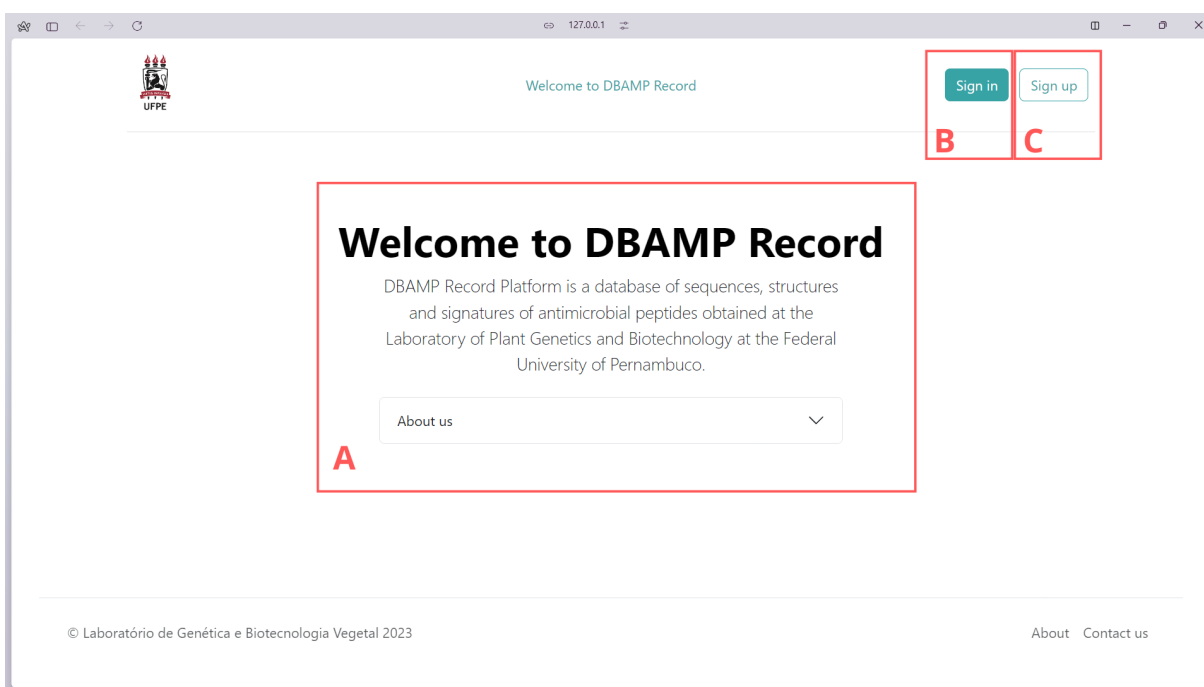
4. RESULTADOS

Neste capítulo será apresentado a aplicação DBAMPRecord desenvolvida a partir da metodologia descrita no capítulo anterior, detalhando as funcionalidade, execução da aplicação e discussão sobre o resultado.

4.1 Tela principal

Na tela principal (Figura 11), o usuário terá acesso a uma descrição detalhada do DBAMPRecord, bem como informações sobre o LGBV (Figura 11, A). Além disso, há dois botões interativos: "Sign In" (Figura 11, B), que redirecionará o usuário para a tela de acesso e "Sign Up" (Figura 11, C), que o redireciona para tela solicitação de acesso.

Figura 11 - Tela principal. A) Área de descrição do DBAMPRecord. B) Botão "Sign in". C) Botão "Sign up".

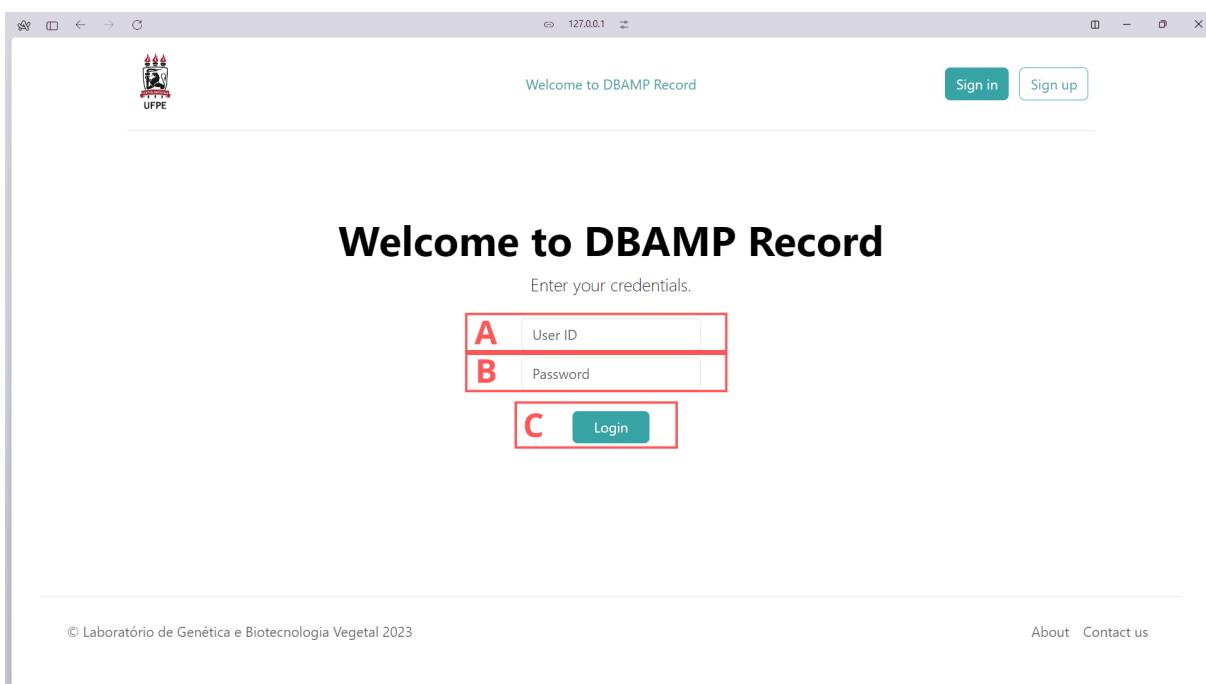


Fonte: O autor (2024)

4.2 Tela de acesso

Na tela de acesso (Figura 12), o usuário encontrará um campo de *input* editável para a identificação do usuário (Figura 12, A) e outro para a senha (Figura 12, B). Adicionalmente, há o botão “Login” (Figura 12, C) que, após a autenticação, direcionará o usuário para a tela do menu principal.

Figura 12 - Tela de acesso. A) Campo de inserção “User ID”. B) Campo de inserção “Password”. C) Botão “Login”.



Fonte: O autor (2024)

4.3 Tela de solicitação de acesso

Na tela de solicitação de acesso (Figura 13), o usuário encontrará um campo de *input* para inserir o e-mail (Figura 13, A). Adicionalmente, há o botão identificado como “Sign Up” (Figura 13, B) com a função de enviar a solicitação de permissão de acesso. Em seguida, o usuário receberá instruções por e-mail para obter o login e senha para acessar o aplicativo.

Figura 13 - Tela de solicitação. A) Campo de inserção “e-mail”. B) Botão “Sign up”.

UFPE

Welcome to DBAMP Record

Sign in Sign up

Welcome to DBAMP Record

We will contact you by email to verify your permission to access the database.

A Email

B Sign up

© Laboratório de Genética e Biotecnologia Vegetal 2023

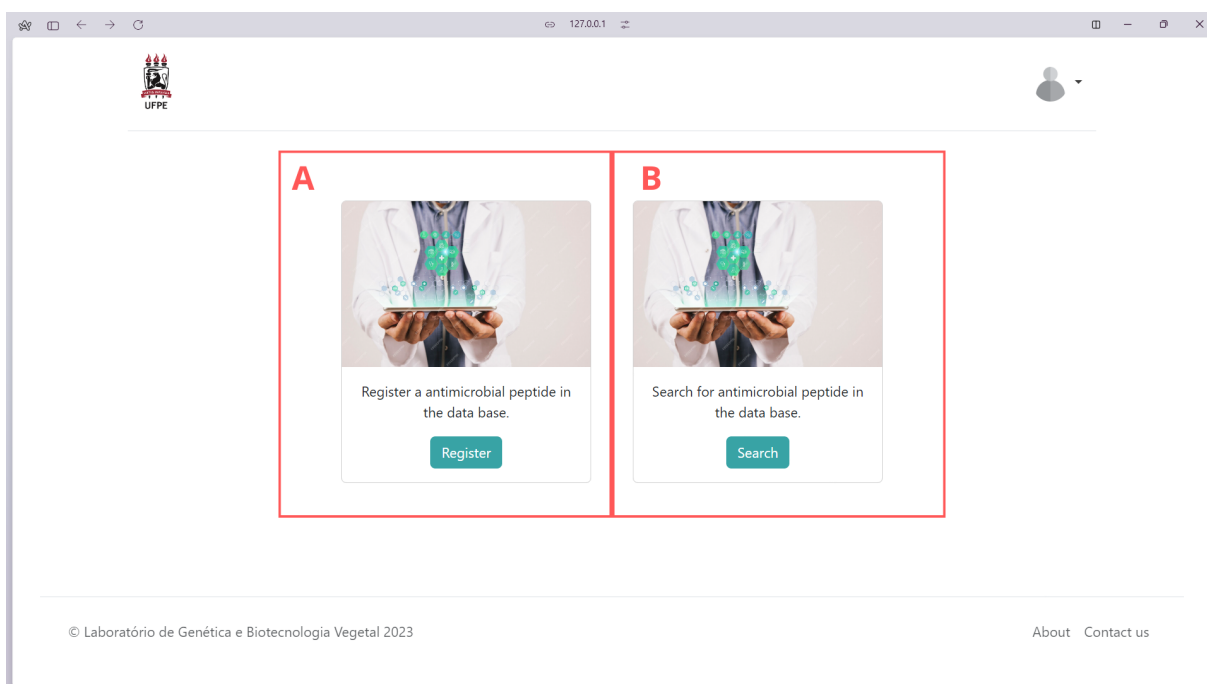
About Contact us

Fonte: O autor (2024)

4.4 Menu principal

Na tela do menu principal (Figura 14), o usuário encontrará dois painéis. O primeiro painel (Figura 14, A) possui o botão “*Register*” que direciona para a tela de cadastramento de sequências. O segundo painel (Figura 14, B) possui o botão “*Search*” que leva para a tela de busca de sequências.

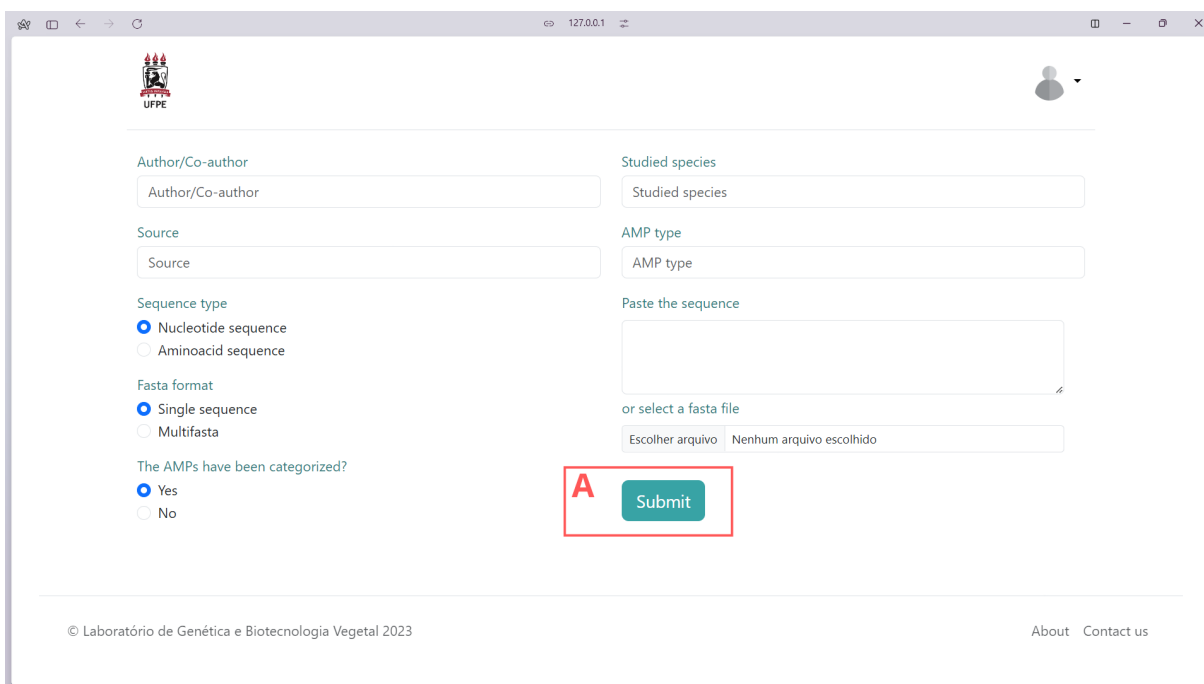
Figura 14 - Tela do menu principal. A) Painel com o botão “Register”. B) Painel com o botão “Search”.



Fonte: O autor (2024)

6.5 Tela de cadastro de sequências

Na tela de cadastro de sequências (Figura 15), o usuário encontrará os campos de inserção de sequência dos AMPs e suas características, que estão contidas e descritas na tabela 4 e ao final da tela o botão “*Submit*” (Figura 15, A), responsável por salvar as informações fornecidas pelo usuário no banco de dados.

Figura 15 - Tela de cadastro de sequências. A) Botão “Submit”.


The screenshot shows a web browser window with a sequence registration form. The form includes several input fields and radio button options. A red box labeled 'A' highlights the 'Submit' button.

Form fields and options:

- Author/Co-author:** Input field with placeholder text 'Author/Co-author'.
- Studied species:** Input field with placeholder text 'Studied species'.
- Source:** Input field with placeholder text 'Source'.
- AMP type:** Input field with placeholder text 'AMP type'.
- Sequence type:** Radio buttons for 'Nucleotide sequence' (selected) and 'Aminoacid sequence'.
- Fasta format:** Radio buttons for 'Single sequence' (selected) and 'Multifasta'.
- The AMPs have been categorized?:** Radio buttons for 'Yes' (selected) and 'No'.
- Paste the sequence:** Text area for pasting the sequence.
- or select a fasta file:** File selection area with buttons 'Escolher arquivo' and 'Nenhum arquivo escolhido'.
- Submit:** Green button highlighted by a red box and labeled 'A'.

Footer: © Laboratório de Genética e Biotecnologia Vegetal 2023. About Contact us

Fonte: O autor (2024)

Tabela 4 - Descrição do elementos da tela de cadastro de sequências

Campo de inserção	Descrição
‘Author/Co-author’	Campo destinado à inserção de autoria sequência a ser cadastrada.
‘Source’	Origem do dado cuja sequência FASTA foi originada.
‘Sequence type’	Seção destinada à indicação de se a sequência será composta por ácidos nucleicos ou aminoácidos.
‘Fasta Format’	Seção destinada à especificação de se o input FASTA conterá uma única sequência ou múltiplas sequências.
‘The AMPs have been categorized?’	Seção destinada à verificação de anotação prévia da sequência.

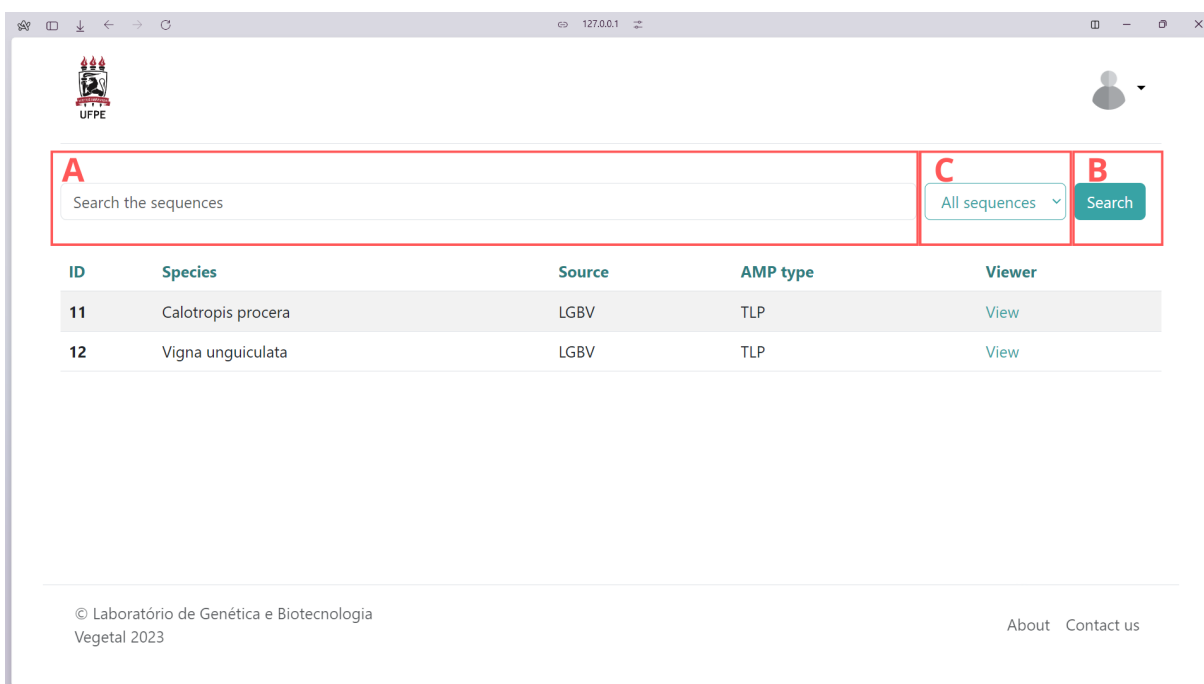
'Studied species'	Área destinada para o nome da espécie na qual o AMP foi obtido.
'Amp type'	Campo especificado à família do AMP a ser cadastrado.
'Paste the sequence'	Espaço alocado para informar a sequência a ser armazenada.
'Select a fasta file'	Botão responsável em receber input do arquivo fasta contendo a sequência a ser cadastrada.

Fonte: O autor (2024)

4.5 Tela de busca de sequências

Na tela de busca de sequências (Figura 16), será mostrado o campo para a pesquisa das sequências armazenadas e a tabela com o resultado da pesquisa. O campo "Search the sequences" (Figura 16, A) recebe informação do AMP (por exemplo, nome da espécie) o qual o usuário deseja obter a sequência. Ao lado esquerdo está o menu de seleção para a filtragem da pesquisa (Figura 16, B) e o botão "Search" (Figura 16, C) responsável por carregar as informações resgatadas do banco de dados. Os resultados são exibidos na tabela de resultados, descrita na tabela 5.

Figura 16 - Tela de busca de sequências. A) Campo “Search the sequences”. B) Menu de filtragem. C) Botão “Search”.



Fonte: O autor (2024)

Tabela 5 - Descrição do elementos da tela de busca de sequências

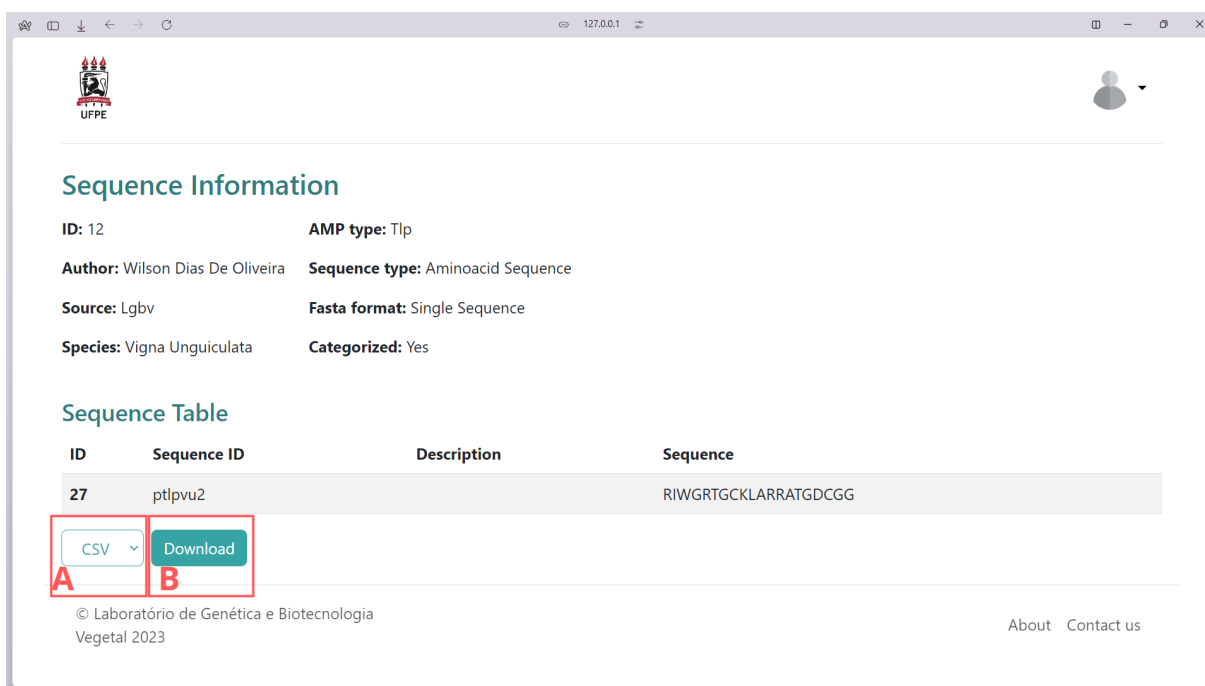
Coluna	Descrição
'ID'	Número de identificação do armazenamento da informação de uma ou mais sequências.
'Species'	Espécie no qual o AMP foi obtido.
'Source'	Origem da sequência armazenada.
'AMP type'	Família do AMP armazenado.
'Viewer'	Coluna responsável por encaminhar o usuário à tela de visualização de sequências, baseado no número de identificação do armazenamento.

Fonte: O autor (2024)

4.6 Tela de visualização de sequências

Na tela de visualização de sequências (Figura 17), serão exibidas as informações recuperadas do banco de dados sobre o armazenamento de uma ou mais sequências. Além disso, serão mostrados os detalhes de anotações da sequência, conforme obtidos no arquivo FASTA, incluindo o identificador, a descrição, e a sequência de ácidos nucleicos ou aminoácidos, conforme detalhado na Tabela 3. Na parte inferior da tela, estará disponível um menu para selecionar o formato de arquivo desejado para download (Figura 17, A), assim como o botão "Download" (Figura 17, B) para realizar a exportação das sequências.

Figura 17 - Tela de visualização de sequências. A) Menu suspenso. B) Botão "Download".



Fonte: O autor (2024)

Tabela 6 - Descrição do elementos da tela de visualização de sequências

Coluna	Descrição
'ID'	Identificação do armazenamento da sequência no banco de dados,.

'Sequence ID'	Identificador da sequência obtido no FASTA da sequência.
'Description'	Descrição da sequência informado na sequência FASTA.
'Sequence'	Sequência de aminoácidos ou de nucleotídeos.

Fonte: O autor (2024)

5. CONCLUSÃO

A plataforma DBAMPRecord foi desenvolvida com sucesso, atingindo seu objetivo principal, ser capaz de armazenar e recuperar sequências de peptídeos antimicrobianos (AMPs) de forma eficiente. A plataforma é capaz de armazenar um número ilimitado de sequências de AMPs, apenas dependendo da capacidade de armazenamento onde a aplicação está hospedada, mostrando-se eficaz no gerenciamento de informações complexas e volumosas. Além disso, o design do banco de dados, utilizando o modelo de relacionamento *one-to-many*, facilitou a organização dos dados, permitindo consultas rápidas e precisas, como a busca por espécie ou tipo de AMP. A interface web, construída com o auxílio da *framework* Bootstrap, foi avaliada quanto à usabilidade e demonstrou ser intuitiva e de fácil navegação. O uso de tecnologias como Python, Flask e Bootstrap garantiu que a aplicação fosse responsiva e funcional, facilitando a interação com o banco de dados. Além disso, a integração de funcionalidades de *upload* de arquivos FASTA simplificou o processo de inserção de dados. Por fim, a DBAMPRecord provou ser uma ferramenta eficaz e escalável para o armazenamento e análise de dados de AMPs, contribuindo significativamente para a organização desses dados e oferecendo uma solução acessível para os membros do Laboratório de Genética e Biotecnologia Vegetal. Futuras melhorias podem incluir a integração com algoritmos de predição de AMPs ou a expansão para incluir outros tipos de peptídeos bioativos.

6. REFERÊNCIAS

1000 GENOMES PROJECT CONSORTIUM *et al.* A global reference for human genetic variation. **Nature**, 1 out. 2015. v. 526, n. 7571, p. 68–74.

AGAMENNONE, M. *et al.* Antiviral Peptides as Anti-Influenza Agents. **International journal of molecular sciences**, 28 set. 2022. v. 23, n. 19. Disponível em: <<http://dx.doi.org/10.3390/ijms231911433>>.

AGRAWAL, P. *et al.* Approach for Prediction of Antifungal Peptides. **Frontiers in microbiology**, 26 fev. 2018. v. 9, p. 323.

BAHAR, A. A.; REN, D. Antimicrobial peptides. **Pharmaceuticals**, 28 nov. 2013. v. 6, n. 12, p. 1543–1575.

BENKO-ISEPPON, A. M. *et al.* Overview on plant antimicrobial peptides. **Current protein & peptide science**, maio. 2010. v. 11, n. 3, p. 181–188.

BENSON, D. A. *et al.* GenBank. **Nucleic acids research**, jan. 2013. v. 41, n. Database issue, p. D36–42.

BOUTET, E. *et al.* UniProtKB/Swiss-Prot. **Methods in molecular biology**, 2007. v. 406, p. 89–112.

CHEN, C. *et al.* TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. **Molecular plant**, 3 ago. 2020. v. 13, n. 8, p. 1194–1202.

CUNNINGHAM, F. *et al.* Ensembl 2022. **Nucleic acids research**, 7 jan. 2022. v. 50, n. D1, p. D988–D995.

DUBOS, R. J.; HOTCHKISS, R. D. THE PRODUCTION OF BACTERICIDAL SUBSTANCES BY AEROBIC SPORULATING BACILLI. **The Journal of experimental medicine**, 30 abr. 1941. v. 73, n. 5, p. 629–640.

FERNÁNDEZ DE ULLIVARRI, M. *et al.* Antifungal Peptides as Therapeutic Agents. **Frontiers in cellular and infection microbiology**, 17 mar. 2020. v. 10, p. 105.

FRANKISH, A. *et al.* GENCODE 2021. **Nucleic acids research**, 8 jan. 2021. v. 49, n. D1, p. D916–D923.

GOODSTEIN, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. **Nucleic acids research**, jan. 2012. v. 40, n. Database issue, p. D1178–86.

GRAY, J. **Notes on Data Base Operating Systems**. [S.l.]: [s.n.], 1978.

HANCOCK, R. E.; SCOTT, M. G. The role of antimicrobial peptides in animal defenses. **Proceedings of the National Academy of Sciences of the United States of America**, 1 ago. 2000. v. 97, n. 16, p. 8856–8861.

HOCH, J. C. *et al.* Biological Magnetic Resonance Data Bank. **Nucleic acids**

research, 6 jan. 2023. v. 51, n. D1, p. D368–D376.

HUAN, Y. *et al.* Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. **Frontiers in microbiology**, 16 out. 2020. v. 11, p. 582779.

KALVARI, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. **Nucleic acids research**, 8 jan. 2021. v. 49, n. D1, p. D192–D200.

KANG, X. *et al.* DRAMP 2.0, an updated data repository of antimicrobial peptides. **Scientific data**, 13 ago. 2019. v. 6, n. 1, p. 148.

KARUPPASAMY, M. P.; VENKATESWARAN, S.; SUBBIAH, P. PDB-2-PBv3.0: An updated protein block database. **Journal of bioinformatics and computational biology**, 11 maio. 2020. Disponível em: <<https://www.worldscientific.com/worldscinet/jbcb>>. Acesso em: 13 set. 2024.

KOZOMARA, A.; BIRGAOANU, M.; GRIFFITHS-JONES, S. miRBase: from microRNA sequences to function. **Nucleic acids research**, 8 jan. 2019. v. 47, n. D1, p. D155–D162.

KRAGOL, G. *et al.* The antibacterial peptide pyrrocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. **Biochemistry**, 13 mar. 2001. v. 40, n. 10, p. 3016–3026.

LANDRUM, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. **Nucleic acids research**, 4 jan. 2018. v. 46, n. D1, p. D1062–D1067.

LEE, B. T. *et al.* The UCSC Genome Browser database: 2022 update. **Nucleic acids research**, 7 jan. 2022. v. 50, n. D1, p. D1115–D1122.

LI, J. *et al.* Plant antimicrobial peptides: structures, functions, and applications. **Botanical studies**, 29 abr. 2021. v. 62, n. 1, p. 5.

LIU, Y. *et al.* DRAVP: A Comprehensive Database of Antiviral Peptides and Proteins. **Viruses**, 23 mar. 2023. v. 15, n. 4, p. 820. . Acesso em: 20 set. 2024.

LI, Z. *et al.* LncBook 2.0: integrating human long non-coding RNAs with multi-omics annotations. **Nucleic acids research**, 6 jan. 2023. v. 51, n. D1, p. D186–D191.

MACDONALD, J. R. *et al.* The Database of Genomic Variants: a curated collection of structural variation in the human genome. **Nucleic acids research**, jan. 2014. v. 42, n. Database issue, p. D986–92.

MA, L. *et al.* Database Commons: A Catalog of Worldwide Biological Databases. **Genomics, proteomics & bioinformatics**, 23 dez. 2022. v. 21, n. 5, p. 1054–1058. . Acesso em: 20 set. 2024.

_____; BAJIC, V. B.; ZHANG, Z. On the classification of long non-coding RNAs. **RNA biology**, jun. 2013. v. 10, n. 6, p. 925–933.

MARKEL, S.; LEÓN, D. **Sequence Analysis In A Nutshell**. [S.l.]: Turtleback Books, 2003.

MEHTA, D. *et al.* ParaPep: a web resource for experimentally validated antiparasitic peptide sequences and their structures. **Database: the journal of biological databases and curation**, 12 jun. 2014. v. 2014. Disponível em: <<http://dx.doi.org/10.1093/database/bau051>>.

MISTRY, J. *et al.* Pfam: The protein families database in 2021. **Nucleic acids research**, 30 out. 2020. v. 49, n. D1, p. D412–D419. . Acesso em: 13 set. 2024.

O'LEARY, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. **Nucleic acids research**, 4 jan. 2016. v. 44, n. D1, p. D733–45.

OTVOS, L., Jr *et al.* Interaction between heat shock proteins and antimicrobial peptides. **Biochemistry**, 21 nov. 2000. v. 39, n. 46, p. 14150–14159.

PANG, K. C. *et al.* RNAdb 2.0--an expanded database of mammalian non-coding RNAs. **Nucleic acids research**, jan. 2007. v. 35, n. Database issue, p. D178–82.

PARK, Y. *et al.* Antinematodal effect of antimicrobial peptide, PMAP-23, isolated from porcine myeloid against *Caenorhabditis elegans*. **Journal of peptide science: an official publication of the European Peptide Society**, maio. 2004. v. 10, n. 5, p. 304–311.

PAYSAN-LAFOSSE, T. *et al.* InterPro in 2022. **Nucleic acids research**, 6 jan. 2023. v. 51, n. D1, p. D418–D427.

PUSHPANATHAN, M. *et al.* Identification of a novel antifungal peptide with chitin-binding property from marine metagenome. **Protein and peptide letters**, dez. 2012. v. 19, n. 12, p. 1289–1296.

QURESHI, A. *et al.* AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. **Nucleic acids research**, 26 nov. 2013. v. 42, n. D1, p. D1147–D1153. . Acesso em: 20 set. 2024.

SANTOS-SILVA, C. A. DOS *et al.* Plant Antimicrobial Peptides: State of the Art, In Silico Prediction and Perspectives in the Omics Era. **Bioinformatics and biology insights**, 2 set. 2020. Disponível em: <<https://journals.sagepub.com/doi/10.1177/1177932220952739>>. Acesso em: 16 set. 2024.

SCHOCH, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. **Database: the journal of biological databases and curation**, 1 jan. 2020. v. 2020. Disponível em: <<http://dx.doi.org/10.1093/database/baaa062>>.

SEEBACH, S. *et al.* Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. **Nucleic acids research**, jan. 2007. v. 35, n. Database issue, p. D265–8.

SHAI, Y. Mode of action of membrane active antimicrobial peptides. **Biopolymers**,

2002. v. 66, n. 4, p. 236–248.

SHERRY, S. T. *et al.* dbSNP: the NCBI database of genetic variation. **Nucleic acids research**, 1 jan. 2001. v. 29, n. 1, p. 308–311.

SHI, G. *et al.* DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. **Nucleic acids research**, 7 jan. 2022. v. 50, n. D1, p. D488–D496.

SHU, Y.; MCCAULEY, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. **Euro surveillance: bulletin European sur les maladies transmissibles = European communicable disease bulletin**, 30 mar. 2017. v. 22, n. 13. Disponível em: <<http://dx.doi.org/10.2807/1560-7917.ES.2017.22.13.30494>>.

SIGRIST, C. J. A. *et al.* PROSITE, a protein domain database for functional characterization and annotation. **Nucleic acids research**, jan. 2010. v. 38, n. Database issue, p. D161–6.

SILLITOE, I. *et al.* CATH: increased structural coverage of functional space. **Nucleic acids research**, 8 jan. 2021. v. 49, n. D1, p. D266–D273.

STEIN, L. D. Integrating biological databases. **Nature reviews. Genetics**, maio. 2003. v. 4, n. 5, p. 337–345. . Acesso em: 13 set. 2024.

THAKUR, M. *et al.* EMBL's European Bioinformatics Institute (EMBL-EBI) in 2023. **Nucleic acids research**, 5 jan. 2024. v. 52, n. D1, p. D10–D17.

The architecture of an active database management system. **ACM SIGMOD Record**, 1 jun. 1989. Disponível em: <<https://dl.acm.org/doi/10.1145/66926.66946>>. Acesso em: 13 set. 2024.

THE RNACENTRAL CONSORTIUM. RNAcentral: a hub of information for non-coding RNA sequences. **Nucleic acids research**, 8 jan. 2019. v. 47, n. D1, p. D1250–D1251.

UNIPROT CONSORTIUM. UniProt: the Universal Protein Knowledgebase in 2023. **Nucleic acids research**, 6 jan. 2023. v. 51, n. D1, p. D523–D531.

VAN EPPS, H. L. René Dubos: unearthing antibiotics. **The Journal of experimental medicine**, 20 fev. 2006. v. 203, n. 2, p. 259.

WHITMORE, L. *et al.* The peptaibol database: a sequence and structure resource. **Journal of peptide science: an official publication of the European Peptide Society**, Nov-Dec. 2003. v. 9, n. 11-12, p. 663–665.

YADAMP: yet another database of antimicrobial peptides. **International journal of antimicrobial agents**, 1 abr. 2012. v. 39, n. 4, p. 346–351. . Acesso em: 20 set. 2024.

YASIN, B. *et al.* Theta defensins protect cells from infection by herpes simplex virus by inhibiting viral adhesion and entry. **Journal of virology**, maio. 2004. v. 78, n. 10, p. 5147–5156.

YOKOYAMA, S. *et al.* The chitin-binding capability of Cy-AMP1 from cycad is essential to antifungal activity. **Journal of peptide science: an official publication of the European Peptide Society**, jul. 2009. v. 15, n. 7, p. 492–497.

YU, Y. *et al.* Plant Public RNA-seq Database: a comprehensive online database for expression analysis of ~45 000 plant public RNA-Seq libraries. **Plant biotechnology journal**, maio. 2022. v. 20, n. 5, p. 806–808.

ZHANG, L.; ROZEK, A.; HANCOCK, R. E. Interaction of cationic antimicrobial peptides with model membranes. **The Journal of biological chemistry**, 21 set. 2001. v. 276, n. 38, p. 35714–35722.

ZHAO, Y. *et al.* NONCODE 2016: an informative and valuable data source of long non-coding RNAs. **Nucleic acids research**, 4 jan. 2016. v. 44, n. D1, p. D203–8.

ZOU, D. *et al.* Biological Databases for Human Research. **Genomics, proteomics & bioinformatics**, 21 fev. 2015. v. 13, n. 1, p. 55–63. . Acesso em: 13 set. 2024.

ZUPIN, L. *et al.* Bioactive Antimicrobial Peptides: A New Weapon to Counteract Zoonosis. **Microorganisms**, 7 ago. 2022. v. 10, n. 8. Disponível em: <<http://dx.doi.org/10.3390/microorganisms10081591>>.