

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE RONDÔNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM PROPRIEDADE INTELECTUAL
E TRANSFERÊNCIA DE TECNOLOGIA PARA INOVAÇÃO

CAROLINA BARROS DA COSTA

INOVAÇÃO NA BIOTECNOLOGIA: DESENVOLVIMENTO DE UM REPOSITÓRIO
ON-LINE (*SOFTWARE*) PARA ESTRUTURAS MOLECULARES DE COMPOSTOS
NATURAIS DA FLORA AMAZÔNICA

Porto Velho/RO

2024

CAROLINA BARROS DA COSTA

**INOVAÇÃO NA BIOTECNOLOGIA: DESENVOLVIMENTO DE UM REPOSITÓRIO
ON-LINE (*SOFTWARE*) PARA ESTRUTURAS MOLECULARES DE COMPOSTOS
NATURAIS DA FLORA AMAZÔNICA**

Trabalho de Conclusão de Curso apresentada como requisito parcial para obtenção do título de Mestre em Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação – PROFNIT – Ponto Focal Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO) - Campus Porto Velho Zona Norte

Orientador (a): Marcio Rodrigues Miranda

Porto Velho/RO

2024

Ficha catalográfica elaborada pelo Sistema Gerador de Ficha Catalográfica do IFRO.

Costa, Carolina Barros da.

Inovação na biotecnologia: desenvolvimento de um repositório on-line (software) para estruturas moleculares de compostos naturais da flora amazônica / Carolina Barros da Costa. - Porto Velho, 2024. 63 f. : il.

Orientador(a): Prof. Marcio Rodrigues Miranda.

Dissertação (Mestrado Profissional em Propriedade Intelectual e Transferência de Tecnologia para Inovação - ProfNIT) – Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO, Porto Velho, 2024.

1. Biodiversidade amazônica. 2. Base de dados. 3. Inovação biotecnológica. 4. Ciência de dados. I. Miranda, Marcio Rodrigues (orient.). II. Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO. III. Título.

Bibliotecário(a) Responsável: Roseni Santos Rodrigues, CRB-11/916

CAROLINA BARROS DA COSTA

**INOVAÇÃO NA BIOTECNOLOGIA: DESENVOLVIMENTO DE UM REPOSITÓRIO
ON-LINE (SOFTWARE) PARA ESTRUTURAS MOLECULARES DE COMPOSTOS
NATURAIS DA FLORA AMAZÔNICA**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação - PROFNIT- Ponto Focal Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO) - Campus Porto Velho Zona Norte

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Marcio Rodrigues Miranda
Presidente da Banca Examinadora, PROFNIT IFRO

Prof. Dr. Márcio Teixeira Oliveira
Membro Externo PROFNIT IFMS

Prof. Dr. Wanderson Roger Azevedo Dias
Membro Externo

DEDICATÓRIA

Dedico este trabalho à minha mãe, Ana Lídia, e ao meu pai, Carlos Clênis, pelo carinho, incentivo e apoio fornecido para a realização deste mestrado.

AGRADECIMENTOS

Aos meus pais, Ana Lúdia e Carlos Clênis, por todo o incentivo e apoio fornecido para a realização deste mestrado.

Ao meu irmão, Felipe Costa, pelo incentivo e apoio no decorrer desta caminhada.

Aos meus amigos, Daniele Anjos e Fernanda Alves, pelo incentivo e apoio durante esse processo.

Ao meu orientador, Dr. Márcio Rodrigues Miranda, pela orientação e contribuição no desenvolvimento deste projeto.

Ao grupo de pesquisa GoTEC - Grupo de pesquisa em soluções tecnológicas pelas contribuições na minha formação como pesquisadora.

Aos meus colegas do mestrado por toda a amizade, troca de experiência e parceria no decorrer do processo.

Ao IFRO - Instituto Federal de Rondônia, pela oportunidade de cursar esse mestrado.

À FORTEC - Associação Fórum Nacional de GEstores de Inovação e Transferência de Tecnologia, proponente do PROFNIT à CAPES.

COSTA, Carolina Barros da. **Inovação na Biotecnologia:** Desenvolvimento de um repositório on-line (*software*) para estruturas moleculares de compostos naturais da flora amazônica. 2024. f. Trabalho de Conclusão de Curso (Mestrado em Propriedade Intelectual e Transferência de Tecnologia para Inovação) – Instituto Federal de Rondônia, Porto Velho, 2024.

RESUMO

As diversas espécies de plantas existentes na Região Amazônica possuem um grande potencial biotecnológico. Nesse contexto, a disponibilidade de um repositório on-line auxiliaria na compreensão desse potencial biotecnológico, pois iria catalogar e disponibilizar as informações de forma que os interessados possam explorar essas riquezas moleculares. No entanto, há uma escassez de um repositório que seja voltado para as plantas amazônicas. Diante disso, o objetivo desse projeto foi desenvolver um repositório on-line (*software*) das estruturas moleculares da flora amazônica. Para o desenvolvimento do repositório, foi utilizado a metodologia por prototipação. A metodologia de desenvolvimento por prototipação envolve a criação iterativa de um protótipo para coletar requisitos, com colaboração entre cliente e desenvolvedores. O protótipo é usado para refinar funcionalidades antes de desenvolver o *software* final, garantindo qualidade e manutenção ao descartar parte do protótipo. Os resultados obtidos reforçaram a escassez de bancos de dados que forneçam informações sobre a flora amazônica, e revelam necessidade urgente de plataformas integradas que centralizem dados biológicos, químicos e taxonômicos, permitindo uma consulta mais eficiente e o avanço da pesquisa científica. A falta de repositórios acessíveis e atualizados limita a prospecção de compostos bioativos, por conseguinte, desperdiçando o vasto potencial biotecnológico e econômico da biodiversidade amazônica. O repositório on-line de moléculas da flora amazônica foi desenvolvido até a quinta etapa do ciclo proposto por Pressman e Maxim (2016), o qual consiste no refinamento do protótipo. O banco de dados do repositório foi construído utilizando técnicas de *web scrapping* e ETL que foram capazes de integrar dados das espécies amazônicas e seus dados biológicos em um único banco de dados. O repositório desenvolvido está ligado diretamente a transferência de tecnologia, pois fornecerá capacidade de armazenamento e gerenciamento eficiente de informações relevantes, como estruturas moleculares e sequências genômicas, proporcionando a facilitação do acesso a informações biológicas relevantes, e permitindo uma compreensão abrangente e aprofundada dos fenômenos biológicos envolvendo a flora amazônica.

Palavras-Chave: biodiversidade amazônica; base de dados; inovação biotecnológica; Ciência de Dados.

COSTA, Carolina Barros da. **Innovation in Biotechnology**: Development of an online repository (software) for molecular structures of natural compounds from the Amazon flora. 2024. f. Trabalho de Conclusão de Curso (Mestrado em Propriedade Intelectual e Transferência de Tecnologia para Inovação) – Instituto Federal de Rondônia, Porto Velho, 2024.

ABSTRACT

The various plant species found in the Amazon region have great biotechnological potential. In this context, the availability of an online repository would help to understand this biotechnological potential, as it would catalog and make information available so that interested parties can explore these molecular riches. However, there is a shortage of a repository focused on Amazonian plants. In view of this, the objective of this project was to develop an online repository (software) of the molecular structures of the Amazonian flora. The prototyping methodology was used to develop the repository. The prototyping development methodology involves the iterative creation of a prototype to collect requirements, with collaboration between the client and developers. The prototype is used to refine functionalities before developing the final software, ensuring quality and maintainability by discarding part of the prototype. The results obtained reinforced the shortage of databases that provide information on the Amazonian flora, and reveal an urgent need for integrated platforms that centralize biological, chemical, and taxonomic data, allowing for more efficient consultation and the advancement of scientific research. The lack of accessible and updated repositories limits the prospecting of bioactive compounds, thus wasting the vast biotechnological and economic potential of Amazonian biodiversity. The online repository of molecules from the Amazonian flora was developed up to the fifth stage of the cycle proposed by Pressman and Maxim (2016), which consists of refining the prototype. The repository database was built using web scrapping and ETL techniques that were able to integrate data from Amazonian species and their biological data into a single database. The developed repository is directly linked to technology transfer, as it will provide storage capacity and efficient management of relevant information, such as molecular structures and genomic sequences, facilitating access to relevant biological information and allowing a comprehensive and in-depth understanding of biological phenomena involving the Amazonian flora.

Keywords: Amazonian biodiversity; database; biotechnological innovation; Data Science.

LISTA DE FIGURAS

FIGURA 1	Representação da Amazônia Internacional e a Amazônia Brasileira (Amazônia Legal)	15
FIGURA 2	Processo de desenvolvimento de fármacos.....	
FIGURA 3	Modelo de processo de prototipação.....	26
FIGURA 4	Identificação dos Bancos de Dados Biológicos.....	32
FIGURA 5	Bancos de Dados utilizados.....	37
FIGURA 6	Busca por filtros.....	38
FIGURA 7	Cards intuitivos.....	40
FIGURA 8	Informações sobre as espécies	41
FIGURA 9	Informações sobre as moléculas	42
FIGURA 10	Comparação entre os dados disponibilizados pelos bancos de dados biológicos levantados e o Moléculas da Amazônia.....	43

LISTA DE TABELAS

TABELA 1	Matriz de amarração do desenvolvimento do repositório.....	30
TABELA 2	Informações Gerais dos Bancos de Dados Biológicos.....	34
TABELA 3	Avaliação da Experiência de Usuário dos Bancos de Dados.....	35
TABELA 4	Tipos de dados para realização das buscas nos campos de busca do repositório.....	39

LISTA DE SIGLAS E ABREVIATURAS

API	Application Programming Interface
CEBio	Centro de Estudos de Biomoléculas Aplicadas à Saúde
CRISP-DM	Cross-Industry Standard Process for Data Mining
DCBD	Descoberta de Conhecimento em Bases de Dados
ETL	Extract, Transform and Load
FORTEC	Associação Fórum Nacional de Gestores de Inovação e Transferência de Tecnologia
GBIF	Global Biodiversity Information Facility
HTTP	Hypertext Transfer Protocol
IFRO	Instituto Federal de Educação, Ciência e Tecnologia de Rondônia
IUPAC	International Union of Pure and Applied Chemistry
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Database
KEGG	Kyoto Encyclopedia of Genes and Genomes
MGI	Genoma Informatics Mouse
MVC	Model View Controller
NCBI	Nacional Center for Biotechnology Information
NuBBEDB	Núcleos de Bioensaios, Ecofisiologia e Biossíntese de Produtos Naturais
OMIM	Online Mendelian Inheritance in Man
PDB	Protein Data Bank
PROFNIT	Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação
UFG	Universidade Federal de Goiás
UML	Unified Modeling Language
UNESP	Universidade Estadual Paulista
TAIR	The Arabidopsis Information Resource

SUMÁRIO

1 APRESENTAÇÃO	13
2 INTRODUÇÃO	15
3 JUSTIFICATIVA	20
3. 1 Lacuna preenchida pelo TCC.....	20
3. 2 Aderência ao PROFNIT	20
3. 3 Impacto.....	21
3. 4 Aplicabilidade	21
3. 5 Inovação	21
3.6 Complexidade.....	21
4 OBJETIVO.....	22
4.1 Objetivo Geral.....	22
4. 2 Objetivos Específicos	22
4. 2. 1 Mapear os bancos de dados biológicos de origem brasileira relacionados à flora	22
4. 2. 2 Descrever as informações obtidas a partir desse mapeamento	22
4. 2. 3 Desenvolvimento do repositório on-line para a análise e exibição dos dados associados	22
5 REFERENCIAL TEÓRICO (ESTADO DA ARTE E DA TÉCNICA)	23
6 METODOLOGIA.....	27
6.1 Mapeamento dos bancos de dados biológicos de origem brasileira relacionadas à flora.....	27
6.2 Descrição das informações obtidas a partir do mapeamento	27
6.3 Desenvolvimento do Repositório.....	29
7 RESULTADOS PARCIAIS	33
7.1 Mapeamento dos bancos de dados biológicos voltados para flora	33
7.2 Descrição das informações obtidas a partir do mapeamento	33
7.3 Desenvolvimento do repositório	37
8 DISCUSSÃO	44
9 ENTREGÁVEIS DE ACORDO COM OS PRODUTOS DO TCC	49
10 CRONOGRAMA.....	Erro! Indicador não definido.
REFERÊNCIAS	51
APÊNDICE A - Matriz FOFA (SWOT).....	56
APÊNDICE B - Modelo de Negócio CANVAS	57
APÊNDICE C - Artigo publicado	58
APÊNDICE D - Artigo publicado	59
ANEXO A - Comprovante de publicação/submissão de artigo	60

1 APRESENTAÇÃO

Os motivos pelos quais este estudo foi desenvolvido originaram-se através das dificuldades vivenciadas por esta pesquisadora durante sua iniciação científica na graduação, período esse em que teve a oportunidade de conhecer o Centro de Estudos de Biomoléculas Aplicadas à Saúde (CEBio), localizado na Fiocruz-Rondônia. A vivência nesse centro de pesquisa tornou evidente a ansiedade por parte dos pesquisadores com a ausência de ferramentas que permitam conhecer e explorar o potencial da biodiversidade amazônica.

A entrada no Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação (PROFNIT), permitiu que esta pesquisadora conhecesse e fizesse parte do Grupo de Pesquisa em Soluções Tecnológicas - GOTEC, localizado no Instituto Federal de Rondônia (IFRO), Campus Calama. Tal ocorrido fez com que tivesse a oportunidade de contribuir para o preenchimento dessa lacuna, a partir do desenvolvimento de um repositório on-line que disponibiliza dados a respeito da flora amazônica.

Devido à sua vasta biodiversidade, a Amazônia é uma fonte de matéria-prima que pode ser explorada de forma sustentável. De acordo com Barbosa (2020), essa variedade de matéria-prima pode ser transformada em uma gama de moléculas e compostos de alto valor agregado, com potencial de serem exploradas por diversos setores, principalmente o biotecnológico. No entanto, há uma dificuldade em coletar e sintetizar essas moléculas.

Logo, o objetivo deste estudo é desenvolver um repositório on-line de estruturas moleculares dos compostos naturais provenientes da flora amazônica, denominado “Moléculas da Amazônia”. Para tal, o mapeamento de bancos de dados biológicos brasileiros relacionados à flora e a descrição das informações obtidas auxiliarão no desenvolvimento do repositório, bem como na exibição dos dados associados. De acordo com (Baxevanis e Bateman, 2011), o desenvolvimento de bancos de dados biológicos impacta positivamente na pesquisa científica, uma vez que fornecem um ambiente propício para a coleta, integração e compartilhamento de dados biológicos. Desse modo, o repositório Moléculas da Amazônia poderá contribuir na prospecção e obtenção de informações a respeito do potencial biotecnológico da flora amazônica.

Para desenvolver o repositório Moléculas da Amazônia, foi utilizada a metodologia de desenvolvimento por prototipação, conforme Pressman e Maxim (2016). Tal abordagem permite identificar e definir requisitos de forma iterativa, ideal

para sistemas que possuem consultas dinâmicas, forte interação com usuários ou processos complexos.

Para auxiliar na construção do repositório, foi utilizado o *Database Commons* para identificar os bancos de dados biológicos de origem brasileira relacionados à flora. Para tal, apenas as bases de dados brasileiras relevantes e acessíveis foram mantidas e agrupadas conforme a natureza dos dados. Em seguida, foram coletadas e analisadas as informações sobre cada banco de dados, como objetivo, tipos de moléculas, idioma e instituição responsável. Além disso, também foi avaliado a qualidade dos serviços destes, seguindo a metodologia proposta por Yang et al. (2004), os quais incluem a usabilidade, utilidade, adequação, acessibilidade e interação dos sistemas a serem avaliados.

A criação desse repositório pode contribuir para a sociedade de diversas maneiras. Dentre elas, a centralização de dados biológicos relacionados à flora brasileira, facilitando o acesso a informações essenciais para pesquisadores, cientistas e instituições. Ademais, o fornecimento de uma plataforma robusta para coleta, análise e comparação de dados sobre moléculas e compostos de plantas nativas da Amazônia brasileira pode auxiliar no desenvolvimento de pesquisas em biotecnologia. Por fim, o fornecimento de dados para a descoberta de novos produtos biotecnológicos pode não só impulsionar iniciativas de desenvolvimento sustentável, como também favorecer a colaboração entre diversas partes interessadas.

2 INTRODUÇÃO

A biodiversidade consiste em uma variedade de formas de vida, a qual está relacionada à sua diversidade de espécies, diversidade genética e diversidade de ecossistemas que juntos desempenham um papel importante não só no ciclo global de nutrientes, mas também na manutenção da biodiversidade em larga escala, abrangendo particularidades como clima, solo, fauna e flora (Wilson e Peter, 1988).

Posto isto, a Conferência a respeito da Biodiversidade realizada no ano de 1988 em Washington propôs o termo Países Megadiversos, utilizado para descrever países no mundo cuja biodiversidade se destaca por sua abundância. Esta nomeou 17 países, dentre os 168 que possuem esta característica, relatando que estes concentram cerca de 70% da biodiversidade do mundo, e, junto deles, o Brasil figura entre os primeiros da lista (Marques *et al.* 2014).

O Brasil possui uma das maiores biodiversidades do mundo, cuja fauna e flora correspondem a cerca de 20% do número total de espécies do planeta (Brasil, 2024). De acordo com o programa Flora e Funga do Brasil, até dezembro de 2024 já foram catalogadas 52.855 espécies, dos quais 20.555 são endêmicas (Reflora, 2024). Isso torna o Brasil o lar de uma fonte de recursos biológicos exclusivos, pois trata-se de espécies que só podem ser encontrados em território brasileiro. Ocupando quase metade do continente sul-americano, essa variedade de espécies e ecossistemas são distribuídos pelos seus seis biomas, sendo eles: amazônica, mata atlântica, pantanal, cerrado, caatinga e pampas (IBGE, 2019).

A Amazônia é reconhecida mundialmente como sendo a maior floresta tropical do mundo, a qual abriga uma biodiversidade abundante composta por uma variedade de espécies de plantas, animais e insetos (Plotkin, 2020). Localizada na América do Sul, seu território abrange o Brasil, o Peru, a Colômbia, a Venezuela, o Equador, a Bolívia, a Guiana, o Suriname e a Guiana Francesa (Aragón, 2018), formando a Amazônia Internacional. Em âmbito nacional, o Brasil detém cerca de 60% da Amazônia e está presente nos estados do Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins e parte do Maranhão (OTCA, 2021; Aragón, 2018), como pode ser visualizado na Figura 1. Essa diversidade biológica coloca o país em uma posição estratégica para o desenvolvimento da bioeconomia.

Figura 1: Representação da Amazônia Internacional e a Amazônia Brasileira (Amazônia Legal).



Fonte: Elaborado pela autora (2024).

A bioeconomia é uma ciência transdisciplinar, cujo intuito é enfrentar os desafios do desenvolvimento sustentável e a necessidade de alternativas para aumentar a produção e preservar o meio ambiente (Mejias, 2019). Por meio da promoção do uso sustentável de recursos biológicos, esta integra princípios relacionados à conservação ambiental e ao desenvolvimento econômico da região (Pimenta e Azevedo, 2020). Ademais, a bioeconomia favorece uma maior conexão entre os setores primários e as indústrias de transformação e de serviços, unindo-os como partes interligadas de um único processo. A partir de recursos biológicos renováveis e resíduos de processos de extração/transformação, é possível criar alimentos, energia, artigos químicos e têxteis e outros itens que podem possuir tanto valor econômico quanto ambiental (Silva *et al.*, 2018).

Nesse contexto, a Amazônia se apresenta como um terreno fértil para o avanço do conhecimento científico e inovação tecnológica. Sua diversidade biológica não só possibilita a exploração sustentável de seus recursos naturais, como também abre portas para avanços significativos em várias áreas, como química, bioquímica, microbiologia e engenharia genética (Mello, 2015). A variedade de matérias primas que podem ser encontradas na região amazônica, quando aliadas a bioinovação, permite a transformação desses recursos naturais em uma gama de moléculas e compostos de alto valor agregado (Barbosa, 2020), os quais são essenciais para o desenvolvimento do campo da biotecnologia. Logo, não só aumentando o valor de seus recursos biológicos, como também os transformando em produtos valiosos. Dessa forma, a bioeconomia exerce um papel essencial não só na bioinovação, como também na capacitação de recursos humanos, sendo estas fundamentais para impulsionar o desenvolvimento de uma ciência aplicada na conversão de produtos da natureza em produtos comerciais (Willerding *et al.*, 2020).

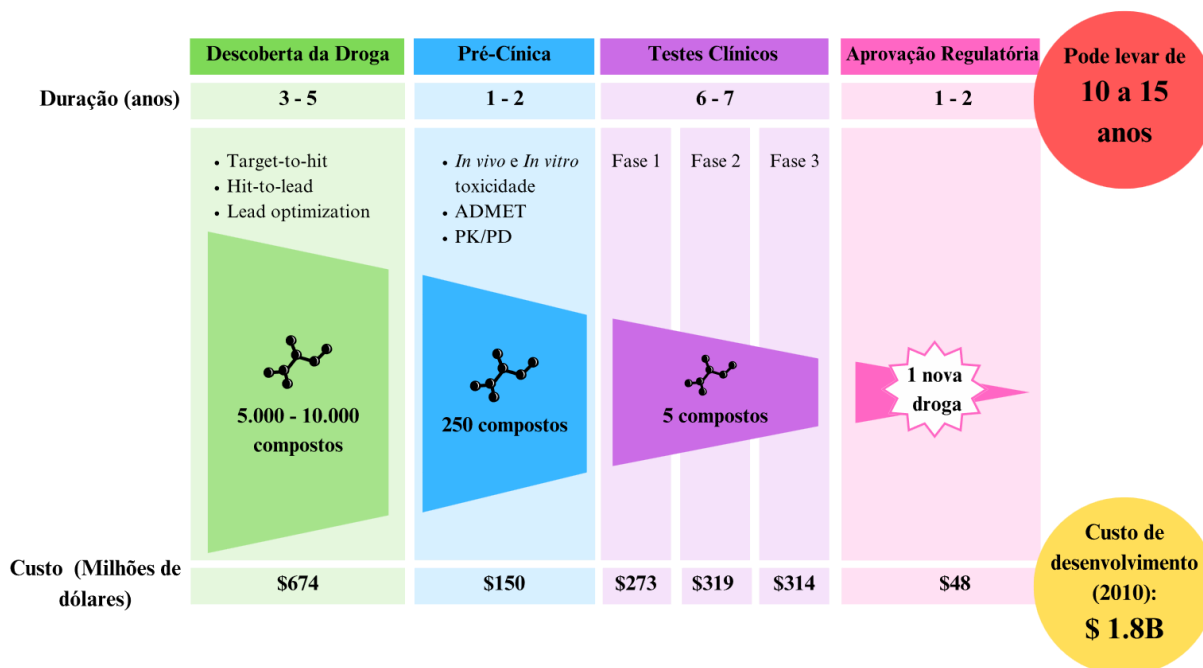
No entanto, apesar do vasto potencial da Amazônia em gerar moléculas e compostos de alto valor agregado, ainda há uma dificuldade em coletar e sintetizar essas moléculas. Dentre os fatores que contribuem para essa realidade, tem-se a complexidade dos processos biológicos envolvidos e a grande variedade de compostos presentes nas espécies amazônicas.

O principal modelo utilizado para o desenvolvimento de fármacos não só é um processo longo e oneroso, como também envolve altos custos (Matthews *et al.*, 2016. Paul *et al.*, 2010). De acordo com Matthews *et al.* (2016), até um fármaco está disponível no mercado, este passa por um processo de desenvolvimento que envolve quatro etapas: descoberta de drogas, pré-clínica, testes clínicos (Fase I, II e III) e aprovação regulatória. Esse processo pode levar de 10 a 15 anos e pode chegar a ter um investimento médio de 1.8 bilhões de dólares (Paul *et al.*, 2010).

Esse tempo e custo elevado está relacionado principalmente a primeira etapa, que envolve testar muitas moléculas, variando de 5 a 10 mil, que pode demorar de 5 a 7 anos para ser concluída (Matthews *et al.*, 2016), e, sozinha, pode ter um custo de 674 milhões de dólares (Paul *et al.*, 2010). Isso ocorre, pois, tais moléculas são coletadas, sintetizadas e testadas para averiguar se estas apresentam alguma atividade biológica. Coletar, sintetizar e testar esse elevado número de moléculas requer não só um gasto de esforço muito grande, como também de tempo e dinheiro, principalmente quando se leva em consideração que apenas 250 moléculas podem

chegar a ir para a próxima etapa e que apenas uma molécula pode chegar a se tornar um fármaco de fato (Matthews *et al.*, 2016), como pode ser visto na Figura 2.

Figura 2: Processo de desenvolvimento de fármacos.



Fonte: Elaborado pela autora a partir da adaptação de Matthews *et al.*, (2016) e Paul *et al.*, (2010).

Isso evidencia a escassez de infraestrutura adequada para o processamento e síntese de estrutura molecular e a necessidade de desenvolver tecnologias avançadas para a extração e análise de moléculas. Esse déficit se torna uma barreira para desenvolvimento do verdadeiro potencial biotecnológico da Região Amazônica. Desta forma, superar esses obstáculos se torna fundamental para desbloquear esse potencial e a bioinformática surge como uma ferramenta promissora para tal.

Nos últimos anos, o campo de biológicas tem testemunhado uma explosão na geração de dados graças a avanços na bioinformática moderna. As primeiras bases de bioinformática foram lançadas no início da década de 60, quando métodos computacionais começaram a ser aplicados na análise de sequências de proteínas. A partir disso, surgiram novos bancos de dados de sequências biológicas, que somado com grandes melhorias nas tecnologias de sequenciamento e a redução de custos, provocaram um aumento exponencial de dados. Esses avanços têm gerado uma quantidade massiva de informações sobre sequências genéticas, estruturas proteicas, redes de interações moleculares e fenótipos biológicos, desafiando os pesquisadores

a armazenar, organizar e analisar esses dados de forma eficiente e significativa (Gauthier *et al.*, 2019).

Nesse contexto, as bases de dados biológicas desempenham um papel crucial na pesquisa científica, fornecendo uma infraestrutura essencial para a coleta, integração e compartilhamento de informações biológicas.

Com isso, o desenvolvimento de um banco de dados biológicos voltado para a flora Amazônica pode não só contribuir para o entendimento mais aprofundado de sua biodiversidade, como também permitir que pesquisadores, empresas e governos identifiquem e explorem de forma sustentável os seus recursos biológicos. Essa realidade pode levar ao desenvolvimento de novos produtos, processos e tecnologias baseadas nas riquezas biológicas dessa região, levando a um impacto significativo no avanço da bioeconomia.

Diante disso, este trabalho tem como intuito desenvolver um repositório on-line (software) voltado para as estruturas moleculares provenientes da flora amazônica, o qual contribuirá para a compreensão do seu potencial científico e tecnológico. A partir da interação com bancos de dados que contêm informações a respeito da flora amazônica, esse repositório será capaz de disponibilizar essas informações de forma que pesquisadores, cientistas ou indústrias possam ter acesso a esses dados e possam aplicá-los em suas pesquisas ou desenvolvimento de produtos. Esse produto preencherá uma lacuna significativa quando se trata de bancos de dados voltados para a flora amazônica. A integração de dados sobre a vasta diversidade de sua flora, a padronização e anotação desses dados, e a garantia da acessibilidade e usabilidade dos sistemas são desafios importantes a serem enfrentados para melhor compreender e preservar essa riqueza biológica única.

3 JUSTIFICATIVA

Não há um banco de dados que centralize informações da flora amazônica que contemple as espécies, as biomoléculas e as estruturas digitais, independente dos formatos de arquivos, nível estrutural ou formato de visualização, estando estas dispersas em diferentes bancos de dados.

Essa fragmentação limita o potencial de prospecção de compostos, pois o pesquisador precisa consultar múltiplas fontes para a obtenção de informações, fazendo com que este não só perca muito tempo nesta atividade, como também tenha um gasto financeiro e de recursos naturais maior realizando a bioprospecção de maneira experimental. Com isso, a criação de um banco de dados centralizado permite que o pesquisador tenha acesso a dados completos de maneira rápida, contribuindo para a redução do tempo e custos envolvidos no processo de pesquisa.

3. 1 Lacuna preenchida pelo TCC

A lacuna a ser preenchida pelo TCC está relacionada ao preenchimento da ausência de repositórios voltados para a flora amazônica existentes no mercado atualmente, por conseguinte, contribuirá para avanço de pesquisas envolvendo plantas amazônicas.

3. 2 Aderência ao PROFNIT

Este projeto se encontra alinhado aos interesses do programa, uma vez que o desenvolvimento de um repositório on-line está ligado diretamente à transferência de tecnologia. O repositório permitirá o compartilhamento de conhecimento entre pesquisadores e cientistas, através do compartilhamento de espécies e estruturas moleculares da flora amazônica de forma acessível e ampla, contribuindo para o avanço da ciência e o desenvolvimento de novas aplicações técnicas. Dessa forma, poderá estimular as colaborações e as parcerias interinstitucionais, uma vez que sua existência incentivará a colaboração entre instituições de pesquisas, empresas e organizações interessadas na flora amazônica, podendo resultar em projetos acordos de cooperação de pesquisa e transferência de tecnologia entre os envolvidos. O repositório poderá contribuir com a inovação, uma vez que será uma fonte para o desenvolvimento de novas tecnologias, produtos e serviços. Por fim, poderá fornecer transferência de tecnologia para comunidades locais, uma vez que pode promover

essa transferência para comunidades locais da região amazônica, como comunidades indígenas ou agricultores locais, fornecendo conhecimento sobre as propriedades medicinais ou outras aplicações das plantas nativas, contribuindo para o desenvolvimento sustentável da região.

3. 3 Impacto

Essa tecnologia poderá ser utilizada por pesquisadores tanto de instituições públicas quanto privadas, que possuam interesse em utilizar moléculas da flora amazônica em suas pesquisas. Ademais, poderá ser utilizada também por pessoas que tenham interesse na área, como empresas, organizações e instituições de pesquisa.

O repositório facilitará o acesso a bases biológicas da flora amazônica, por conseguinte, irá aumentar os estudos nessa área, gerando um alto impacto na área científica e biotecnológica.

3. 4 Aplicabilidade

Essa produção tecnológica é de fácil aplicabilidade, podendo abranger todo o Brasil, ou falantes da língua portuguesa, podendo ter um alto potencial na área industrial, na pesquisa científica, na saúde, ambiental, e agropecuária, sendo de fácil replicabilidade como produção técnica.

3. 5 Inovação

Este projeto possui uma produção com alto teor inovativo, sendo uma combinação de conhecimentos pré-estabelecidos que serão utilizados na construção de um repositório on-line que será responsável por tornar a informação mais acessível, em que o usuário poderá utilizá-lo de forma intuitiva.

3.6 Complexidade

Este projeto possui alta complexidade, uma vez que é o resultado da combinação de conhecimento pré-estabelecidos, que podem ser utilizados por pesquisadores, cientistas e indústrias que possuam interesse na área. Não requer, necessariamente, a participação de outros autores, sendo de fácil execução e adaptação.

4 OBJETIVO

4.1 Objetivo Geral

Desenvolver um repositório on-line das estruturas moleculares dos compostos naturais provenientes da flora amazônica.

4. 2 Objetivos Específicos

4. 2. 1 Mapear os bancos de dados biológicos de origem brasileira relacionados à flora

4. 2. 2 Descrever as informações obtidas a partir desse mapeamento

4. 2. 3 Desenvolvimento do repositório on-line para a análise e exibição dos dados associados

5 REFERENCIAL TEÓRICO

A bioinformática é um campo científico interdisciplinar que desenvolve métodos para armazenar, recuperar, organizar e analisar dados biológicos (Ogbe, 2016) e pode ser dividida em dois grupos: i) associado à sequência; ii) associado à estrutura biomolecular. A integração desses dados permite o uso de redes, tornando possível identificar grupos gênicos relacionados a uma determinada rota metabólica e associá-los com um determinado fenótipo resultante desta via, contribuindo para a melhor compreensão de uma determinada característica, como o caso da análise de suposição proporcionada pelo *software Weighted Gene Co-Expression Network Analysis* - WGCNA (Diniz e Canduri, 2017).

Os experimentos *in silico* podem ser usados em estágios iniciais do processo de desenvolvimento de medicamentos para compostos planejados a serem sintetizados, para os quais nenhum ou apenas um pequeno composto está disponível, ou também para impurezas ou produtos de rápida degradação posteriormente no processo de desenvolvimento de medicamentos (Bah *et al.*, 2018). Todavia, é essencial o desenvolvimento de um design experimental bem aplicado com as devidas validações para a melhor acurácia do teste *in silico* para que seja introduzido no processo de desenvolvimento de medicamentos (Amberg, 2013).

Os bancos de dados biológicos têm desempenhado um papel central na bioinformática. Eles nos oferecem a oportunidade de acessar uma ampla variedade de dados biologicamente relevantes, incluindo as sequências genômicas de uma gama cada vez maior de organismos. Os principais bancos de dados e portais de sequência, como o *GenBank*, o *UCSC Genome Browser* e o *Ensembl*. Banco de dados de organismos modelo, incluindo *WormBase*, *The Arabidopsis Information Resource* - TAIR, e aqueles disponibilizados através do recurso *Genoma Informatics Mouse* - MGI, Bancos de dados centrados em não sequências, como o *Online Mendelian Inheritance in Man* - OMIM, o *Protein Data Bank* - PDB, o *MetaCyc* e o *Kyoto Encyclopedia of Genes and Genomes* - KEGG (Baxevanis e Bateman, 2011).

A crescente disponibilidade de proteínas, estruturas de ligantes com afinidades de ligação medidas e conjuntos dados contendo compostos ativos, estão aumentando o uso de técnicas sofisticadas de aprendizado de máquina para obter melhores funções de pontuação (Guedes *et al.*, 2018).

Com o aumento de informações e a crescente disponibilidade de proteínas, estruturas de ligantes com afinidades de ligação medidas e conjuntos dados contendo compostos ativos, o uso de técnicas computacionais tem sido de grande relevância (Guedes *et al.*, 2018).

Experimentos *in silico* já são conhecidos por seu uso extensivo em experimentações farmacocinéticas, além disso, complexos modelos computacionais têm sido aplicados a problemas fisiopatológicos para fornecer informações que não podem ser obtidas praticamente ou eticamente pelos métodos tradicionais de pesquisa clínica (Colquitt *et al.*, 2011; Kaushik *et al.*, 2020).

Uma grande diversidade de produtos naturais tem demonstrado atividades farmacológicas ou biológicas que podem ser de benefício terapêutico no tratamento de doenças e também são uma importante fonte de inspiração para o desenvolvimento de novos medicamentos em potencial. As últimas décadas testemunharam extenso estudo de produtos naturais para suas perspectivas promissoras na aplicação da química médica, biologia molecular e ciências farmacêuticas (Sorokina e Steinbeck, 2020; Xie *et al.*, 2015).

O interesse pelos produtos naturais pela comunidade científica ainda tem um constante crescimento, tendo um exponencial aumento de bancos de dados e coleções disponíveis. Dados recentes têm trazidos aproximadamente 120 diferentes bancos de dados de compostos naturais depositados, dentre os mais acessíveis destacam-se, o ZINC e o COCONUT (Sorokina e Steinbeck, 2020).

A fim de integrar e caracterizar compostos naturais brasileiros, alguns bancos de dados foram criados, como o Núcleos de Bioensaios, Ecofisiologia e Biossíntese de Produtos Naturais - NuBBEDB que vem possibilitando o acesso a diversos compostos da biodiversidade do país (Pilon *et al.*, 2017). Entretanto, há poucos bancos de dados sobre compostos naturais oriundos de plantas da região amazônica, no qual pode-se citar a base de dados de Maia & Andrade (2009), que cataloga informações da composição química dos óleos e aromas de mais de 350 espécies, associadas a um maior número de tipos químicos.

Os óleos essenciais de muitas espécies oferecem ótimas condições para exploração econômica e utilização no mercado nacional e internacional de fragrâncias, cosméticos, agrotóxicos e domésticos (Maia e Andrade, 2009). Deve-se incentivar os

bancos de dados construídos a partir de recursos minoritários, acreditando que esses esforços serão finalmente recompensados e integrados aos bancos de dados abrangentes para um uso mais profundo. Desta forma, para que seja possível desenvolver melhores bancos de dados de produtos naturais e, assim, apoiar melhor a descoberta e o desenvolvimento de drogas, deve-se fazer um esforço para visualizar as drogas de produtos naturais em redes e enfatizar mais as abordagens orientadas a sistemas (Chen *et al.*, 2005; Xie *et al.*, 2015).

Algumas técnicas possibilitam o melhor desenvolvimento e criação de bancos de dados. O processo capaz de gerar conhecimento a partir de dados estruturados nomeia-se de *Knowledge Discovery in Database* (KDD) ou Descoberta de Conhecimento em Bases de Dados (DCBD). Esse processo combina diversas áreas da descoberta do conhecimento, tais como Aprendizagem de Máquina, Reconhecimento de Padrões, Estatística e Inteligência Artificial, com o objetivo de extrair, de forma automática, informação útil em bases de dados.

A técnica de *Cross-Industry Standard Process for Data Mining* (CRISP-DM), foi concebida originalmente para mineração de dados. Para a CRISP-DM, o ciclo de vida do processo de DCBD segue uma sequência de etapas (Chapman *et al.*, 2000). Essas etapas são executadas de forma interativa. Assim, pelas entradas e respostas providas pelo usuário, a sequência da execução pode ser alterada. O encadeamento das ações, dependendo do objetivo e de como as informações se encontram, permite retorno a passos já realizados.

A etapa de Compreensão do Negócio procura identificar as necessidades e os objetivos do negócio do cliente e converter esse conhecimento numa tarefa de mineração de dados. Em adição, detecta eventuais problemas e/ou restrições que, se desconsideradas, poderão implicar na perda de tempo e esforço em obter respostas corretas para questões erradas. Essa tarefa compreende ainda descrição do cliente, seus objetivos e descrição dos critérios utilizados para determinar o sucesso do seu negócio.

Já a etapa de Compreensão dos Dados visa a identificação de informações que possam ser relevantes para o estudo e uma primeira familiarização com seu conteúdo, descrição, qualidade e utilidade. A coleção inicial dos dados procura adquirir a informação com a qual se irá trabalhar, relacionando suas fontes, o procedimento de leitura e os problemas detectados. Nessa tarefa, descreve-se ainda a forma como os

dados foram adquiridos, listando seu formato, volume, significado e toda informação relevante. Durante essa etapa, são realizadas as primeiras descobertas.

A etapa de Preparação dos Dados consiste numa série de atividades destinadas a obter o conjunto final de dados, a partir do qual será criado e validado o modelo. Nessa fase, são utilizados programas de extração, limpeza e transformação dos dados. Compreende a junção de tabelas e a agregação de valores, modificando seu formato, sem mudar seu significado a fim de que reflitam as necessidades dos algoritmos de aprendizagem. Na etapa de Modelagem, são selecionadas e aplicadas as técnicas de mineração de dados mais apropriadas, dependendo dos objetivos pretendidos. A criação de um conjunto de dados para teste permite construir um mecanismo para comprovar a qualidade e validar os modelos que serão obtidos. A modelagem representa a fase central da mineração, incluindo escolha, parametrização e execução de técnicas sobre o conjunto de dados visando à criação de um ou vários modelos.

A etapa de Avaliação do Modelo consiste na revisão dos passos seguidos, verificando se os resultados obtidos vão ao encontro dos objetivos, previamente, determinados na Compreensão do Negócio, como também as próximas tarefas a serem executadas. De acordo com os resultados alcançados, na revisão do processo, decide-se pela sua continuidade ou se deverão ser efetuadas correções, voltando às fases anteriores ou ainda, iniciando novo processo.

A etapa de Distribuição (Aplicação) é o conjunto de ações que conduzem à organização do conhecimento obtido e à sua disponibilização de forma que possa ser utilizado eficientemente pelo cliente. Nessa fase, gera-se um relatório final para explicar os resultados e as experiências, procurando utilizá-los no negócio.

6 METODOLOGIA

6.1 Mapeamento dos bancos de dados biológicos de origem brasileira relacionadas à flora

O mapeamento dos bancos de dados foi realizado através do *Database Commons*, que fornece um extenso catálogo de bancos de dados biológicos globais, o qual não só permite que os usuários acessem e recuperem facilmente uma variedade completa de bases de dados biológicos em todo o mundo, como também proporciona uma visão global para uma compreensão mais profunda do amplo impacto desses recursos na vida, na medicina e nas ciências da saúde (Ma *et al.*, 2023).

Como o objetivo é identificar os bancos de dados biológicos de origem brasileira, inicialmente foi feita uma delimitação espacial a fim de retornar somente os que foram desenvolvidos por Instituições do Brasil. Após essa etapa, foram eliminados os bancos de dados que não se destinavam a informações relacionadas a plantas, os que estão inacessíveis e os que tem como foco espécies não nativas.

Ao fim, os resultados foram agrupados de acordo com a natureza dos dados tratados pelas bases de dados biológicas aceitas.

6.2 Descrição das informações obtidas a partir do mapeamento

Após o mapeamento, foi feita uma coleta detalhada das informações a respeito das bases de dados encontradas. O objetivo desta coleta foi levantar informações como: objetivo, os tipos de moléculas que utilizam, o idioma e a instituição responsável pelo desenvolvimento, resultando em uma análise descritiva das características de cada base de dados, destacando vantagens e limitações.

Outra informação levada em consideração foi a experiência de usuário. A medição da experiência de usuário é uma ferramenta capaz de prover *feedback* a respeito de como o usuário interage com o sistema (Maia e Furtado, 2014). Ramírez-Acosta (2017) afirma que a experiência do usuário depende das necessidades deste, uma vez que ao desenvolver uma ferramenta de difícil utilização, o usuário pode se frustrar e desistir de usá-la. Isso faz com que seja importante garantir um sistema que seja intuitivo, autoexplicativo e que seus elementos estejam dispostos de forma lógica, de forma que vá utilizá-lo consiga realizar suas tarefas (Ramírez-Acosta, 2017).

Dessa forma, para avaliar essa experiência foi escolhida a metodologia utilizada por Yang *et al.* (2004), cujo instrumento de avaliação é composto por cinco dimensões, os quais envolvem a usabilidade, utilidade do conteúdo, adequação da informação, acessibilidade e interação. Cada um deles possui um conjunto de fatores que serão medidos. A medição dessas dimensões é feita por meio da Escala Likert de 5 pontos, em que “1” representa “discordo totalmente” e “5” representa “concordo totalmente”.

Segue abaixo os fatores a serem medidos de acordo com cada dimensão:

1) Usabilidade:

- *Hiperlinks* bem-organizados
- Adequação dos recursos de segurança
- Facilidades de pesquisa
- Funções de pesquisa personalizadas
- Apresentação de informações personalizadas
- Confidencialidade das informações do cliente

2) Adequação da informação:

- Abrangência das informações em relação a outros portais
- Conteúdo completo
- Informações suficientes para clientes potenciais e existentes
- Descrição completa do produto/serviço
- Informações detalhadas de contato

3) Utilidade do conteúdo:

- Conteúdo exclusivo
- Informações relevantes para o cliente
- Dicas valiosas sobre produtos/serviços
- Opiniões profissionais confiáveis
- Informações atualizadas

4) Acessibilidade:

- Acessibilidade do site
- Alta velocidade de carregamento de página

5) Interação:

- *Feedback* interativo entre clientes e a empresa
- Serviços de acompanhamento para clientes

- Fórum de quadro de mensagens para clientes para clientes/empresa

Ao fim, a pontuação dos fatores será somada e dividida pela sua quantidade, a fim de obter uma média. O valor resultante será adaptado para uma melhor compreensão, em que:

- 1 ou 2 pontos: Insatisfatório
- 3 ou 4 pontos: Relevante
- 5 pontos: Satisfatório

Essas informações serão representadas em uma tabela de forma clara e concisa e servirão para auxiliar no levantamento de requisitos e funcionalidades que o repositório deverá ter.

6.3 Desenvolvimento do Repositório

O desenvolvimento do sistema foi de acordo com a metodologia de desenvolvimento por prototipação. Segundo Pressman e Maxim (2016), trata-se de uma técnica que tem como intuito identificar e definir os requisitos necessários para a construção de um *software*, por conseguinte, auxilia no melhor entendimento do que está sendo desenvolvido. Tal metodologia pode ser empregada desde que o sistema que será desenvolvido tenha como características consultas dinâmicas, forte interação com as pessoas ou possua um algoritmo ou processamento combinatório que tenha a necessidade de ser desenvolvido de forma evolucionária. O seu processo deve ser executado de forma iterativa, obedecendo a um ciclo (Figura 3).

Figura 3: Modelo de processo de prototipação

Fonte: Adaptado de Pressman e Maxim, 2016.

Dessa forma, como etapa inicial, tem-se a coleta de requisitos. Nesta, os envolvidos no projeto, usuários e desenvolvedores, estão em constante interação, com o intuito de definir os objetivos gerais do *software*, e levantar os requisitos e funcionalidades do sistema, permitindo identificar áreas que necessitam de uma definição mais detalhada. A partir do que é definido pelo usuário, ocorre a elaboração de um “projeto rápido”. Este projeto leva à construção de um protótipo, o qual é avaliado pelo usuário, gerando então *feedbacks* que são utilizados para refinar os requisitos. O protótipo é criado de modo a servir como base para a definição de

requisitos. Ao fim, este será descartado, ao menos em parte, e o software real será desenvolvido (Pressman e Maxim, 2016).

Uma variedade de ferramentas foi utilizada durante o processo de desenvolvimento, sendo responsáveis por auxiliar na execução das etapas. Na “Coleta de Requisitos” foi utilizado o sistema StarUML, cuja função é gerar diagramas necessários para uma manutenção eficiente do processo de desenvolvimento. Trata-se de um sistema de uso livre que utiliza a Linguagem de Modelagem Unificada (*Unified Modeling Language/UML*) para realizar a documentação. Somado a isso, foi utilizado o *software Mysql WorkBench* no processo de diagramação da base de dados a ser construída. Também foram definidos os critérios de busca e seleção dos bancos de dados a serem utilizados.

Para correlacionar os bancos de dados selecionados, foram utilizadas as técnicas de *Web Scrapping* e *Extract, Transform and Load (ETL)*. A partir disso, foi desenvolvido um script em Python para a realização da extração de dados com o módulo *scrapy*, que permite extrair dados de páginas web de forma automatizada e eficiente. Essas extrações foram feitas por meio de consultas nas APIs dos bancos de dados escolhidos. A *Application Programming Interface/Interface* de programação de aplicações (API) permite realizar consultas e obter respostas em formato *JavaScript Object Notation (JSON)*, que podem ser manipuladas e analisadas com Python. Após isso, foi utilizado o módulo *request*, o qual permite enviar e receber requisições em *Hypertext Transfer Protocol/ Protocolo de Transferência de Hipertexto (HTTP)*. Já os processos ETL entraram com o intuito de integrar os dados. Este atua combinando os dados de diversas fontes, resultando em um único conjunto de dados, os quais poderão ser carregados em um *data warehouse, data lake* ou outro sistema.

A seguir, na construção do “Projeto Rápido”, foi utilizado o *software* Figma para a criação das telas do sistema, de forma a facilitar a identificação dos elementos gráficos a serem utilizados. O “Projeto Rápido” criado no Figma resultou em um protótipo, o qual foi avaliado e definido os requisitos a partir dos dados levantados.

6.4 Matriz de Validação

Abaixo segue a matriz de amarração que apresenta as relações entre os objetivos, a metodologia e os produtos desenvolvidos referentes ao desenvolvimento do repositório (Tabela 1).

Tabela 1: Matriz de amarração do desenvolvimento do repositório

Objetivos	Metodologia	Produtos
Mapear os bancos de dados biológicos de origem brasileira relacionados à flora	Mapeamento dos bancos de dados biológicos de origem brasileira relacionadas à flora	Texto Dissertativo
Descrever as informações obtidas a partir desse mapeamento	Descrição das informações obtidas a partir do mapeamento	Artigos
Desenvolvimento do repositório on-line para a análise e exibição dos dados associados	Desenvolvimento do repositório on-line	Moléculas da Amazônia
		Canvas
		SWOT

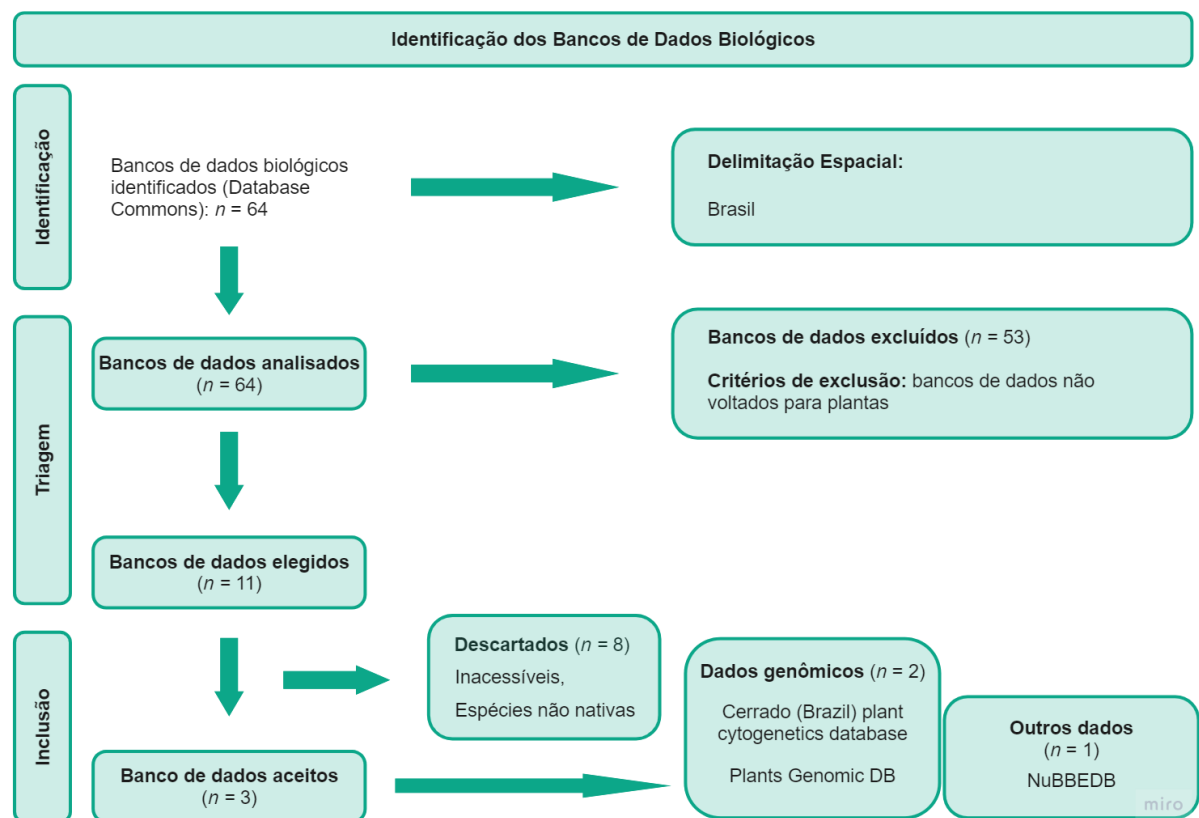
Fonte: elaborado pela autora (2025).

7 RESULTADOS

7.1 Mapeamento dos bancos de dados biológicos voltados para flora

Com a delimitação espacial, foram encontrados ao todo 64 bancos de dados biológicos de origem brasileira. Ao utilizar o critério de exclusão, houve um retorno de 11 bancos de dados voltados para plantas. Desses onze, 7 foram descartados, uma vez que já não estavam mais acessíveis ou que não trabalhavam com espécies nativas brasileiras, resultando em 4 bancos de dados. Este resultado pode ser dividido em três naturezas de dados, sendo eles genômicos, toxicológicos e outros (Figura 4).

Figura 4: Identificação dos Bancos de Dados Biológicos



Fonte: Elaborado pela autora (2024).

7.2 Descrição das informações obtidas a partir do mapeamento

Os bancos de dados encontrados desempenham papéis distintos na pesquisa científica. O "Cerrado (Brazil) *plant cytogenetics database*" é um banco de dados desenvolvido com o intuito de avaliar o conhecimento citogenético das espécies vegetais do cerrado, fornecendo um banco de dados que inclui todos os atributos nessa área.

Dentre as informações citogenéticas das plantas, temos o número cromossômico, fórmula cariotípica, bandas C, CMA e AgNOR, hibridização *in situ* fluorescente (FISH), tamanho do genoma (valor C) e referências bibliográficas (Roa; Telles, 2017). Disponível em inglês, esse banco de dados difere da proposta do "Moléculas da Amazônia", uma vez que além de ter como foco espécies da flora do Cerrado e seus respectivos dados citogenéticos, a falta de informações sobre ecologia, distribuição geográfica detalhada dentre outros dados moleculares tornam esse sistema dependente de outros em relação à obtenção de informações complementares.

O "Plant Genomics DB", da Universidade Estadual de São Paulo (UNESP), foi desenvolvido com o intuito de catalogar e conceder o acesso aos recursos genômicos vegetais. Disponível em inglês, o site te permite acessar informações como: modelos genéticos acompanhados de anotações integradas na visualização InterProScan de cada região codificantes, anotação de elementos transponíveis e repetições, assembleias de PASA, anotação de tRNA e rRNA e dados de RNAseq alinhados, além de outros recursos.

Por fim, o "NuBBEDB", também da UNESP, oferece dados químicos e biológicos de produtos naturais encontrados na biodiversidade brasileira, com informações em português e inglês. Este é o sistema mais próximo do que é proposto por este projeto. O NuBBEDB fornece informações detalhadas sobre as estruturas químicas de compostos naturais brasileiros, permitindo a visualização e a análise de suas propriedades moleculares. A partir deste banco de dados é possível analisar dados físico-químicos das estruturas moleculares, dados sobre suas atividades biológicas, dados espectroscópicos e as referências científicas relacionadas. No entanto, apesar de ser possível realizar buscas por espécie e por localização, não é possível obter informações mais detalhadas sobre as espécies buscadas e sobre a distribuição geográfica, tornando necessário um constante acesso a múltiplos bancos de dados para obter essas informações adicionais, bem como os bancos de dados anteriormente mencionados. Segue abaixo as informações gerais (Tabela 2) e a experiência do usuário (Tabela 3) referente aos bancos de dados levantados.

Tabela 2: Informações Gerais dos Bancos de Dados Biológicos

Base de dados	Objetivo	Tipo de dados	Idioma	Instituição	Referência
Cerrado (Brazil) plant cytogenetics database	Integrar e compartilhar todas as informações citogenéticas publicadas para espécies do Cerrado	RNA	Inglês	Universidade Federal de Goiás - UFG	Roa; Telles, 2017
Plant Genomics DB	Catalogar e fornecer acesso a todos os recursos genômicos vegetais, bem como genomas relacionados	DNA	Inglês	Universidade Estadual Paulista - UNESP	Abreu <i>et al.</i> , 2023
NuBBEDB	Fornecer informações químicas e biológicas da biodiversidade brasileira	Outro	Português/Inglês	Universidade Estadual Paulista - UNESP	Pilon <i>et al.</i> , 2017

Fonte: Elaborado pela autora (2024).

Tabela 3: Avaliação da Experiência de Usuário dos Bancos de Dados

Base de dados	Usabilidade	Utilidade do conteúdo	Adequação da informação	Acessibilidade	Interação
Cerrado (Brazil) plant cytogenetics database	Relevante	Satisfatório	Relevante	Insatisfatório	Insatisfatório
Plant Genomics DB	Satisfatório	Satisfatório	Insatisfatório	Relevante	Relevante
NuBBEDB	Satisfatório	Satisfatório	Relevante	Satisfatório	Satisfatório

Fonte: Elaborado pela autora (2024) baseado em Yang *et al.* (2004).

7.3 Desenvolvimento do repositório

Foi utilizado a metodologia de desenvolvimento por prototipação para o desenvolvimento do “Moléculas da Amazônia”. Esse repositório está sendo desenvolvido a fim de integrar em um único sistema todas as informações referentes às espécies da flora Amazônica e suas respectivas moléculas. Baseado em Pressman e Maxim (2016), a prototipação é realizada em seis etapas: coleta de requisitos, projeto rápido, construção do protótipo, avaliação do protótipo, refinamento do protótipo e construção do produto.

Para a etapa de coleta de requisitos foi necessário o uso do sistema *StarUML* para a geração dos diagramas utilizados para a manutenção eficiente do processo de desenvolvimento. Somado a isso, o *software* *Mysql WorkBench* foi utilizado para a diagramação da base de dados do Moléculas da Amazônia.

Ao todo, quatro bancos de dados foram escolhidos para compor o repositório, dos quais um, o GBIF, um banco de dados que fornece informações a respeito das espécies presentes na flora brasileira e as suas ocorrências, e três, NCBI, ChEMBL e NuBBEDB, bancos de dados biológicos que fornece informações a respeito das moléculas presentes nelas.

Por meio da ferramenta API, foi realizada a extração de dados automatizada no NuBBEDB de informações referentes às biomoléculas presentes na flora amazônica. Utilizando o Scrapy em Python, foi desenvolvido um script que percorreu as páginas deste banco de dados, coletando informações sobre quais biomoléculas estão presentes na flora da Amazônia e em quais espécies estas podem ser encontradas. Esta etapa teve como intuito obter uma lista preliminar dessas moléculas e suas respectivas espécies. A partir desses dados levantados, a próxima etapa consistiu em realizar consultas complementares em outras fontes de dados.

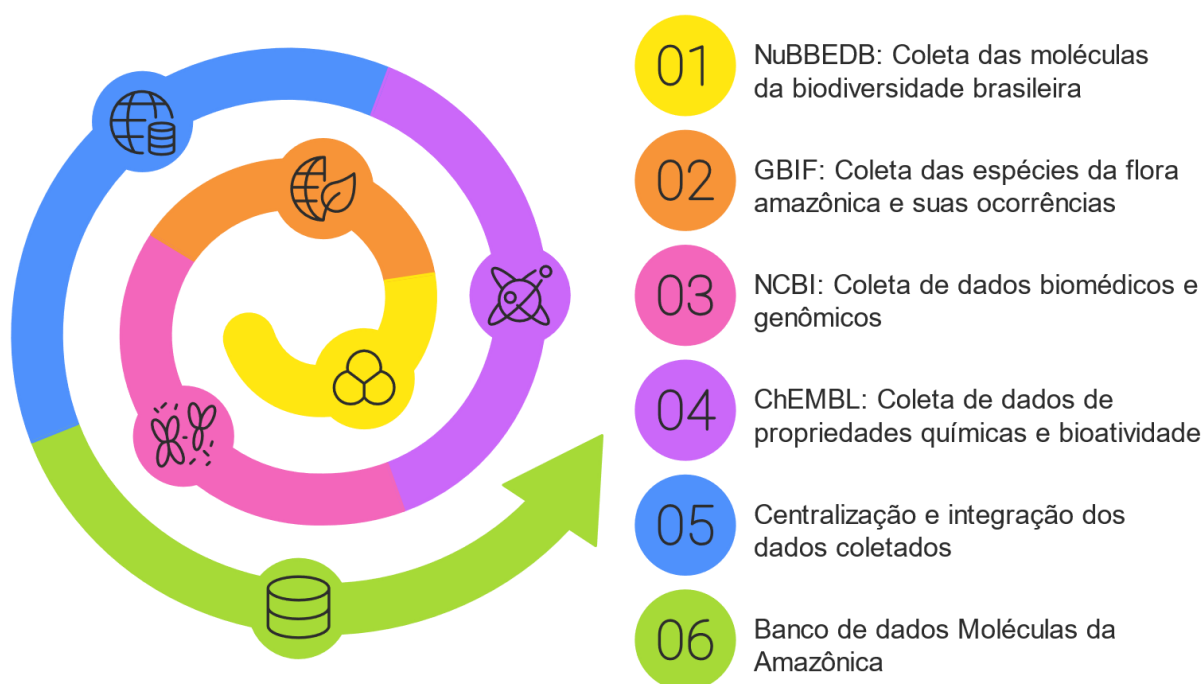
Com as espécies coletadas no NuBBEDB, foi feita uma consulta na API disponibilizada pelo GBIF. O objetivo desta consulta foi complementar os dados obtidos na etapa anterior com informações taxonômicas detalhadas (família, gênero e espécie) e dados geográficos que indicam a distribuição dessas espécies na Amazônia.

No que se refere às moléculas bioativas coletadas, posteriormente foram feitas para consultas no NCBI e ChEMBL. O ChEMBL contribuiu com informações adicionais sobre as propriedades físico-químicas das moléculas, bem como sua bioatividade. Já as consultas no NCBI, disponibilizaram dados genéticos e moleculares,

particularmente aquelas relacionadas a sequências de DNA e RNA associadas às espécies e moléculas amazônicas. Tais consultas permitiram a integração de dados biológicos e químicos complementares no Moléculas da Amazônia, contribuindo para o enriquecimento do conjunto de informações disponíveis.

Além disso, para a conversão dos dados coletados em formatos compatíveis para o carregamento destes no banco de dados, foi utilizada os processos ETL por meio de scripts em Python. Isso permitiu padronizar e integrar os dados de forma correta (Figura 5).

Figura 5: Bancos de Dados utilizados.



Fonte: Elaborado pela autora (2025).

Com os dados coletados, o projeto seguiu para a segunda etapa, que consistiu na construção do “Projeto Rápido” por meio do *software* Figma. Através deste, foram criadas as telas do sistema com o intuito de otimizar a identificação dos elementos gráficos que seriam utilizados. Em seguida, foi feita a construção do protótipo, terceira etapa.

O protótipo do repositório Moléculas da Amazônia conta com filtros que te permitem realizar buscas por moléculas, espécies, localização específica, referências e informações químicas (Figura 6). A busca por moléculas pode ser feita através do

seu nome comum, nome IUPAC, inchikey, tipo de moléculas e/ou fórmula moléculas. Nessa mesma linha, a busca pelas informações disponíveis te dá a opção de procurar por massa molar, CLOGP, violações de lipinski, doadores de h-bond, massa monoisotópica, TPSA, aceitadores de h-bond e ligações rotáveis. Já a busca por espécie é feita através da sua família, do seu gênero e/ou da espécie.

Além disso, é possível realizar a busca selecionando a forma de vida desejada. Com o filtro de localização específica, o usuário poderá buscar por bioma, região, estado ou cidades e ter acesso às espécies desses locais e seus respectivos compostos. Este emerge como o principal diferencial desse repositório, pois com ele é possível obter informações a respeito da distribuição geográfica das espécies e de suas moléculas, função essa não presente nos repositórios existentes atualmente. As opções disponíveis para a busca por referência são o nome da revista, o título, o ano de publicação e/ou nome do autor.

Figura 6: Busca por filtros.

The screenshot displays the search interface for 'MOLÉCULAS DA AMAZÔNIA'. The header includes navigation links: Home, Sobre o projeto, Nossa equipe, Entre em contato, Pesquisar molécula, Entrar, and Cadastrar-se. The main search area is organized into several sections:

- Molécula:** Includes input fields for 'Nome comum', 'InchiKey da molécula', 'Tipo da molécula', 'Nome IUPAC', and 'Fórmula molecular'.
- Espécie:** Includes a 'Família' field, 'Gênero', 'Espécie', and a dropdown for 'Formas de vida'.
- Localização específica:** Features a 'Filtrar por: Cidade' dropdown and a 'Selecione as cidades' dropdown. Below these are two active filters: 'Porto Velho' and 'Vilhena'.
- Referência:** Includes a 'Jornal' field, 'Título', 'Ano' (with 'Até' and another 'Ano' field), and 'Autor'.
- Informações químicas:** A grid of filters for chemical properties:

<input type="text"/>	≥	Massa Molar	≤	<input type="text"/>	≥	Massa Monoisotrópica	≤	<input type="text"/>
<input type="text"/>	≥	CLOGP	≤	<input type="text"/>	≥	TPSA	≤	<input type="text"/>
<input type="text"/>	≥	Violações de lipinski	≤	<input type="text"/>	≥	Aceitadores de h-bond	≤	<input type="text"/>
<input type="text"/>	≥	Doadores de h-bond	≤	<input type="text"/>	≥	Ligações rotáveis	≤	<input type="text"/>

A green 'Pesquisar' button is located at the bottom right of the form.

Fonte: Elaborado pela autora (2024).

A tabela a seguir apresenta uma descrição dos campos utilizados para a busca no repositório, incluindo o tipo de dado e exemplos de valores para cada um deles (Tabela 4).

Tabela 4: Tipos de dados para realização das buscas nos campos de busca do repositório.

Campo	Tipo de dado	Exemplo
Nome comum	Texto (Chave alfanumérica)	3,4-dihydroxybenzoic acid geranyl ester
Inchikey	Texto (Chave alfanumérica)	OVDDPLULMGGUFM-UKTHLTGXSA-N
Tipo de Molécula	Texto	Aromatic derivatives
Nome IUPAC	Texto (Chave alfanumérica)	(2E)-3,7-dimethylocta-2,6-dien-1-yl 3,4-dihydroxybenzoate
Fórmula Molecular	Texto (Chave alfanumérica)	C17H22O4
Família	Texto	Piperaceae
Gênero	Texto	Piper
Espécie	Texto	crassinaervium
Formas de vida	Texto	Arbusto
Filtra por Cidade	Texto	Porto Velho
Journal	Texto	Journal of Medicinal Chemistry
Título	Texto	Evaluation of Curcumin's Antioxidant Activity
Ano	Texto	2023
Autor	Texto	Silva
Massa Molar	Número (Decimal)	368.40
CLOGP	Número (Decimal)	3.25
Violações de Lipinski	Número (Inteiro)	1
Doadores de H-bond	Número (Inteiro)	3

Massa Monoisotópica	Número (Decimal)	368.2292
TPSA (Topological Polar Surface Area)	Número (Decimal)	70.5
Aceitadores de H-bond	Número (Inteiro)	4
Ligações rotáveis	Número (Inteiro)	8

Fonte: Elaborado pela autora (2025).

As informações pesquisadas são apresentadas através de cards intuitivos. Estes organizam as moléculas e permitem que o usuário visualize os principais dados de forma rápida, clara e objetiva (Figura 7).

Figura 7: Cards Intuitivos.

<p>Piperaceae Piper crassinervium 3-O-geranyl-4-hydroxybenzoic acid Aromatic Acid</p>	<p>Piperaceae Piper crassinervium p-hydroxybenzoic acid-geranyl ester Aromatic Derivatives</p>
<p>Piperaceae Piper crassinervium 3,4-dihydroxybenzoic acid geranyl ester Aromatic Derivatives</p>	<p>Piperaceae Piper crassinervium 3,4,5-Trimethoxydihydrocinnamic acid Phenylpropanoids</p>
<p>Piperaceae Piper crassinervium Sakuranetin; 5,4'-dihydroxy-7-methoxyflavanon Flavonoids Flavanone</p>	<p>Piperaceae Piper crassinervium 4-Hydroxy-(3',7'-dimethyl-1'-oxo-octa-2'-E-6'-dienyl)benzoic acid Aromatic Acid</p>

Fonte: Elaborado pela autora (2025).

Diferente dos bancos de dados biológicos existentes, com o Moléculas da Amazônia será possível visualizar tanto os dados referentes às espécies (Figura 8), quanto das moléculas (Figura 9) a serem pesquisadas. Tais informações são apresentadas através de blocos organizados, contendo a maior quantidade de informações catalogadas. Tal funcionalidade permite que o usuário obtenha em um

único lugar as informações desejadas, tirando a necessidade de acessar múltiplos bancos de dados para obtê-las.

Figura 8: Informações sobre as espécies.

The screenshot displays a web application interface for 'MOLÉCULAS DA AMAZÔNIA'. The header includes navigation links: Home, Sobre o projeto, Nossa equipe, Entre em contato, **Pesquisar molécula**, Entrar, and Cadastrar-se. The breadcrumb trail shows: Home > Pesquisa molécula > Piperaceae Piper crassinervium. A 'Plantas' button is active, and 'Moléculas relacionadas' is visible. The main heading is 'Piperaceae Piper crassinervium'. Below it, the 'Informações Gerais' section lists: 'Dados básicos', 'Forma de vida: Arbusto', 'Habitat: Terrícola', 'Origem: Nativa', and 'Endêmica: Não endêmica'. The 'Arquivos e imagens' section has a 'Colapsar' button. The 'Incidência geográfica' section is currently empty. At the bottom, there is a note 'Adquirida quando isolada da planta' and a 'Galeria de Imagens' section with left and right navigation arrows.

Fonte: Elaborado pela autora (2025).

Figura 9: Informações sobre as moléculas.

MOLÉCULAS DA AMAZÔNIA Home Sobre o projeto Nossa equipe Entre em contato [Pesquisar molécula](#) Entrar Cadastrar-se

Home > Pesquisa molécula > Piperaceae Piper crassinervium

Plantas [Moléculas relacionadas](#)

Piperaceae Piper crassinervium

[Colapsar](#)

Informações Gerais

Dados básicos

Nome comum: 3,4-dihydroxybenzoic acid geranyl ester
 Nome IUPAC: (2E)-3,7-dimethylocta-2,6-dien-1-yl 3,4-dihydroxybenzoate
 Fórmula molecular: C17H22O4
 Inchi: 1S/C17H22O4/c1-12(2)5-4-6-13(3)9-10-21-17(20)14-7-8-15(18)16(19)11-14/h5,7-9,11,18-19H,4,6,10H2,1-3H3/b13-9+
 Inchikey: OVDDPLULMGGUFM-UKTHLTGXSA-N
 Smile: O=C(OC=C(C)/CC/C=C(C)/C)C1=CC(O)=C(O)C=C1
 Espécie: Piperaceae Piper crassinervium
 Propriedades biológicas: Inibição de protease; Antiviral
 NCBI Compound ID: 11358243
 ChEMBL ID: ChEMBL340845

Arquivos e imagens

[Colapsar](#)

Estrutura

Arquivos relacionados

Nome	Estrutura	Tipo de arquivo	Link de Download
Modelo 1	Estrutura 2D	PNG	Download
Modelo 2	Estrutura 3D	PNG	Download

< 1 >

Fonte: Elaborado pela autora (2025).

O usuário poderá ter acesso ao seu histórico de pesquisa, de modo a facilitar o seu desenvolvimento durante o uso do sistema

8 DISCUSSÃO

A flora amazônica possui um vasto potencial biotecnológico, através de uma diversidade biológica que fornece inúmeras oportunidades para o desenvolvimento de produtos inovadores e sustentáveis. Reforçando essa visão, um estudo produzido por Calazans *et al.* (2021) revela o potencial tecnológico presente nas sementes florestais, o qual é explorado por diferentes setores, como o alimentício, cosmetológico e o farmacêutico. Espécies muito presentes na cultura amazônica como a Jambu (*Acmella oleracea*) e o Cacau (*Theobroma cacao*) também são amplamente explorados por esses setores, devido a gama de atividades biológicas presentes nestas, principalmente a atividade antioxidante (Costa *et al.*, 2024; Reis *et al.*, 2024).

Sampaio Neto *et al.* (2020) enfatizam esse potencial através da utilização de óleos extraídos da biodiversidade brasileira para a criação de cadeias produtivas não madeireiras. Os autores também afirmam que os avanços tecnológicos desempenham um papel crucial na preservação dessas espécies, ao mesmo tempo em que aumentam o valor agregado de seus derivados. Logo, a riqueza amazônica proporciona não só uma fonte de recursos naturais, como também um campo fértil para o avanço na biotecnologia.

No entanto, o mapeamento realizado revelou uma lacuna no que tange a disponibilidade de dados biológicos sobre a flora amazônica. A identificação e a categorização dos bancos de dados demonstraram um foco em dados genômicos, e reforçaram a falta de integração de dados. Um banco de dados biológicos que compreendesse não só sequências genéticas, mas também informações sobre as espécies coletadas, suas estruturas proteicas e redes de interações moleculares, por exemplo, seria fundamental para preencher essa lacuna. Desse modo, o desenvolvimento de um banco de dados que integre informações variadas a respeito da flora amazônica não apenas enriquece o seu entendimento, como também otimiza o processo de busca.

Ao comparar os diferentes bancos de dados selecionados, fica evidente a dispersão de dados. Essa dispersão torna necessário que o pesquisador acesse múltiplas fontes para obter as informações desejadas, ocasionando um gasto de tempo e esforço maior e o Moléculas da Amazônia se diferencia ao integrar todas essas informações em um único ambiente. Essa abrangência permite que o repositório

atenda tanto demandas científicas quanto tecnológicas, possibilitando identificar biomoléculas com potencial de aplicações biotecnológicas (Figura 10).

Figura 10: Comparação entre os dados disponibilizados pelos bancos de dados biológicos levantados e o Moléculas da Amazônia.

Banco de Dados	Dados de Espécies	Ocorrências	Dados Biológicos	Dados Bibliográficos	Dados Patentários
Reflora	✓	✓	✗	✓	✗
GBIF	✓	✓	✗	✓	✗
Cerrado (Brazil) plant cytogenetics database	✗	✗	✓	✗	✗
Plant Genomics DB	✗	✗	✓	✗	✗
NuBBEDB	✗	✗	✓	✓	✗
Moléculas da Amazônia	✓	✓	✓	✓	✓

Fonte: Elaborado pela autora (2025).

Além disso, o mapeamento evidenciou a ausência de instituições de pesquisa do Norte do Brasil. De acordo com Negri e Squeff (2016), a infraestrutura de pesquisa em território nacional é fragmentada e insuficiente, composto por uma quantidade significativa de laboratório de tamanho limitado e mal equipados.

Os autores também afirmam que há uma concentração no número de infraestrutura nas regiões Sul e Sudeste. Tal fato revela a falta de infraestrutura nos campos científico e tecnológico presente na região Norte, acarretando na limitação da capacidade local de conduzir pesquisas de ponta sobre a biodiversidade amazônica, apesar de estar diretamente inserida neste bioma.

Ademais, a desigualdade na distribuição de recursos e investimentos também contribuem para esse cenário. Segundo Negri e Squeff (2016), há um favorecimento de regiões mais desenvolvidas e regiões como o Norte recebem uma fração inferior de investimento comparado com outras regiões como o Sul e o Sudeste. Essa disparidade não só aumenta a dificuldade de executar projetos inovadores na região, como também compromete a geração de novos conhecimentos e tecnologia, os quais poderiam desempenhar um importante papel no desenvolvimento socioeconômico e ambiental local.

Esses fatores podem afetar o entendimento da Amazônia, uma vez que os recursos limitados impedem a exploração de sua biodiversidade. Com isso, o conhecimento sobre sua flora permanece fragmentado e subdesenvolvido.

Quanto à experiência de usuário, os resultados obtidos revelam um panorama variado referente à qualidade e eficácia. Cada um apresentou pontos fortes e fracos, indicando que, embora os conteúdos num geral sejam úteis e relevantes, há uma deficiência considerável em relação à acessibilidade, interação e adequação das informações. Isso pode afetar diretamente a experiência e o interesse por parte do usuário, tendo em vista que ao não ter suas necessidades supridas, este pode desistir de utilizar a ferramenta (Ramírez-Acosta, 2017).

Levando em consideração que o sistema a ser desenvolvido deve suprir as necessidades do usuário (Maia e Furtado, 2014), essa análise revelou a importância de um desenvolvimento adequado das dimensões avaliadas. Portanto, para o desenvolvimento do “Moléculas da Amazônia”, foi crucial identificação dessas falhas, pois a partir deste resultado foi possível notar a necessidade de um foco maior quanto a acessibilidade e interatividade do sistema, de modo a garantir que o repositório seja de fácil acesso e que tenha uma interação efetiva do conteúdo por parte do usuário.

A metodologia de Pressman e Maxim (2026) auxiliou na construção de um protótipo funcional do repositório Moléculas da Amazônia. Quanto ao seu desenvolvimento, a adoção das técnicas *Web Scrapping* e ETL possibilitou a coleta, a integração e o armazenamento de dados dispersos em quatro diferentes bancos de dados: o banco de dados GBIF para a extração de dados sobre espécies e suas ocorrências, e os bancos de dados biológicos NuBBEDB, ChEMBL e NCBI para a coleta de dados biológicos. Vale ressaltar que devido a instabilidades no sistema, o

banco de dados Re flora não foi utilizado nesse primeiro momento. Ademais, por meio do Figma foi possível ter uma visualização mais clara e organizada das telas a serem desenvolvidas.

Diante disso, o repositório Moléculas da Amazônia tem o potencial de gerar impactos significativos em diversos aspectos da sociedade e do meio ambiente.

Do ponto de vista ambiental, o acesso a informações detalhadas sobre a flora amazônica pode contribuir para uma melhor compreensão dos ecossistemas amazônicos e para a identificação de espécies com potencial medicinal, alimentício, ou industrial, sem comprometer a sua preservação. Isso pode levar a práticas de manejo mais sustentáveis e à conservação de habitats naturais.

No âmbito econômico, a exploração responsável dos recursos da flora amazônica pode gerar oportunidades de emprego e renda para as comunidades locais, especialmente para aquelas que dependem dos recursos naturais da região. Além disso, o desenvolvimento de novos produtos e tecnologias a partir das moléculas da flora amazônica pode impulsionar a economia regional e nacional, criando indústrias e mercados.

Do ponto de vista da saúde, a descoberta e o desenvolvimento de compostos medicinais e terapêuticos derivados de plantas amazônicas podem levar a avanços significativos no tratamento de doenças e na melhoria da qualidade de vida das pessoas. Além disso, a utilização desses recursos naturais pode contribuir para a medicina tradicional e para a diversificação das opções terapêuticas disponíveis.

Em termos de conhecimento, o acesso a dados biológicos da flora amazônica pode estimular a pesquisa científica e a inovação em uma variedade de campos, desde a biologia e a biotecnologia até a farmacologia e a ecologia. Isso pode levar a novas descobertas e avanços tecnológicos que beneficiam não apenas o Brasil, mas também a comunidade científica global.

De modo geral, o acesso a dados biológicos da flora amazônica através desse repositório on-line tem o potencial de gerar uma ampla gama de impactos positivos, desde a conservação do meio ambiente até o avanço da ciência, a melhoria da saúde pública e o desenvolvimento econômico e social.

9 CONCLUSÃO

Os resultados obtidos reforçaram a escassez de bancos de dados que forneçam informações sobre a flora amazônica, e revelam necessidade urgente de plataformas integradas que centralizem dados biológicos, químicos e taxonômicos, permitindo uma consulta mais eficiente e o avanço da pesquisa científica. A falta de repositórios acessíveis e atualizados limita a prospecção de compostos bioativos, por conseguinte, desperdiçando o vasto potencial biotecnológico e econômico da biodiversidade amazônica.

Neste trabalho, o repositório on-line de moléculas da flora amazônica foi desenvolvido até a quinta etapa do ciclo proposto por Pressman e Maxim (2016), o qual consiste no refinamento do protótipo. O banco de dados do repositório foi construído utilizando técnicas de *web scrapping* e ETL que foram capazes de integrar dados das espécies amazônicas e seus dados biológicos em um único banco de dados.

A sexta e última etapa do ciclo de desenvolvimento, que consiste na construção do produto, será executada dentro do doutorado em Biodiversidade e Biotecnologia ofertado pela Bionorte entre os anos de 2025 e 2029, onde pretende-se realizar a construção do repositório, bem como tornar a plataforma funcional e adaptada às necessidades dos pesquisadores.

Espera-se que o repositório se torne uma ferramenta acessível, contribuindo para a prospecção de novas moléculas com potencial biotecnológico e a utilização sustentável da biodiversidade amazônica. Assim, o projeto servirá como um ponto de partida para futuros avanços na biotecnologia, impactando na pesquisa farmacológica e no desenvolvimento de bioinsumos. O trabalho realizado até aqui estabelece as bases para o crescimento contínuo do repositório, que será expandido e aprimorado ao longo do doutorado, com o objetivo de atender a uma demanda da comunidade científica.

10 ENTREGÁVEIS DE ACORDO COM OS PRODUTOS DO TCC

O presente projeto resultou em dois artigos publicados, sendo eles: “**Prospecção Tecnológica do potencial terapêutico de moléculas extraídas de *Acmella oleracea*” (L.) R.K. Jansen (*Spilanthes oleracea*)**” (Apêndice B), publicado na Revista de Administração de Roraima - RARR, e “**Prospecção Tecnológica: potencial terapêutico de moléculas presentes no veneno de serpentes do gênero *Bothrops* sp., com ênfase na espécie *Bothrops jararaca***” (Apêndice A), apresentado em congresso e publicado no livro dos Anais do ProspeCT&I 2023 e na revista Caderno de Prospecções. Este artigo foi premiado como o melhor trabalho da sessão. A Revista de Administração de Roraima - RARR e o Cadernos de Prospecção possuem qualis B1 e B2, respectivamente, cumprindo um dos entregáveis obrigatórios exigidos pelo programa, que consiste na submissão/publicação de artigos em revista qualis B2 ou superior.

A seguir são apresentados os produtos desenvolvidos durante o desenvolvimento do TCC:

- 1) Matriz de SWOT (FOFA): Apêndice A;
- 2) Modelo de Negócio CANVAS: Apêndice B;
- 3) Artigo em avaliação por revista qualis B2: Apêndice C e D
- 4) Texto Dissertativo - TCC.
- 5) Registro de programa de computador (Moléculas da Amazônia).

Ademais, também foi escrito um terceiro artigo, cujo título é **Tendências científicas e tecnológicas da *Carapa guianensis*, popularmente conhecida como andiroba**. Este foi apresentado no congresso ProspeCT&I 2024 e atualmente está em processo de avaliação para publicação na revista Caderno de Prospecções (Anexo A).

Alguns outros trabalhos foram apresentados em Simpósios e Conferências. O artigo curto intitulado “**Moléculas da Amazônia: Construção do protótipo de um repositório on-line das espécies e biomoléculas da flora amazônica**” foi apresentado no 11ª Conferência Ibero Americana Computação Aplicada (CIACA 2024) e publicado no Livro de Atas do evento, e o resumo intitulado

“Molecules of the Amazon: Integration and Centralization of Data on the Amazon Flora and its Biomolecules” foi apresentado na sessão de posters do 17th Brazilian Symposium on Bioinformatics (BSB2024).

No momento, dois artigos estão em processo de análise para apresentação em evento: o artigo **Innovation in Biotechnology: Integrating and Centralizing Biological and Molecules Data from the Amazon Flora** para apresentar no Simpósio Brasileiro de Sistemas de Informação, e o artigo **Promoção da bioeconomia em território amazônico: utilização de técnicas de Web Scraping e ETL para mapeamento das moléculas da flora amazônica** para apresentar no Computer on the Beach 2025.

REFERÊNCIAS

ABREU, C. G. *et al.* Decoding the chromosome-scale genome of the nutrient-rich *Agaricus subrufescens*: a resource for fungal biology and biotechnology. **Res. Microbiol.**, [s.l.], v. 174, n. 8, 104116, nov./dez. 2023. DOI: 10.1016/j.resmic.2023.104116. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37573924/>. Acesso em: 13 abr. 2024.

AMBERG, A. (2013). *In silico* methods. In: VOGEL, H. G.; MAAS, J.; HOCK, F. J.; MAYER, D. **Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays**, 2. ed., [s.l.]: Springer Link, 2012. DOI: 10.1007/978-3-642-25240-2_55. Disponível em: https://doi.org/10.1007/978-3-642-25240-2_55. Acesso em: 07 mar. 2024.

ARAGÓN, L. E. A dimensão internacional da Amazônia: um aporte para sua interpretação. **Rev. NERA**, Presidente Prudente, n. 42, p. 14-33, mar. 2018. DOI: 10.47946/rnera.v0i42.5676. Disponível em: <https://doi.org/10.47946/rnera.v0i42.5676>. Acesso em: 10 abr. 2024.

BAH, S. Y.; MORANG'A, C. M.; KENGNE-OUAFO, J. A.; AMENGA-ETEGO, L.; AWANDARE, G. A. Highlights on the Application of Genomics and Bioinformatics in the Fight Against Infectious Diseases: Challenges and Opportunities in Africa. **Frontiers in Genetics**, [s.l.], v. 9, p. 1-15, nov 2018. DOI: 10.3389/fgene.2018.00575. Disponível em: <https://doi.org/10.3389/fgene.2018.00575>. Acesso em: 10 abr. 2024.

BARBOSA, R. A bioeconomia e a Amazônia. **Interesse Nacional**, São Paulo, v. 13, n. 1, p. 31-37, ago. 2020. Disponível em: https://www.escolhas.org/wp-content/uploads/2020/07/Revista-Interesse-Nacional_BIOECONOMIA_Irice-e-Escolhas_agosto2020.pdf. Acesso em: 10 abr. 2024.

BAXEVANIS, A. D.; BATEMAN, A. The importance of biological databases in biological discovery. **Current Protocols in Bioinformatics**, [s.l.], v. 50, n. 1, p. 1.1.1-1.1.6, jun. 2011. DOI: 10.1002/0471250953.bi0101s50. Disponível em: <https://doi.org/10.1002/0471250953.bi0101s50>. Acesso em: 07 mar. 2024.

BRASIL, Ministério do Meio Ambiente. **Biodiversidade Brasileira**. 2024. Disponível em: <https://antigo.mma.gov.br/biodiversidade/biodiversidade-brasileira.html>. Acesso em: 07 mar. 2024.

CALAZANS, C. C.; NUNES, V. V.; SOUZA, J. L.; SILVA-MANN, R. Sementes Florestais e seu Potencial Tecnológico: uma análise de metadados. **Cadernos de Prospecção**, Salvador, v. 14, n. 3, p. 794-809, set. 2021. DOI: 10.9771/cp.v14i3.42765. Disponível em: <https://periodicos.ufba.br/index.php/nit/article/view/42765>. Acesso em: 24 jan. 2024.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **Step-by-step data mining guide**. [s.l.]: SPSS Inc, 2000. 78p.

CHEN, J.; SWAMIDASS, S. J.; DOU, Y., BRUAND, J.; BALDI, P. ChemDB: A public database of small molecules and related cheminformatics resources.

Bioinformatics, [s.l.], v. 21, n. 22, p. 4133-4139, nov. 2005. DOI: 10.1093/bioinformatics/bti683. Disponível em: <https://doi.org/10.1093/bioinformatics/bti683>. Acesso em: 07 mar. 2024.

COLQUITT, R. B.; COLQUHOUN, D. A.; THIELE, R. H. (2011). In silico modelling of physiologic systems. **Best Practice and Research Clinical Anaesthesiology**, [s.l.], v. 25, n. 4, p. 499-510, dez. 2011. DOI: 10.1016/j.bpa.2011.08.006. Disponível em: <https://doi.org/10.1016/j.bpa.2011.08.006>. Acesso em: 07 mar. 2024.

COSTA, C. B.; SILVA, K. A.; MIRANDA, M. R. Prospecção Tecnológica do potencial terapêutico de moléculas extraídas de *Acmella oleracea* (L.) R.K. Jansen (*Spilanthes oleracea*). **Revista de Administração de Roraima - RARR**, Boa Vista, v. 15, n. 1, p. 1-17, mar. 2024. DOI: 10.18227/2237-8057rarr.v15i1.8088. Disponível em: <https://doi.org/10.18227/2237-8057rarr.v15i1.8088>. Acesso em: 10 jun. 2024.

DIAS, R. F.; DE CARVALHO, C. A. A. Bioeconomia no Brasil e no mundo: Panorama atual e perspectivas. **Rev. Virtual Quim.**, [s.l.], v. 9, n. 1, p. 410-430, dez. 2016. DOI: 10.21577/1984-6835.20170023. Disponível em: <https://rvq.s bq.org.br/pdf/v9n1a23>. Acesso em: 12 abr. 2024.

DINIZ, W. J. S.; CANDURI, F. (2017). Bioinformatics: An overview and its applications. **Genetics and Molecular Research**, [s.l.], v. 16, n. 1, p. 01-21, mar. 2017. DOI: doi.org/10.4238/gmr16019645. Disponível em: [10.4238/gmr16019645](https://doi.org/10.4238/gmr16019645). <https://doi.org/10.4238/gmr16019645>. Acesso em: 20 abr. 2024.

GAUTHIER, J.; VINCENT, A. T.; CHARETTE, S. J.; DEROME, N. A brief history of bioinformatics. **Briefings in Bioinformatics**, [s.l.], v. 20, n. 6, p. 1981-1996, nov. 2019. DOI: 10.1093/bib/bby063. Disponível em: <https://doi.org/10.1093/bib/bby063>. Acesso em: 20 abr. 2024.

GBIF - Global Biodiversity Information Facility. **What is GBIF**, 2024. Disponível em: <https://www.gbif.org/what-is-gbif>. Acesso em: 10 jun. 2024.

GUEDES, I. A.; PEREIRA, F. S. S.; DARDENNE, L. E.. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. **Frontiers in Pharmacology**, v. 9, p. 1–18, set. 2018. DOI: 10.3389/fphar.2018.01089. Disponível em: <https://doi.org/10.3389/fphar.2018.01089>. Acesso em: 07 mar. 2024.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Biomass e sistema costeiro-marinho do Brasil**: compatível com a escala 1:250 000. Rio de Janeiro: IBGE, 2019.

KAUSHIK, A. C.; MEHMOOD, A.; DAI, X.; WEI, D. Q. (2020). A comparative chemogenic analysis for predicting Drug-Target Pair via Machine Learning Approaches. **Scientific Reports**, [s.l.], v. 10, 6870, p. 1–11, abr. 2020. DOI: 10.1038/s41598-020-63842-7. Disponível em: <https://doi.org/10.1038/s41598-020-63842-7>. Acesso em: 07 mar. 2024.

MA, L. *et al.* Database Commons: A Catalog of Worldwide Biologic Database. **Genomics Proteomics Bioinformatics**, [s.l.], v. 21, n. 5, p. 1054-1058, out. 2023. DOI: 10.1016/j.gpb.2022.12.004. Disponível em: <https://doi.org/10.1016/j.gpb.2022.12.004>. Acesso em: 12 abr. 2024.

MAIA, C. L. B.; FURTADO, E. S. Uma revisão sistemática sobre medição da experiência do usuário. **Foz do Iguaçu: HC '14: Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems**, out. 2014. p. 369-372. DOI: 10.5555/2738055.2738122. Disponível em: <https://dl.acm.org/doi/10.5555/2738055.2738122>. Acesso em: 21 ago. 2024.

MAIA, J. G. S.; ANDRADE, E. H. A. Database of the Amazon aromatic plants and their essential oils. **Química Nova**, [s.l.], v. 32, n. 3, p. 595–622, mar. 2009. DOI: 10.1590/S0100-40422009000300006. Disponível em: <https://doi.org/10.1590/S0100-40422009000300006>. Acesso em: 07 mar. 2024.

MATTHEWS, H.; HANISON, J., NIRMALAN, N. “Omics”- Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. **Proteomes**, [s.l.], v. 4, n. 3, p. 1-12, 28, set. 2016. DOI: 10.3390/proteomes4030028. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28248238/>. Acesso em: 16 jan. 2025.

MARQUES, L. G. A.; PESSOA, C.; SANTOS, M. R. M. C. O valor econômico da biodiversidade e os países megadiversos. *In*: SILVA, G. F.; RUSSO, S. L. (Org.). **Capacite: Os caminhos para a inovação tecnológica**. São Cristóvão: Editora UFS, 2014. p. 41-58.

MARTINS, S. E.; BIANCHINI, A. Toxicity tests aiming to protect Brazilian aquatic systems: current status and implications for management. **J. Environ. Monit.**, [s.l.], v. 13, n. 7, p. 1866-1875, jul. 2011. DOI: 10.1039/c0em00787k. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21643562/>. Acesso em: 13 abr. 2024.

MEJIAS, R. G. Bioeconomia e suas aplicações. **R. ÎANDÉ - Ciências e Humanidades**, São Bernardo do Campo, v. 2, n. 3, p. 105-121, jul. 2019. DOI: 10.36942/iande.v2i3.87. Disponível em: <https://doi.org/10.36942/iande.v2i3.87>. Acesso em: 10 abr. 2024.

MELLO, A. F.. Dilemas e desafios do desenvolvimento da Amazônia: o caso brasileiro. **Revista Crítica de Ciências Sociais**, Coimbra, n. 107, p. 91-108, set. 2015. DOI: 10.4000/rccs.6025. Disponível em: <https://journals.openedition.org/rccs/6025>. Acesso em: 13 abr. 2024.

NEGRI, F.; SQUEFF, F. H. S. O mapeamento da infraestrutura científica e tecnológica no Brasil. *In*: NEGRI, F.; SQUEFF, F. H. S. (Org.). **Sistemas setoriais de inovação e infraestrutura de pesquisa no Brasil**. Brasília: IPEA, 2016. cap. 1. p. 15-62.

OTCA - Organização do Tratado de Cooperação Amazônica. **Avaliação rápida da biodiversidade e dos serviços exossistêmicos da região amazônica**. [s.l.]: OTCA, 2021. 52p.

PAUL, S. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. **Nat. Rev. Drug Discov.**, [s.l.], v. 9, p. 203-214, fev./mar. 2010. DOI: 10.1038/nrd3078. Disponível em: <https://doi.org/10.1038/nrd3078>. Acesso em: 16 jan. 2025.

PILON, A. C. *et al.* NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. **Sci. Rep.**, [s.l.], v. 7, n. 1, 7215, ago. 2017. DOI: 10.1038/s41598-017-07451-x. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28775335/>. Acesso em: 13 abr. 2024.

PIMENTA, C.; AZEVEDO, A. Por uma bioeconomia inclusiva e que mantenha de pé a floresta. **Interesse Nacional**, São Paulo, v. 13, n. 1, p. 31-37, ago. 2020. Disponível em: https://www.escolhas.org/wp-content/uploads/2020/07/Revista-Interesse-Nacional_BIOECONOMIA_Irice-e-Escolhas_agosto2020.pdf. Acesso em: 10 abr. 2024.

PLOTKIN, M. J. **The Amazon**: what everyone needs to know. New York: Oxford University Press, 2020. 248p.

PRESSMAN, ROGER S; MAXIM, B. R. (2016). **Engenharia de Software - Uma Abordagem Profissional**. 8. ed. [s.l.]: AMGH. 968p.

OGBE, R. J. D. BIOINFORMATICS ADVANCES IN GENOMICS – A REVIEW. **International Journal of Current Research and Review**, [s.l.], v. 8, n. 10, p. 5-11, maio 2016. Disponível em: https://www.researchgate.net/publication/304648097_BIOINFORMATICS_ADVANCES_IN_GENOMICS_-_A_REVIEW. Acesso em: 07 mar. 2024.

QUEIROZ, L. F. P.; FLORES, M. S. A.; SOBRINHO, M. V. A bioeconomia e sua relação com a Amazônia paraense: uma revisão a partir do conceito de desenvolvimento. **Editora Científica**, [s.l.], v. 1, p. 11-24, jan. 2023. DOI: 10.37885/221110922. Disponível em: <https://www.editoracientifica.com.br/articles/code/221110922>. Acesso em: 13 abr. 2024.

RAMÍREZ-ACOSTA, K. Interfaz y experiencia de usuario: parámetros importantes para un diseño efectivo. **Tecnología en Marcha**, Cartago, v. 30, suppl. 1, p. 49-54, out.;nov. 2017. DOI: 10.18845/tm.v30i5.3223. Disponível em: <http://dx.doi.org/10.18845/tm.v30i5.3223>. Acesso em: 21 ago. 2024.

REFLORA. **Programa REFLORA**, 2024. Disponível em: <https://floradobrasil.jbrj.gov.br/reflora/PrincipalUC/PrincipalUC.do;jsessionid=338A800C5FF588B8C9739395B540A97E>. Acesso em: 10 jun. 2024.

REIS, R. M. C.; PEREIRA, N. P.; RABÊLO, O. S. Estudo Prospectivo sobre o Potencial Uso do Cacau no Setor de Cosméticos: análise das tendências atuais para PD&I. **Cadernos de Prospecção**, Salvador, v. 17, n. 2, p. 639–653, 2024. DOI: 10.9771/cp.v17i2.56025. Disponível em: <https://periodicos.ufba.br/index.php/nit/article/view/56025>. Acesso em: 21 ago. 2024.

ROA, F.; TELLES, M. P. C. The Cerrado (Brazil) plant cytogenetics database. **Comp. Cytogenet.**, [s.l.], v. 11, n. 2, p. 285-297, abr. 2017. DOI: 10.3897/CompCytogen.11(2).11395. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28919965/>. Acesso em: 13 abr. 2024.

SAMPAIO NETO, O. Z.; BATISTA, E. A. C.; MEIRELLES, A. J. A. Potencial de oleaginosas nativas do desenvolvimento de cadeias produtivas da biodiversidade brasileira. **Desenvolvimento e Meio Ambiente**, [s.l.], v. 54, p. 537-559, jul./dez. 2020. DOI: 10.5380/dma.v54i0.71934. Disponível em: <http://dx.doi.org/10.5380/dma.v54i0.71934>. Acesso em: 24 jan. 2024

SOROKINA, M.; STEINBECK, C. Review on natural products databases: where to find data in 2020. **Journal of Cheminformatics**, [s.l.], v. 12, n. 1, p. 1-51, abr. 2020. DOI: 10.1186/s13321-020-00424-9. Disponível em: <https://doi.org/10.1186/s13321-020-00424-9>. Acesso em: 07 mar. 2024.

SILVA, M. F. O.; PEREIRA, F. S.; MARTINS, J. V. B. A bioeconomia em números. **BNDES Setorial**, Rio de Janeiro, n. 47, p. 277-331, mar. 2018. Disponível em: <http://web.bndes.gov.br/bib/jspui/handle/1408/15383>. Acesso em: 12 abr. 2024.

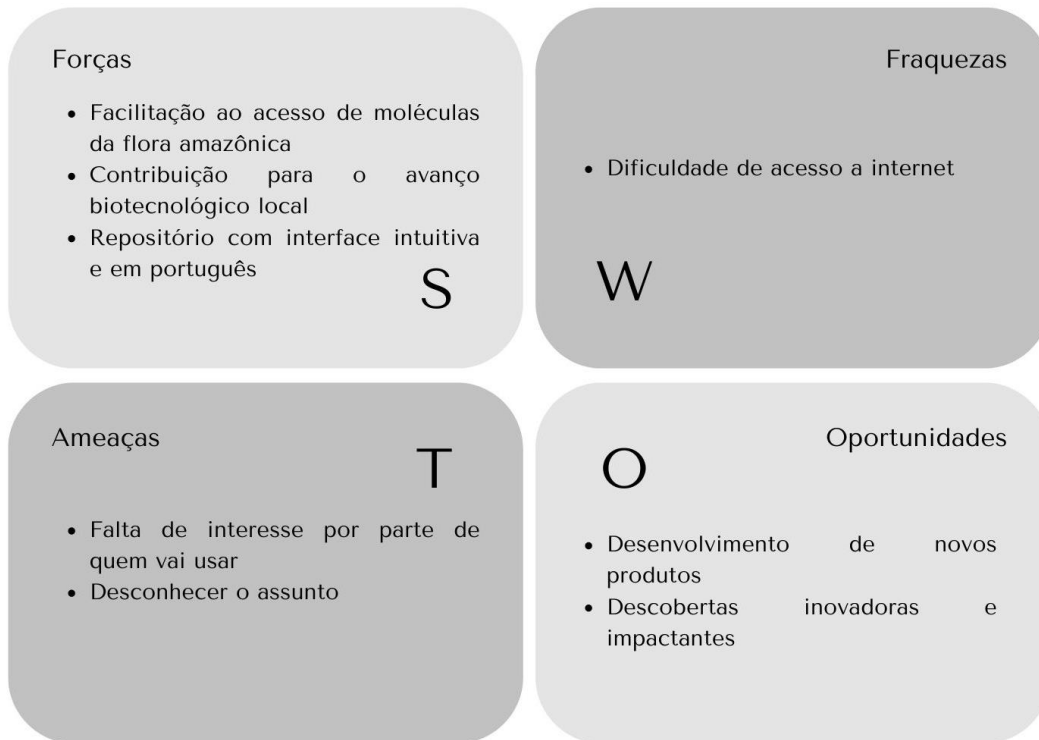
XIE, T.; SONG, S.; LI, S.; OUYANG, L.; XIA, L.; HUANG, J. (2015). Review of natural product databases. **Cell Proliferation**, [s.l.], v. 48, n. 4, p. 398–404, ago. 2015. DOI: doi.org/10.1111/cpr.12190. Disponível em: <https://doi.org/10.1111/cpr.12190>. Acesso em: 07 mar. 2024.

WILLERDING, A. L.; SILVA, L. R.; SILVA, R. P.; ASSIS, G. M. O.; PAULA, E. V. C. M. Estratégias para o desenvolvimento da bioeconomia no estado do Amazonas. **Estudos avançados**, [s.l.], v. 34, n. 98, p. 144-165, jan./abr. 2020. DOI: 10.1590/s0103-4014.2020.3498.010. Disponível em: <https://doi.org/10.1590/s0103-4014.2020.3498.010>. Acesso em: 12 abr. 2024.

WILSON, E. O.; PETER, F. M. (ed.). **Biodiversity**. Washington, D.C.: National Academy of Sciences, 1988. 538p.

YANG, Z.; CAI, S.; ZHOU, Z.; ZHOU, NAN. Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. **Information & Management**, [s.l.], v. 42, p. 575-589, jun. 2004. DOI: 10.1016/j.im.2004.03.001. Disponível em: <https://www.researchgate.net/publication/222513969>. Acesso em: 20 jun. 2024.

APÊNDICE A - Matriz FOFA (SWOT)



Fonte: Elaborado pela autora (2025).

APÊNDICE B - Modelo de Negócio CANVAS

Parceiros Chave	Atividades Chave	Propostas de Valor	Relacionamento	Segmentos de Clientes		
FAPERO	<ul style="list-style-type: none"> -Coleta e integração de dados -Desenvolvimento da plataforma 	<ul style="list-style-type: none"> -Centralização de dados biológicos da flora amazônica -Plataforma on-line -Facilitação de pesquisa Inovação biotecnológica 	<ul style="list-style-type: none"> -Suporte -E-mail 	<ul style="list-style-type: none"> Pesquisadores Estudantes Empresas 		
	<th>Recursos Chaves</th> <td></td> <td> <th>Canais</th> <td></td> </td>	Recursos Chaves		<th>Canais</th> <td></td>	Canais	
	<ul style="list-style-type: none"> -Plataforma Tecnológica -Banco de Dados Biológicos -Equipe multidisciplinar 		<ul style="list-style-type: none"> -Plataforma on-line -Conferências -Publicações científicas 			
<th>Estrutura de Custos</th> <td colspan="2"> <th>Fluxos de Receita</th> </td>			Estrutura de Custos	<th>Fluxos de Receita</th>		Fluxos de Receita
<ul style="list-style-type: none"> -Desenvolvimento e manutenção da plataforma -Equipe -Aquisição e Processamento de dados 			Edital 53/2023			

Fonte: Elaborado pela autora (2025).

APÊNDICE C - Artigo publicado

Cadernos de Prospecção

DOI: <https://doi.org/10.9771/cp.v17i2.56536>

Prospecção Tecnológica: potencial terapêutico de moléculas presentes no veneno de serpentes do gênero *Bothrops* sp., com ênfase na espécie *Bothrops jararaca*

Technological Prospection: therapeutic potential of molecules present in the venom of snakes of the genre Bothrops, especially Bothrops jararaca

Carolina Barros da Costa¹

Kaio Alexandre da Silva¹

Marcio Rodrigues Miranda¹

¹Instituto Federal de Educação, Ciência e Tecnologia de Rondônia, Porto Velho, RO, Brasil

Resumo

O objetivo do presente estudo foi realizar um levantamento tecnológico de documentos de patentes relacionadas às serpentes do gênero *Bothrops* sp., analisando os documentos relacionados para fins medicinais. Para a execução deste levantamento, foi utilizada a base de patentes Orbit Intelligence, buscando o termo “*Bothrops*” e direcionando a busca para fins medicinais usando a classificação de patentes A61K, sem delimitação temporal ou espacial. Foram encontradas 67 famílias de patentes, as quais possuem Brasil, China e México como os países com as maiores quantidades de depósitos de documentos de patentes. As principais instituições depositantes dos pedidos de patente foram a empresa suíça Pentapharm, a empresa brasileira Biolab Sanus Farmacêutica e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Os principais países depositantes foram o Brasil, a China e o México. Foi possível notar um predomínio de patentes mortas, arquivadas por motivos diversos, em relação às vivas (depositadas e/ou concedidas), indicando que se trata de uma área tecnológica com baixa dinâmica de inovação.

Palavras-chave: Bioeconomia; Biomoléculas; P&D&I.

Abstract

The objective of the present study was to carry out a technological survey of patent documents related to snakes of the genus *Bothrops* sp., analyzing documents related to medicinal purposes. To carry out this survey, the Orbit Intelligence patent database was used, searching for the term “*Bothrops*” and directing the search to medicinal purposes using the A61K patent classification, without temporal or spatial delimitation. 67 patent families were found, with Brazil, China and Mexico as the countries with the largest amounts of patent document deposits. The main institutions filing patent applications were the Swiss company Pentapharm, the Brazilian company Biolab Sanus Farmacêutica and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). The main depositing countries were Brazil, China and Mexico. It was possible to notice a predominance of filed patents in relation to current ones, indicating that this is a technological area with low innovation dynamics.

Keywords: Bioeconomy; Biomolecules; R&D&I.

Área Tecnológica: Farmacêutica. Tecnologia Médica. Biotecnologia.



Direito autorial e licença de uso: Este artigo está licenciado sob uma Licença Creative Commons. Com esta licença você pode compartilhar, adaptar, para qualquer fim, desde que atribua a autoria da obra, forneça um link para a licença, e indicar se foram feitas alterações.

Recebido: 18/09/2023
Aceito: 11/01/2024

APÊNDICE D - Artigo publicado

Revista de Administração de Roraima - RARR



Revista de Administração de Roraima, v. 15, 2024
 CADECON - Universidade Federal de Roraima, Boa Vista, RR
<https://revista.ufr.br/adminrr/>
 DOI: 10.18227/2237-8057rarr.v15i0.8088

Artigo
 Original

ISSN 2237-8057

PROSPECÇÃO TECNOLÓGICA DO POTENCIAL TERAPÊUTICO DE MOLÉCULAS EXTRAÍDAS DE *ACMELLA OLERACEA* (L.) R.K. JANSEN (*SPILANTHES OLERACEA*)

TECHNOLOGICAL PROSPECTION OF THE THERAPEUTIC POTENTIAL OF
 MOLECULES EXTRACTED FROM *Acmella oleracea* (L.) R.K. Jansen (*Spilanthes
 oleracea*).

Carolina Barros da Costa¹

¹ Discente do PROFNIT do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia
 Email: carolinabc962@gmail.com

Kaio Alexandre da Silva²

Docente do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia
 Email: kaio.silva@ifro.edu.br

Marcio Rodrigues Miranda³

Docente do PROFNIT do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia
 Email: marcio.miranda@ifro.edu.br

RESUMO:

O objetivo deste estudo é realizar uma prospecção tecnológica de documentos de patentes relacionados à utilização de compostos extraídos da *Acmella oleracea* (L.) R.K. Jansen para fins medicinais. A prospecção tecnológica foi realizada por meio da base de dados Orbit Intelligence, utilizando as palavras-chave “*acmella oleracea*” e “*spilanthes oleracea*”, e a classificação de patentes A61K, sem delimitação temporal ou espacial, com o intuito de identificar e compreender as inovações que estão impulsionando esse setor. Foram encontradas ao todo 114 famílias de patentes, as quais apresentaram uma tendência crescente de investimento até o ano de 2022, com queda em 2023. No que se refere às patentes vigentes, Estados Unidos, China e Alemanha se destacaram como principais depositantes, enquanto o Brasil surge em quarto lugar. A forte presença dos Estados Unidos é reforçada com a presença das empresas Nulixir, BodyBio e Mary Kay entre as cinco principais empresas depositantes. No Brasil, as IES se destacam como as principais depositantes.

Palavras chaves: Biomoléculas. Jambu. Biodiversidade Brasileira.

ABSTRACT

The objective of this study is to carry out a technological search for patent documents related to the use of compounds extracted from *Acmella oleracea* (L.) R.K. Jansen for medicinal purposes. Technological prospecting was carried out using the Orbit Intelligence database, using the keywords “*acmella oleracea*” and “*spilanthes oleracea*”, and the A61K patent classification, without temporal or spatial delimitation, with the aim of identifying and understanding the innovations that are driving this sector. A total of 114 patent families were found, which showed a growing investment trend until 2022, with a drop in 2023. With regard to current patents, the United States, China and Germany stand

ANEXO A - Comprovante de publicação/submissão de artigo

Título do Artigo publicado: Prospecção Tecnológica: potencial terapêutico de moléculas presentes no veneno de serpentes do gênero *Bothrops* sp. , com ênfase na espécie *Bothrops jararaca*

Revista: Cadernos de Prospecção

Publicado em 01-04-2024

Cadernos de Prospecção: <https://periodicos.ufba.br/index.php/nit/article/view/56536>

DOI: <https://doi.org/10.9771/cp.v17i2.56536>

Título do Artigo publicado: Prospecção tecnológica do potencial terapêutico de moléculas extraídas de *Acmella oleracea* (L.) R.K. Jansen (*Spilanthes oleracea*)

Revista: Revista de Administração de Roraima - RARR

Publicado em 06-03-2024

Revista de Administração de Roraima - RARR:

<https://revista.ufrr.br/adminrr/article/view/8088>

DOI: <https://doi.org/10.18227/2237-8057rarr.v15i1.8088>

Título do Artigo Publicado: Moléculas da Amazônia: Construção do protótipo de um repositório on-line das espécies e biomoléculas da flora amazônica.

Evento: 11ª Conferência Ibero Americana Computação Aplicada (CIACA 2024).

Livro de Atas da CIACA 2024

ISBN: 978-989-8704-64-1

MOLÉCULAS DA AMAZÔNIA: CONSTRUÇÃO DE UM REPOSITÓRIO ON-LINE DAS ESPÉCIES E BIOMOLÉCULAS DA FLORA AMAZÔNICA

Carolina Barros da Costa, Márcio Rodrigues Miranda e Kaio Alexandre da Silva
Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO
IFRO, Câmpus Porto Velho Calama, Porto Velho/RO, Brasil

RESUMO

A Amazônia possui uma biodiversidade abundante, cuja flora abriga uma variedade de moléculas e compostos de alto valor agregado que podem ser explorados por diversos setores, como o farmacêutico e o biotecnológico. No entanto, há uma fragmentação de dados referentes as espécies da flora amazônica e suas biomoléculas, tornando o pesquisador dependente de acessar múltiplos bancos de dados para a obtenção das informações desejadas. Logo, o objetivo deste estudo é apresentar o desenvolvimento de um repositório on-line que disponibiliza informações sobre as espécies amazônicas e suas respectivas biomoléculas. Para isso, foi utilizada a metodologia de desenvolvimento por prototipação associada a técnicas de raspagem de dados (web scraping) e ETL (Extract, Transform and Load). Com isso, se pode correlacionar bancos de dados que continham não só informações sobre biodiversidade da Amazônia Legal Brasileira, como também de seus compostos químicos. Isso resultou em um protótipo de um repositório que conta com filtros de buscas personalizados capazes de facilitar a localização de informações específicas e permitir uma exploração mais direcionada e eficiente para os recursos biológicos da flora Amazônica.

PALAVRAS-CHAVE

Repositório On-Line, Biomoléculas, Flora, Amazônia, Protótipo, Banco de Dados

1. INTRODUÇÃO

O vasto território amazônico abriga uma biodiversidade abundante composta por uma variedade de espécies de plantas, animais e insetos [Plotkin, 2020]. Essa diversidade biológica fornece uma gama de recursos naturais, os quais, por meio da bioinovação, podem ser transformados em uma variedade de moléculas e compostos de alto valor agregado [Barbosa 2020].

As moléculas e composto extraídos da biodiversidade podem ser explorados por diversos setores. Calazans *et al.* [2021] afirma que as sementes florestais possuem um potencial tecnológico significativo. Os autores afirmam que estas podem ser exploradas pelos setores alimentício, cosmetológico e farmacêutico. Ademais, Costa *et al.* [2024] realizou uma prospeção tecnológica de patentes que revelou uma ampla exploração da atividade biológica presente no Jambu (*Acmella oleracea*), espécie muito presente na cultura amazônica. Estes afirmam que essa espécie pode promover atividades antioxidante, analgésica e até mesmo anti-inflamatória, os quais são explorados principalmente pela indústria farmacêutica e cosmetológica. Desse modo, a exploração desse potencial pode não só viabilizar avanços na industrialização, como também promover o desenvolvimento socioeconômico local [Queiroz *et al.* 2023]. Diante disso, a bioinformática pode desempenhar um papel

Título do Resumo apresentado: Molecules of the Amazon: Integration and Centralization of Data on the Amazon Flora and its Biomolecules

Evento: 17th Brazilian Symposium on Bioinformatics (BSB2024)



Certificate of Poster Presentation

This is to certify that

Carolina Barros da Costa, Gustavo Casagrande Borges,
Marcos Reis Dutra, Márcio Rodrigues Miranda, Kaio
Alexandre da Silva

are the authors of the poster entitled:

**Molecules of the Amazon: Integration and Centralization of Data on the Amazon
Flora and its Biomolecules**

which was presented during the **XVII Brazilian Symposium on Bioinformatics
BSB2024**, held from December 2 to 4, 2024, in **Vitória, Espírito Santo, Brazil**.



Marcelo da Silva Reis
Prof. Marcelo da Silva Reis
Program Chair

Sérgio Nery Simões
Prof. Sérgio Nery Simões
General Chair



Organized by the Special Interest Group in Computational Biology (CE-BioComp) of the Brazilian Computer Society (SBC), the symposium brought together researchers and professionals from around the world to discuss the latest advances in Bioinformatics and Computational Biology.

Título do Artigo submetido: Tendências científicas e tecnológicas da *Carapa guianensis*, popularmente conhecida como andiroba.

The screenshot shows a web interface for managing submissions. The main content area is titled 'Minhas Submissões Designadas' and contains a table with one entry:

ID	Author	Title	Status	Actions
61103	Costa et al.	TENDÊNCIAS CIENTÍFICAS E TECNOLÓGICAS DA Carapa guianensis, POPULARMENTE CONHECIDA COMO ANDIROBA	Avaliação	1/4

This is a zoomed-in view of the submission management interface, showing the same table as above:

ID	Author	Title	Status	Actions
61103	Costa et al.	TENDÊNCIAS CIENTÍFICAS E TECNOLÓGICAS DA Carapa guianensis, POPULARMENTE CONHECIDA COMO ANDIROBA	Avaliação	1/4

61103 / Costa et al. / TENDÊNCIAS CIENTÍFICAS E TECNOLÓGICAS DA *Carapa guianensis*, POPULARMENTE CONHECIDA COMO ANDIRO

Biblioteca da Submissão

Fluxo de Trabalho Publicação

Submissão Avaliação Edição de Texto Editoração


Rodada 1







Situação da rodada 1
Uma avaliação está atrasada.

Discussão da avaliação [Adicionar comentários](#)


Nome	De	Última resposta	Respostas	Fechado
Nenhum item				

Título do Artigo submetido: Tendências científicas e tecnológicas da *Carapa guianensis*, popularmente conhecida como andiroba.

 Gmail Q jems ✕ ≡

←      

Your BSB 2024 paper 246335 Caixa de entrada ✕

 **JEMS** <jems@sbc.org.br>
para mim, msreis, Kaio, Márcio, Marcos, Gustavo ▾

Dear Ms. Carolina Costa:

Congratulations - your work "Molecules of the Amazon: Integration and Centralization of Data on the Amazon Flora and its Biomolecules" has been accepted for the Poster Abstract session at BSB 2024.

Please address the reviewers suggestions for the preparation of the final file for publication. Remember to submit the revised file with all corrections before the 24th October.

The reviews are below or can be found at
<https://jems.sbc.org.br/PaperShow.cgi?m=246335>

The format to be used is SBC-OpenLib (SOL) in its Bioinformatics series, from SBC. Authors should consult SOL author guidelines and use their proceedings templates, either for LaTeX or for Word, for the preparation of their papers. The actual submission must be in PDF format (maximum 20 Mb). LaTeX or Word versions (on which the PDF must be based) will be required later, for accepted papers, for inclusion in the SBC-OpenLib proceedings. BSB encourages authors to include their ORCIDs in their papers.

Author guidelines: <https://bsb.sbc.org.br/2024/call-for-papers/>

Título do Artigo submetido: Promoção da bioeconomia em território amazônico: utilização de técnicas de Web Scraping e ETL para mapeamento das moléculas da flora amazônica.

[COTB16] Registro da submissão nº5640 (Promoção da bioeconomia em território amazônico: utilização de técnicas de Web Scraping e ETL para mapeamento das moléculas da flora amazônica) Caixa de entrada ✕

jems3@sbc.org.br

para mim, kaio.silva, borgesgustavo360, reis17265, marcio.miranda, cottb.tpc ▾

dom., 15 de dez. de 2024, 14:54 ☆ ⓘ

Prezado(a) Carolina Barros da Costa,

Obrigado por registrar sua submissão #5640 intitulada "Promoção da bioeconomia em território amazônico: utilização de técnicas de Web Scraping e ETL para mapeamento das moléculas da flora amazônica" no **Computer on the Beach** 2025.

Você pode modificar sua submissão em:

<https://jems3.sbc.org.br/submissions/5640/>

Você visualizar todas as suas submissões em:

<https://jems3.sbc.org.br/submissions/>

Cordialmente,

Coordenação do Comitê de Programa

Computer on the Beach 2025

cottb.tpc@gmail.com