

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 prospecção 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 fundamental na identificação desse potencial.

A bioinformática é uma área interdisciplinar responsável por combinar biologia, informática e estatística de modo a desenvolver métodos para armazenar, recuperar, organizar e armazenar dados biológicos (Obge, 2016). Além disso, esta pode ser dividida em dois grupos: um associado à sequência e o outro à estrutura biomolecular. Essa integração desses dados não só resulta na facilitação de estudos de sequência genéticas, estruturas moleculares, interações proteicas, dentre outras possibilidades, como também na melhor compreensão dos processos biológicos complexos [Diniz e Canduri 2017].

Os avanços proporcionados pela bioinformática têm transformado de forma significativa a biologia experimental. Dentre as ferramentas que contribuem para esses avanços, pode-se mencionar o docking molecular e a dinâmica molecular. Ao utilizar o docking molecular, é possível simular o encaixe de um ligante em uma estrutura alvo, permitindo prever a interação entre eles e identificar potenciais moléculas de interesse e suas aplicações [Fan *et al.* 2019]. Já a dinâmica molecular proporciona uma análise mais detalhada a respeito dessas interações a partir da simulação do comportamento de moléculas ao longo do tempo. Ademais, também leva em consideração fatores externos, como temperatura e pressão [Namba *et al.* 2008]. Essas técnicas reduzem a necessidade de testes físicos extensivos, diminuindo o uso de materiais e economizando tempo e custos associados aos experimentos tradicionais.

No entanto, não há um banco de dados que integre e centralize dados sobre as espécies da flora amazônica e suas moléculas/compostos bioativos. Apesar de iniciativas como o ReFlora, responsável por disponibilizar informações sobre a flora e fungo do Brasil [ReFlora, 2024], e o NuBBEDE, responsável por disponibilizar dados químicos e biológicos da biodiversidade brasileira [Pilon, 2017], as informações se apresentam dispersas em diferentes bancos de dados. Isso torna o pesquisador dependente de acessar múltiplos bancos de dados para a obtenção das informações desejadas, gerando maior gasto de tempo, de recursos naturais com testes experimentais, e de custo. Logo, o objetivo deste estudo é apresentar o desenvolvimento de um repositório on-line que centralize e integre dados a respeito das espécies e biomoléculas da flora amazônica. A seguir, a seção 2 apresenta a metodologia utilizada para o desenvolvimento do protótipo, a seção 3 apresenta e discute os resultados obtidos, seguida da seção 4 que finaliza apresentando a conclusão.

2. METODOLOGIA

O desenvolvimento do sistema foi de acordo com a metodologia de desenvolvimento por prototipação. Segundo Pressman e Maxim [2016], trata-se de um processo que tem como intuito identificar e definir os requisitos necessários para a construção de um software, o processo deve ser executado de forma iterativa, obedecendo as etapas: Coleta de requisitos, Projeto Rápido, Construção do Protótipo, Avaliação do Protótipo, Refinamento do Protótipo e Construção do produto.

No processo de coleta de requisitos, foram definidas as ferramentas a serem utilizadas para a coleta de dados, bem como as fontes de entrada de dados que iriam alimentar o repositório Moléculas da Amazônia. Diante disso, os dados que compõem o Moléculas da Amazônia foram obtidos por meio da técnica *Web scraping*, ou raspagem de dados. O intuito da raspagem de dados é criar rotinas responsáveis por automatizar a coleta de dados disponíveis da web, bem como salvar esses dados para recuperação e/ou análise posterior [Zhao, 2017]. Desse modo, essas técnicas permitiram a extração de forma automatizada e a estruturada de dados de fontes que alimentam o repositório Moléculas da Amazônia.

Os bancos de dados escolhidos para alimentar esse repositório foram o GBIF, para a obtenção de dados sobre as espécies da Amazônia, e o NuBBEDB, NCBI e ChEMBL, para a obtenção de dados sobre as biomoléculas que podem ser encontradas nessas espécies. De modo a garantir a integração dos dados, foi utilizada a ferramenta de ETL (Extract, Transform and Load) através de scripts desenvolvidos em Python. Estes organizaram os dados obtidos dos diferentes bancos de dados e os converteram em formatos compatíveis com o modelo do banco de dados.

Logo, os dados extraídos foram carregados no banco por meio da ferramenta ETL. A padronização dos nomes e atributos foi feita com o intuito de garantir que as informações fossem carregadas de forma correta no banco e que dados taxonômicos coletados do GBIF fossem corretamente integrados com os dados moleculares do NuBBEDB, ChEMBL e sequências genéticas do NCBI, por exemplo.

Ademais, 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. Ao finalizar essa primeira etapa, foi feito o uso do 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 serviu de base para a construção do protótipo do repositório on-line Moléculas da Amazônia.

3. RESULTADOS

O protótipo construído possui filtros que te permite realizar buscas por moléculas, espécies, localização específica, referências e informações químicas (Figura 1). 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 à 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. Por fim, também é possível realizar buscar por meio de referências bibliográficas e patentárias associadas às espécies e/ou moléculas presentes no repositório.

Figura 1. Busca por filtros e Cards intuitivos

Ao efetuar a busca, as informações pesquisadas são apresentadas através de cards intuitivos. A proposta destes cards é apresentar as informações de forma organizada e permitir que o usuário visualize os principais dados associados à sua busca de forma rápida, clara e objetiva (Figura 2).

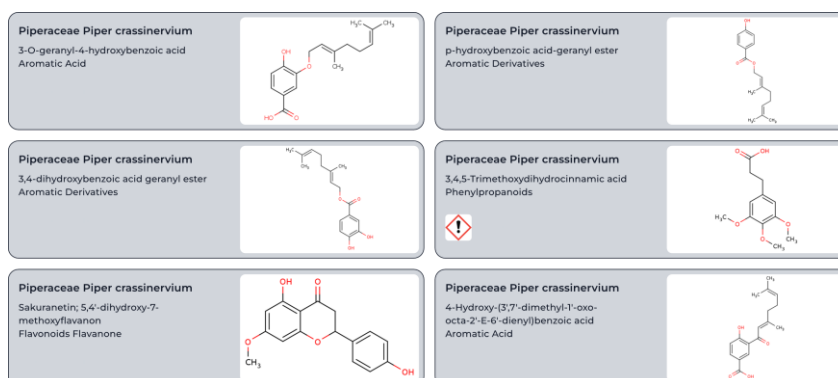


Figura 2. Cards intuitivos

Diferente dos bancos de dados biológicos existentes, o Moléculas da Amazônia permite visualizar tanto os dados referentes às espécies (Figura 3), quanto das moléculas (Figura 4) 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. Desse modo, o usuário poderá ter acesso a essas informações de forma mais fácil e rápida, contribuindo para uma melhor compreensão dos dados pesquisados.

Além disso, o usuário pode acessar o seu histórico de pesquisa, de modo a permitir o acompanhamento e a organização das buscas já efetuadas. Isso facilita o desenvolvimento de análises ao longo do tempo, uma vez que o usuário tem a possibilidade de retomar as pesquisas anteriores de maneira rápida e eficiente.

The screenshot shows the 'Moléculas da Amazônia' website interface. The main content area is titled 'Piperaceae Piper crassinervium'. Under 'Informações Gerais', it lists: 'Forma de vida: Arbusto', 'Habitat: Terrestre', 'Origem: Nativa', and 'Endêmica: Não endêmica'. Below this is a 'Galeria de imagens' section with three photos of the plant. A world map shows the geographical distribution of the species with yellow and orange markers.

Figura 3. Informações sobre as espécies

The screenshot shows the 'Moléculas da Amazônia' website interface for molecular data. The main content area is titled 'Piperaceae Piper crassinervium'. Under 'Informações Gerais', it lists: 'Nome comum: 3-(4-dihydroxybenzoato) ácido geranyl éster', 'Nome IUPAC: (2E)-3,7-dimethylocta-2,6-dien-1-yl 3,4-dihydroxybenzoato', 'Fórmula molecular: C17H22O4', 'InChI: 1S(C17H22O4)1-12(2E)-4-6-13(2E)-10-21-17(20)14-7-8-15(18)16(19)11-14-h,3-9,11,8-19H,4,6,10-12,13-h3,13-9+', 'InChIKey: OVDOPULMGGUUFM-UKHTLTKRSA-N', 'Smile: O=C(O)C=C(C)C(C)C=C(C)C(C)C=C(C)C(O)C(O)C=C1', 'Espécie: Piperaceae Piper crassinervium', 'Propriedades biológicas: Inibição de protease; Antibal', 'NCBI Compound ID: 1358243', and 'ChEMBL ID: ChEMBL340845'. Below this is a chemical structure diagram and a table of related files.

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

Figura 4. Informações sobre as moléculas

4. CONCLUSÃO

A construção do protótipo do repositório on-line Moléculas da Amazônia resultou em uma plataforma que integra e centraliza dados de diversas fontes, tais como o NuBBEDB, GBIF, ChEMBL e NCBI. Este protótipo contribui de forma significativa para os conceitos relacionados a bancos de dados, visto que apresenta a relevância da modelagem e integração de dados de múltiplas fontes em um único ambiente. Somado a isso, este contribui para a gestão e distribuição de conhecimento, pois facilita o acesso e a utilização de informações sobre a flora amazônica e suas biomoléculas. Logo, ao integrar esses dados em um único ambiente, este repositório facilita o acesso a dados completos sobre a flora amazônica e seus compostos bioativos, resultando na aceleração do processo de pesquisa e suprimindo a necessidade de avanço científico e tecnológico relacionados à biodiversidade amazônica.

REFERÊNCIAS

- Aragón, L. E., (2018). A dimensão internacional da amazônia: Um aporte para sua interpretação. *Revista NERA*, Vol. 42, No. 21, pp. 14-33.
- Barbosa, R., (2020). A bioeconomia e a Amazônia. *Interesse Nacional*, Vol. 13, No. 1, pp. 31-37.
- Calazans, C. C. *et al.*, (2021). Sementes Florestais e seu Potencial Tecnológico: uma análise de metadados. *Cadernos de Prospecção*, Vol. 14, No. 3, pp. 794-809.
- Costa, C. B. *et al.*, (2021). Prospecção Tecnológica do potencial terapêutico de moléculas extraídas de *Acmella oleracea* (L.) R.K. Jansen (*Spilanthes oleracea*). *Revista de Administração de Roraima - RARR*, Vol. 15, No. 1, pp. 1-17.
- Diniz, W. J. S. & Canduri, F., (2017). Bioinformatics: An overview and its applications. *Genetics and Molecular Research*, Vol. 16, No. 1, pp. 01-21.
- Fan, J. *et al.*, (2019). Progress in molecular docking. *Quantative Biology*, Vol. 7, pp. 83-89.
- Namba, A. M. *et al.*, (2008). Dinâmica molecular: teoria e aplicações em planejamento de fármacos. *Eclética Química*, Vol. 33, No. 4, pp. 13-23.
- Obge, R. J. D., (2016). Bioinformatics advances in genomics – A review. *International Journal of Current Research and Review*, Vol. 8, No. 10, pp. 5-11.
- Plotkin, M. J., (2020). *The Amazon: what everyone needs to know*. Oxford University Press, New York, USA.
- Pilon, A. *et al.*, (2017). NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.*, Vol. 7, 7215, pp. 1-12.
- Pressman, R. S. & Maxim, B. R., (2016). Engenharia de Software - Uma Abordagem Profissional. 8. AMGH.
- Queiroz, L. F. P. *et al.*, (2023). A bioeconomia e sua relação com a Amazônia paraense: uma revisão a partir do conceito de desenvolvimento. *Editores Científicos*, Vol. 1, pp. 11-24.
- Reflora, (2024). *Programa Reflora*.
- Zhao, B., (2017). Web Scraping. *Springer*, pp. 1-3.