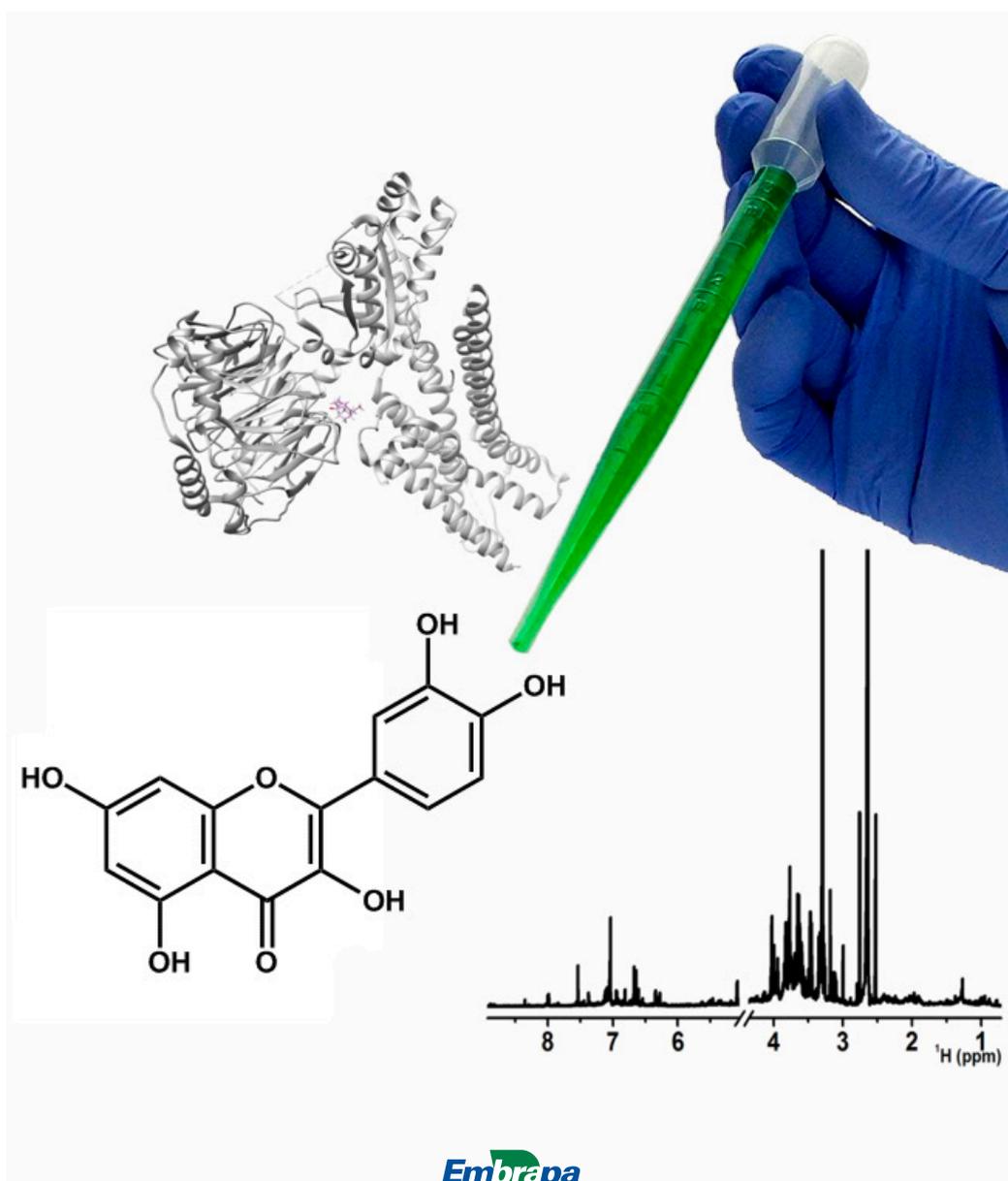


Fortaleza, CE / Abril, 2025

Base de dados de compostos naturais da biodiversidade brasileira do Laboratório Multiusuário de Química de Produtos Naturais

OBJETIVOS DE
DESENVOLVIMENTO
SUSTENTÁVEL**Embrapa**

Agroindústria Tropical

**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Agroindústria Tropical
Ministério da Agricultura e Pecuária**

ISSN 1677-1915 / e-ISSN 2179-8184

Documentos 203

Abril, 2025

Base de dados de compostos naturais da
biodiversidade brasileira do Laboratório
Multiusuário de Química de Produtos Naturais

*Natália Florêncio Martins
Cléber de Freitas Fernandes
Kirley Marques Canuto
Lorena Mara Alexandre e Silva
Paulo Riceli Vasconcelos Ribeiro
Marcos José Andrade Viana
Yasmim Aquino Milhome
Iago Souza Vila Nova*

Embrapa Agroindústria Tropical
Fortaleza, CE
2025

Embrapa Agroindústria Tropical
Rua Pernambuco, 2.270, Pici
60.511-110 Fortaleza, CE
www.embrapa.br/agroindustria-tropical
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações

Presidente

José Roberto Vieira Junior

Secretária-executiva

Celli Rodrigues Muniz

Membros

Afrânio Arley Teles Montenegro, Aline

Saraiva Teixeira, Eveline de Castro

Menezes, Francisco Nelsieudes

Sombra Oliveira, Helenira Ellery

Marinho Vasconcelos, Kirley Marques

Canuto, Laura Maria Bruno, Marlon

Vagner Valentim Martins, Pablo Busatto

Figueiredo, Roselayne Ferro Furtado e

Sandra Maria Morais Rodrigues

Edição executiva

Celli Rodrigues Muniz

Revisão de texto

José Cesamildo Cruz Magalhães

Normalização bibliográfica

Rita de Cassia Costa Cid

Projeto gráfico

Leandro Sousa Fazio

Diagramação

José Cesamildo Cruz Magalhães

Imagens da capa

Natália Florêncio Martins

Publicação digital: PDF

Todos os direitos reservados

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Agroindústria Tropical

Base de dados de compostos naturais da biodiversidade brasileira do Laboratório Multiusuário de Química de Produtos Naturais / Natália Florêncio Martins... [et al.]. – Fortaleza : Embrapa AgroindústriaTropical, 2025.

PDF (14 p.) : il. color. – (Documentos / Embrapa Agroindústria Tropical, ISSN 2179-8184 ; 203).

1. Compostos químicos. 2. Laboratório. 3. Coleta de dados. 4. Gestão de dados. 5. *Chemical compounds*. 6. *Laboratory*. I. Martins, Natália Florêncio. II. Fernandes, Cléberson de Freitas. III. Canuto, Kirley Marques. IV. Silva, Lorena Mara Alexandre e. V. Ribeiro, Paulo Riceli Vasconcelos. VI. Viana, Marcos José Andrade. VII. Milhome, Yasmim Aquino. VIII. Nova, Iago Souza Vila. IX. Série.

CDD 542.1

Rita de Cassia Costa Cid (CRB-3/624)

© 2025 Embrapa

Autores

Natália Florêncio Martins

Bióloga, doutora em Bioquímica e Imunologia, pesquisadora da Embrapa Agroindústria Tropical, Fortaleza, CE

Cléber de Freitas Fernandes

Farmacêutico, doutor em Bioquímica, pesquisador da Embrapa Agroindústria Tropical, Fortaleza, CE

Kirley Marques Canuto

Farmacêutico, doutor em Química, pesquisador da Embrapa Agroindústria Tropical, Fortaleza, CE

Lorena Mara Alexandre e Silva

Química, doutora em Ressonância Magnética Nuclear, analista da Embrapa Agroindústria Tropical, Fortaleza, CE

Paulo Riceli Vasconcelos Ribeiro

Químico, doutor em Química, analista da Embrapa Agroindústria Tropical, Fortaleza, CE

Marcos José Andrade Viana

Bioinformata, doutorando em Bioinformática, analista da Embrapa Agroindústria Tropical, Fortaleza, CE

Yasmim Aquino Milhome

Técnica em Informática, graduanda em Farmácia, bolsista da Embrapa Agroindústria Tropical, Fortaleza, CE

Iago Souza Vila Nova

Graduando em Agronomia pela Universidade Federal do Ceará, bolsista da Embrapa Agroindústria Tropical, Fortaleza, CE

Apresentação

O Laboratório Multiusuário de Química de Produtos Naturais (LMQPN) da Embrapa desempenha um papel importante na investigação de compostos químicos presentes em plantas, utilizando tecnologias de ponta para análise e identificação. Com áreas especializadas e instrumentação analítica avançada, o laboratório realiza atividades de extração, fracionamento e caracterização molecular por meio de cromatografia gasosa e líquida, espectrometria de massas e ressonância magnética nuclear. Essas metodologias possibilitam a análise precisa de uma ampla gama de substâncias, permitindo o estudo detalhado de sua composição química e suas possíveis aplicações biotecnológicas.

Para assegurar a rastreabilidade e o uso contínuo das informações obtidas, desenvolvemos a

TropicalDB, uma base de dados relacional voltada à sistematização e preservação das informações químicas geradas pelas análises laboratoriais. A base organiza dados estruturais e composicionais, metadados relevantes, como taxonomia, atributos de coleta e coordenadas geográficas, promovendo o registro completo e interconectado das diferentes matrizes de origem.

Essa abordagem permite não apenas o armazenamento seguro e acessível, mas também facilita a prospecção de novos ativos tecnológicos, beneficiando os esforços de inovação da Embrapa e fortalecendo parcerias estratégicas.

O presente documento apresenta os argumentos e a concepção da base de dados, ressaltando o cenário do seu potencial inovador.

Gustavo Adolfo Saavedra Pinto
Chefe-Geral da Embrapa Agroindústria Tropical

Sumário

Introdução	9
Coleta de dados	9
Análises	10
Gestão de dados	10
Modelagem da estrutura do banco de dados	11
Enriquecimento de dados	12
Classificação dos dados no TropicalDB	12
Personalização e controle de acesso	12
Impactos da base de dados na Embrapa	13
Considerações finais	13
Referências	13

Introdução

O Laboratório Multiusuário de Química de Produtos Naturais (LMQPN) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) está localizado na Embrapa Agroindústria Tropical (CNPAT), em Fortaleza, CE. O LMQPN é um setor com áreas especializadas na extração, fracionamento, isolamento e identificação de substâncias químicas presentes em plantas. Essas análises contam com o uso de instrumentação analítica avançada, como cromatógrafos gasosos e líquidos, associados a poderosos espectrômetros de massas, ressonância magnética nuclear e infravermelho. Para garantir a rastreabilidade dos dados e seu potencial uso futuro, criamos o TropicalDB, uma base de dados relacional com o objetivo de sistematizar as informações sobre os compostos químicos identificados pelas análises laboratoriais, assegurando sua preservação e disponibilidade em longo prazo. Os parâmetros são registrados em planilhas que contêm metadados de análise, como taxonomia, atributos de coleta, coordenadas geográficas e a porcentagem composicional e respectivas estruturas de forma relacional, conforme as diferentes matrizes de origem. A organização da base de dados visa facilitar a prospecção de novos ativos tecnológicos pela Embrapa e seus colaboradores.

O LMQPN foi estabelecido seguindo normas e diretrizes definidas pela Deliberação publicada no Boletim de Comunicações Administrativas (BCA) nº 45/2012, em 15 de outubro de 2012. Posteriormente, essas normas foram homologadas pela Diretoria da Embrapa por meio do documento intitulado “Regras e procedimentos para uso do Laboratório Multiusuário de Química de Produtos Naturais (LMQPN)”, o qual se baseou na Norma nº 037.008.001.001. Esta norma foi ratificada pela Deliberação publicada no BCA nº 05/2021, de 16 de março de 2021.

O LMQPN foi criado com o propósito de potencializar o aproveitamento dos recursos naturais brasileiros, promovendo pesquisas que demandam instrumentação específica e de alta resolução. Sua missão é agregar valor à biodiversidade nacional e desenvolver tecnologias sustentáveis para diversas áreas da agroindústria.

Como laboratório multiusuário da Embrapa, especializado em Química de Produtos Naturais, o LMQPN oferece acesso a procedimentos de extração, fracionamento, isolamento e identificação de substâncias químicas presentes em plantas, utilizando instrumentação analítica sofisticada, tais

como cromatógrafos gasosos e líquidos, acoplados ou não a poderosos espectrômetros de massas, ressonância magnética nuclear e infravermelho.

As pesquisas transdisciplinares realizadas no LMQPN geram uma vasta quantidade de dados, colaborando com parceiros internos, universidades e empresas em áreas como Fitopatologia, Sanidade Animal, Entomologia, Farmacologia, Processos Agroindustriais, Ciência e Tecnologia de Alimentos e Materiais.

Assim, em razão da necessidade de gerenciar esses dados, foi criada uma base de dados denominada TropicalDB. A plataforma é acessível e segura para a gestão e pesquisa de compostos químicos provenientes de diversas fontes naturais, como óleos essenciais e extratos vegetais. Esses extratos são analisados em diferentes instrumentos analíticos, incluindo ressonância magnética nuclear (RMN), cromatografia gasosa acoplada à espectrometria de massas (GC-MS) e cromatografia líquida acoplada à espectrometria de massas (LC-MS). Portanto, a organização do TropicalDB está em consonância com as diretrizes do projeto Implementação e Monitoramento de Requisitos de Qualidade em Laboratórios Multiusuários da Embrapa (QualiMulti) que visa à implementação dos Requisitos de Qualidade da Embrapa (RQEs), para que eles operem segundo um padrão de qualidade único e internacional, garantindo excelência aos ensaios realizados e às infraestruturas compartilhadas.

A fim de suprir essa necessidade e facilitar o acesso a informações sobre produtos naturais, o presente trabalho tem como objetivo principal a geração de um banco de dados contendo informações relacionais de compostos químicos, visando identificar e auxiliar no gerenciamento dos dados analíticos produzidos pelo LMQPN. Este banco de dados, denominado TropicalDB, busca facilitar o acesso e apoiar a pesquisa de novos ativos agroindustriais e seus respectivos metadados, fornecendo acessibilidade, rastreabilidade e acesso sistemático aos dados. Essa abordagem possibilitará a identificação estrutural de substâncias químicas e contribuirá significativamente para o avanço da pesquisa nesse campo.

Coleta de dados

O material botânico de diversas matrizes fornece uma enorme variedade de moléculas e compostos. As matrizes geram extratos que, por sua vez,

analisados por técnicas de alta precisão, produzem perfis químicos correspondentes aos compostos majoritários e minoritários. Os produtos de origem natural são fracionados em componentes voláteis, solúveis, óleos essenciais, entre outros. Esse processamento inicial é crucial para preparar as amostras para análises subsequentes, garantindo que os

compostos de interesse sejam isolados e concentrados adequadamente. A plataforma de prospecção de compostos naturais do LMQPN, representada na Figura 1, corresponde às etapas de análises, identificação de perfis químicos, determinação dos compostos químicos e respectivos metadados que formam a base de dados TropicalDB.

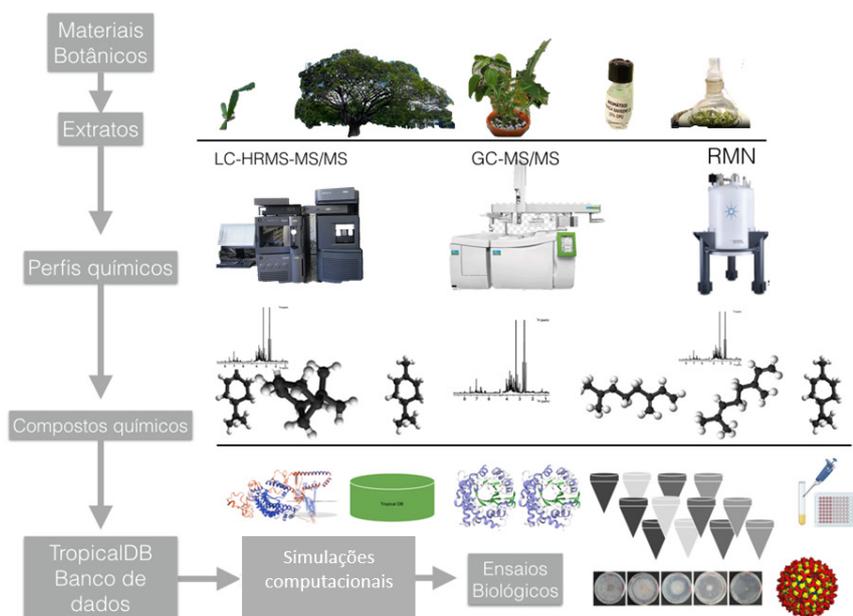


Figura 1. Esquema representativo da plataforma de obtenção e tratamento de dados realizados no Laboratório Multiusuário de Química de Produtos Naturais – LMQPN. Fonte: Natália Florêncio Martins (2023).

Análises

A caracterização e identificação dos compostos presentes nas amostras são realizadas por meio de várias técnicas analíticas, conforme representação na Figura 2.

Entre as técnicas empregadas, destacam-se:

- Cromatografia Líquida de Ultra Eficiência acoplada à Espectrometria de Massas de Alta Resolução (UPLC-QTOF-MS/MS): técnica empregada para separar, detectar e caracterizar metabólitos secundários complexos. A UPLC-QTOF-MS/MS é particularmente eficaz na obtenção de dados rápidos e precisos sobre a massa dos compostos, permitindo uma análise detalhada das amostras.

- Cromatografia Gasosa acoplada à Espectrometria de Massas (CG-EM): utilizada principalmente para a análise de substâncias voláteis, a CG-EM proporciona alta sensibilidade e especificidade na identificação dos componentes presentes nas amostras. Essa técnica é especialmente útil para estudar compostos em matrizes diversas.

- Espectroscopia de Ressonância Magnética Nuclear (RMN): fundamental para a elucidação de estruturas moleculares, a RMN é aplicada tanto a moléculas isoladas quanto a misturas complexas. Além disso, a RMN pode ser utilizada de forma quantitativa, oferecendo informações detalhadas sobre a composição química das amostras.

Gestão de dados

Os resultados das análises são organizados no banco de dados relacional visando auxiliar a gestão e o acesso a informações detalhadas sobre compostos químicos, facilitando a prospecção e o desenvolvimento de novos ativos agroindustriais. A gestão da informação segue as etapas descritas abaixo e ilustradas na Figura 1.

- Identificação e Catalogação: após a análise, os componentes moleculares das amostras são identificados e organizados em planilhas parametrizadas. Cada amostra é associada a um QR-Code

que permite o rastreamento e a recuperação de dados específicos.

- **Enriquecimento:** as tabelas de metadados são enriquecidas com informações taxonômicas, localização geográfica, parte da planta analisada e referências bibliográficas associadas. A coleta de metadados é automatizada por meio de Application Programming Interface (APIs), como a do *PubChem*,

que fornece atualizações regulares e acesso gratuito a informações quimiotaxonômicas.

- **Armazenamento:** os dados são armazenados fisicamente na extratoteca local e digitalmente no banco de dados TropicalDB. A organização desses dados permite a rápida recuperação e análise, apoiando pesquisas futuras.

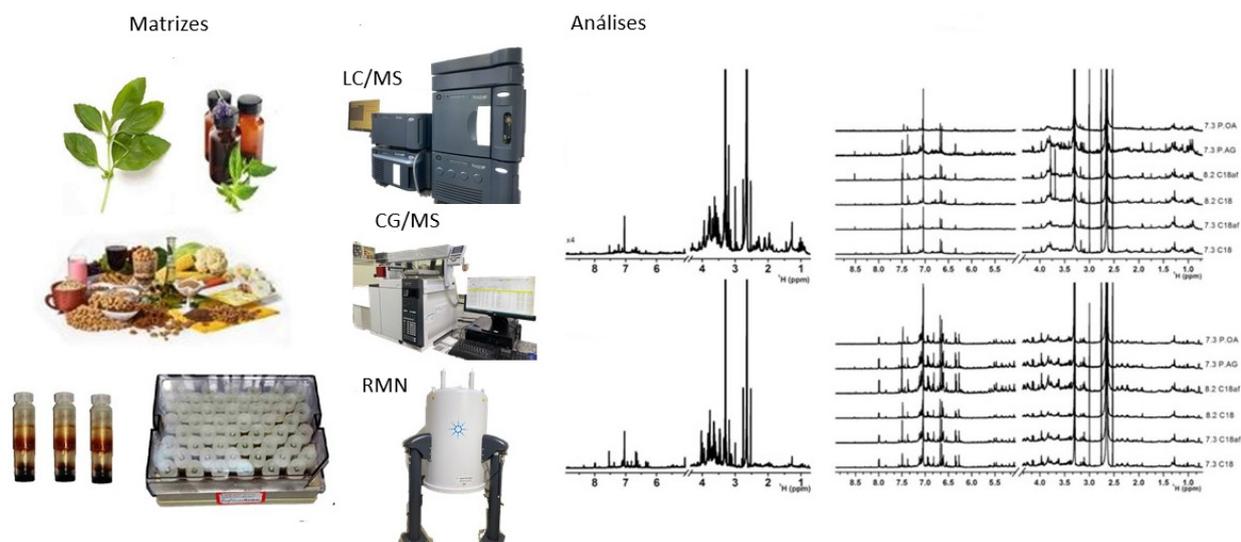


Figura 2. Esquema de análise de matrizes e geração de dados do Laboratório Multiusuário de Química de Produtos Naturais – LMQPN. Fonte: Lorena Mara Alexandre e Silva (2022).

Modelagem da estrutura do banco de dados

A modelagem da estrutura do banco de dados define como as informações são organizadas em tabelas e estabelece a hierarquia das palavras catalogadas. Esse processo meticuloso envolve a diagramação do fluxo de dados, que determina como as informações serão organizadas, armazenadas e recuperadas. Com essa abordagem, assegura-se a catalogação eficiente das informações em todos os níveis de busca, como, por exemplo, moléculas. A modelagem do banco é fundamental para garantir a integridade e a funcionalidade da plataforma.

A conexão entre o banco de dados e a planilha TropicalDB é estabelecida utilizando-se o modelo de banco de dados descrito em *Structured Query Language* (SQL). No núcleo desse modelo, encontra-se uma tabela central que armazena informações detalhadas sobre cada molécula. Entre os campos

principais dessa tabela, destaca-se o identificador único (ID), que é um número sequencial gerado automaticamente pelo sistema no momento da inserção de um novo registro. Além do ID, a tabela central inclui o nome e a fórmula molecular de cada molécula. Para complementar a tabela central, existem tabelas específicas que contêm propriedades físico-químicas e dados de análise espectroscópica. Essas tabelas são estrategicamente vinculadas à tabela-mestre por meio de chaves primárias e estrangeiras, garantindo a organização e a inter-relação dos dados de maneira eficiente.

A eficiência operacional do TropicalDB é significativamente aprimorada por meio da automação da extração de metadados. Esse processo é realizado utilizando-se SQL. As ferramentas SQL são comandos pré-compilados que executam tarefas repetitivas de maneira eficiente, otimizando as operações do banco de dados. As instruções são automaticamente executadas em resposta a eventos, como inserções ou atualizações de dados, garantindo a consistência e a integridade das informações.

Enriquecimento de dados

As tabelas de metadados do TropicalDB foram enriquecidas a partir da análise criteriosa das principais bases de dados padronizadas. O *PubChem* (2023) foi a base de dados escolhida como protocolo de coleta de metadados devido ao seu acesso gratuito e atualizações regulares de 3.529 compostos. Para tal, desenvolvemos um código em *JavaScript* para acessar as informações via API¹, permitindo a integração de dados do *PubChem* aos metadados do TropicalDB para cada composto químico.

Classificação dos dados no TropicalDB

A base de dados tem três categorias específicas:

Banco Ativo de Germoplasma (BAG)

O módulo BAG contém dados químicos associados às espécies depositadas nos Bancos Ativos de Germoplasma da Embrapa, disponíveis no sistema ALELO (Padua et al., 2020). Na primeira versão do TropicalDB, as espécies representadas nesta divisão são: feijão (*Phaseolus vulgaris*) (Bento et al., 2021); abacaxi (*Ananas comosus* L. Merr.); umbu (*Spondias tuberosa*); cajá (*Spondias mombin*) (Silva, 2018); alecrim (*Rosmarinus officinalis*); chambá (*Justicia pectoralis*); e curumaru (*Dipteryx odorata*). Espera-se que outras espécies sejam representadas nas futuras versões.

Banco de dados da biodiversidade brasileira

Nesta versão, a biodiversidade brasileira é representada nesta divisão, estando em constante construção e atualização. São duas subcategorias: o Depósito de Extratos (ExtractDB) e o Depósito de Óleos Essenciais (OEDB). As respectivas matrizes (extratos e óleos essenciais) são armazenadas e

preservadas em freezer, garantindo condições adequadas para a conservação dos componentes ativos de plantas de interesse agroquímico da biodiversidade brasileira.

Extrato de frutas (FruDB)

A base de dados FruDB é o módulo de acesso à composição química de frutas nativas e exóticas do Brasil que reúne dados de composição química de metabólitos primários, secundários e compostos voláteis das frutas.

Personalização e controle de acesso

Por meio da organização da base de dados no TropicalDB, tornamos viável o acesso aos compostos encontrados na biodiversidade brasileira, simplificando sua consulta e utilização em pesquisas voltadas para o desenvolvimento de novos ativos tecnológicos na Embrapa. O TropicalDB mostra registro de estudos, numa interface adaptável, garantindo acessibilidade, tanto em dispositivos móveis quanto em computadores. O TropicalDB oferece dois tipos de acesso: i) Visitantes – Este tipo de acesso permite apenas a visualização dos estudos cadastrados pelos pesquisadores; ii) Pesquisadores – Concede acesso a áreas restritas do site, onde podem realizar tarefas específicas relacionadas à pesquisa e ao gerenciamento de dados.

Essa estrutura flexível de acesso permite uma experiência adaptada às necessidades e permissões de cada usuário, otimizando a utilização do TropicalDB em diferentes contextos de pesquisa.

Adicionalmente, vale ressaltar que a base de dados abrange não apenas informações bidimensionais, mas também representações tridimensionais de moléculas, com base na API do *PubChem*. Esse recurso amplia ainda mais o entendimento e a utilidade do banco de dados, proporcionando uma visão mais completa das propriedades químicas das substâncias analisadas para fortalecer a base para futuras pesquisas e aplicações em melhoramento genético e bioeconomia.

⁽¹⁾ Interface de Programação de Aplicação (APIs) são mecanismos que permitem que dois componentes de software se comuniquem usando um conjunto de definições e protocolos.

Impactos da base de dados na Embrapa

A implementação do TropicalDB proporciona um ambiente de pesquisa e gestão de ativos da biodiversidade brasileira. Bancos de dados já consolidados, como, por exemplo, o ChEMBL (características físicas de substâncias), são estimados em milhões de euros, simbolizando a triagem de bibliotecas economicamente viáveis para a descoberta de medicamentos, pesticidas e outros compostos ativos.

O banco de dados espectrais chamado Amix (Bruker) (2023), que possibilita o acesso à base de dados de espectros de RMN de moléculas de baixo peso (aminoácidos, ácidos orgânicos, açúcares, etc.), cobra de seus usuários da academia aproximadamente 5 mil dólares por uma assinatura. Existe também o MNova + Wiley, que possui uma base de dados de ¹H e ¹³C de compostos orgânicos, cuja licença custa cerca de 5 mil Euros por ano (Jeannerat et al., 2021).

O TropicalDB v. 1.0 cria um ambiente de pesquisa da biodiversidade brasileira destinado a empresas, academia e instituições de pesquisa interessadas. Essa ferramenta visa facilitar a consulta e utilização de extratos, frutas e acessos, aumentando a eficiência na triagem, validação, extração e aplicação de produtos naturais. Além disso, promove o avanço do conhecimento sobre a quimiodiversidade e variabilidade metabólica desses produtos, oferecendo uma oportunidade única de agregar valor aos ativos provenientes da biodiversidade.

A construção dessa base de dados pelo LM-QPN é uma proposta ampla, multidimensional e atemporal. Diante do cenário de crescente geração de dados na era do conjunto de dados complexos, volumosos e variados que exigem tecnologias avançadas para análise e processamento, a ampliação em magnitude e conteúdo do TropicalDB pode torná-lo acessível para toda a empresa. No entanto, o crescimento da base exigirá recursos computacionais e investimentos para a manutenção do sistema. A utilização e a segurança dos dados serão regulamentadas por uma política de acesso, por meio da qual o TropicalDB será disponibilizado aos usuários cadastrados.

Considerações finais

A química de produtos naturais desempenha um papel essencial na agricultura sustentável, permitindo o desenvolvimento de biofertilizantes e biopesticidas que reduzem impactos ambientais, promovem a saúde do solo e fortalecem a resistência das plantas, contribuindo para a segurança alimentar global.

No LMQPN, a geração de dados e o uso de ferramentas computacionais permitem identificar compostos naturais promissores para aplicações agrícolas. Essa abordagem integra triagens virtuais e testes experimentais, potencializando a descoberta de soluções inovadoras baseadas na biodiversidade brasileira.

A organização e proteção desses dados são fundamentais para garantir competitividade, segurança e colaboração no setor. Com isso, a Embrapa e seus parceiros impulsionam a pesquisa e o desenvolvimento de tecnologias avançadas, promovendo uma agricultura mais moderna e eficiente.

Referências

AMIX. Disponível em: <https://www.bruker.com/en/products-and-solutions/mr/nmr-software/amix.html>.

Acesso em: 27 dez. 2023.

BENTO, J. A. P.; RIBEIRO, P. R. V.; SILVA, L. M. A. e; ALVES FILHO, E. G.; BASSINELLO, P. Z.; BRITO, E. S. de; CALIARI, M.; SOARES JÚNIOR, M. S. Chemical profile of colorful bean (*Phaseolus vulgaris* L) flours: Changes influenced by the cooking method. **Food Chemistry**, v. 356, p. 129718, set. 2021.

EMBRAPA AGROINDÚSTRIA TROPICAL. Deliberação aprovada em 16 de março de 2021. "Regras e procedimentos para uso do Laboratório Multiusuário de Química de Produtos Naturais (LMQPN)". **Boletim de Comunicação Administrativa**, Brasília, DF, nº 05, 2021.

JEANNERAT, D.; COBAS, C. Application of multiplet structure deconvolution to extract scalar coupling constants from 1D nuclear magnetic resonance spectra. **Magnetic Resonance**, v. 2, n. 2, p. 545-555, 2021.

PADUA, J. G.; ALBUQUERQUE, M. do S. M.; MELLO, S. C. M. de (ed.). **Bancos e coleções de germoplasma da Embrapa**: conservação e uso. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 2020. 64 p. (Embrapa Recursos Genéticos e Biotecnologia. Documentos, 371). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1124923>. Acesso em: 27 dez. 2023.

PUBCHEM. **PubChem**. Disponível em: <https://pubchem.ncbi.nlm.nih.gov/>. Acesso em: 27 dez. 2023.

SILVA, H. N. **Criação de um modelo de biblioteca de extratos (extratoteca) aplicado a banco ativo de germoplasma de folhas de cajazeira**. 2018. 83 f. Trabalho de Conclusão de Curso (Graduação em Química Bacharelado) – Universidade Federal do Ceará, Fortaleza.

