

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO  
CAMPUS CRISTALINA  
BACHARELADO EM AGRONOMIA**

JÚLIO CÉSAR CAETANO DE OLIVEIRA

**BIOTECNOLOGIA NA RESISTÊNCIA DE PLANTAS A DOENÇAS:** uma  
análise exploratória a partir de trabalhos de conclusão de curso

Cristalina

2025

JÚLIO CÉSAR CAETANO DE OLIVEIRA

**BIOTECNOLOGIA NA RESISTÊNCIA DE PLANTAS A DOENÇAS:** uma  
análise exploratória a partir de trabalhos de conclusão de curso

Trabalho de curso apresentado ao curso de Bacharelado em Agronomia do Instituto Federal Goiano – Campus Cristalina, como requisito parcial para a obtenção do título de Bacharel em Agronomia.

Orientador(a): Edivaldo Barbosa de Almeida Júnior

Cristalina

2025

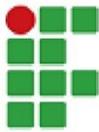
**Ficha de identificação da obra elaborada pelo autor, através do  
Programa de Geração Automática do Sistema Integrado de Bibliotecas do IF Goiano - SIBi**

O48b Oliveira, Júlio César Caetano de  
BIOTECNOLOGIA NA RESISTÊNCIA DE PLANTAS A  
DOENÇAS: UMA ANÁLISE EXPLORATÓRIA A PARTIR  
DE TRABALHOS DE CONCLUSÃO DE CURSO / Júlio César  
Caetano de Oliveira. Cristalina 2025.

40f. il.

Orientador: Prof. Dr. Edivaldo Barbosa de Almeida Júnior.  
Tcc (Bacharel) - Instituto Federal Goiano, curso de 1020024 -  
Bacharelado em Agronomia - Cristalina (Campus Cristalina).

I. Título.



**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

**Identificação da Produção Técnico-Científica**

- |  |   |
|--|---|
| <input type="checkbox"/> Tese  | <input type="checkbox"/> Artigo Científico              |
| <input type="checkbox"/> Dissertação                                 | <input type="checkbox"/> Capítulo de Livro              |
| <input type="checkbox"/> Monografia – Especialização                 | <input type="checkbox"/> Livro                          |
| <input checked="" type="checkbox"/> TCC - Graduação                  | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ |   |

Nome Completo do Autor: Júlio César Caetano de Oliveira

Matrícula: 2021110200240165

Título do Trabalho: Biotecnologia na resistência de plantas a doenças: uma análise exploratória a partir de trabalhos de conclusão de curso

**Restrições de Acesso ao Documento**

Documento confidencial:  Não  Sim, justifique:

Informe a data que poderá ser disponibilizado no RIIF Goiano: 01/01/2027

O documento está sujeito a registro de patente?  Sim  Não

O documento pode vir a ser publicado como livro?  Sim  Não

**DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA**

O/A referido/a autor/a declara que:

- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Cristalina, 15/09/2025.  
Local Data

Documento assinado digitalmente  
**gov.br** JULIO CESAR CAETANO DE OLIVEIRA  
Data: 15/09/2025 18:40:09-0300  
Verifique em <https://validar.iti.gov.br>

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:

Documento assinado digitalmente  
**gov.br** EDIVALDO BARBOSA DE ALMEIDA JUNIOR  
Data: 15/09/2025 18:30:29-0300  
Verifique em <https://validar.iti.gov.br>

Assinatura do(a) orientador(a)



SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Documentos 19/2025 - GE-CRT/CMPCRIS/IFGOIANO

BACHARELADO EM AGRONOMIA

**Biotechnologia na resistência de plantas a doenças: uma análise exploratória a partir de trabalhos de conclusão de curso**

Autor: Júlio César Caetano de Oliveira

Orientador: Edivaldo Barbosa de Almeida Júnior

TITULAÇÃO: **Bacharel em Agronomia**

**APROVADA** em 15 de Agosto de 2025

Prof. Dr. Edivaldo Barbosa de Almeida Júnior

Presidente da Banca

IF Goiano Campus Cristalina

Profa. Dra. Geisiane Alves Rocha

Membro da Banca

IF Goiano Campus Cristalina

Prof. Dra. Giselle Anselmo de Souza Goncalves

Membro da Banca

IF Goiano Campus Cristalina

Documento assinado eletronicamente por:

- **Edivaldo Barbosa de Almeida Junior**, PROFESSOR ENS BASICO TECN TECNOLOGICO , em 15/08/2025 13:23:35.
- **Giselle Anselmo de Souza Goncalves**, PROFESSOR ENS BASICO TECN TECNOLOGICO , em 18/08/2025 08:41:47.
- **Geisiane Alves Rocha**, PROFESSOR ENS BASICO TECN TECNOLOGICO , em 25/08/2025 15:45:51.

Este documento foi emitido pelo SUAP em 15/08/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.igoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

**Código Verificador:** 733054

**Código de Autenticação:** be1813070a



INSTITUTO FEDERAL GOIANO  
Campus Cristalina  
Rua Araguaia, Loteamento 71, SN, Setor Oeste, CRISTALINA / GO, CEP 73850-000  
(61) 3612-8500

## AGRADECIMENTOS

Dedico esse trabalho ao meu Senhor e Salvador, Jesus Cristo!

A realização deste trabalho não seria possível sem o apoio, incentivo e amor de pessoas muito especiais, às quais dedico minha mais sincera gratidão.

Ao meu Professor e Orientador Edivaldo, por sua dedicação, paciência e disponibilidade ao longo deste percurso. Sua orientação foi essencial para o desenvolvimento deste trabalho, sempre oferecendo contribuições valiosas, com firmeza e sensibilidade. Sou grato por ter tido a oportunidade de aprender sob sua orientação.

À minha mãe Célia, exemplo de força e dedicação. Agradeço por todo o amor, conselhos e orações, que me sustentaram mesmo nos momentos em que o cansaço parecia maior que a vontade de continuar. Tudo isso é também fruto do seu esforço e sacrifício.

À minha namorada Letícia, pelo carinho constante, pela paciência nos momentos difíceis e por acreditar em mim mesmo quando eu duvidava. Sua presença foi um alento nos dias mais desafiadores, e sua companhia tornou esse caminho mais leve e significativo.

Aos meus irmãos, companheiros de vida, que com suas palavras, exemplos e até mesmo silêncios, contribuíram para minha formação pessoal. Obrigado por cada gesto de incentivo, por cada demonstração de afeto e pelo suporte incondicional.

A todos vocês, meu profundo agradecimento. Este trabalho é dedicado também a cada um que fez parte dessa caminhada.

Muito obrigado!

Não temas, porque eu sou contigo; não te assombres, porque eu sou o teu Deus; eu te fortaleço, e te ajudo, e te sustento com a destra da minha justiça.

Isaías 41:10

## RESUMO

O presente trabalho tem como objetivo analisar, sob uma perspectiva exploratória, a produção científica brasileira relacionada ao uso da biotecnologia na resistência de plantas a doenças, com ênfase no avanço do melhoramento genético vegetal. A partir de um levantamento bibliográfico realizado em bases institucionais indexadas no Google Acadêmico, foram identificados e analisados trabalhos acadêmicos publicados entre 2013 e 2023. A pesquisa utilizou métodos quantitativos e ferramentas de mineração de texto para examinar a frequência de termos, padrões de co-ocorrência, agrupamentos temáticos e variações temporais na linguagem científica. Os resultados apontam a centralidade de conceitos como resistência, marcadores moleculares, genes e doenças vegetais, refletindo o protagonismo das ferramentas biotecnológicas no desenvolvimento de estratégias sustentáveis para o controle fitopatológico. Além disso, identificou-se uma concentração significativa da produção em dissertações de mestrado e uma tendência de queda nas publicações nos últimos anos, relacionada às restrições orçamentárias que impactaram diretamente a pós-graduação no Brasil. Os dados obtidos reforçam a relevância da biotecnologia como aliada no enfrentamento de desafios da agricultura contemporânea e a necessidade de políticas públicas que assegurem a continuidade da produção científica nesse campo estratégico.

**Palavras-chave:** Biotecnologia; Controle de doenças de plantas; Ferramentas moleculares; Melhoramento genético vegetal; Transformação genética.

## ABSTRACT

This study aims to analyze, from an exploratory perspective, Brazilian scientific production related to the use of biotechnology in plant resistance to diseases, with an emphasis on the advancement of plant genetic improvement. Based on a bibliographic survey carried out in institutional databases indexed in Google Scholar, academic papers published between 2013 and 2023 were identified and analyzed. The research used quantitative methods and text mining tools to examine the frequency of terms, co-occurrence patterns, thematic groupings, and temporal variations in scientific language. The results point to the centrality of concepts such as resistance, molecular markers, genes, and plant diseases, reflecting the leading role of biotechnological tools in the development of sustainable strategies for phytopathological control. In addition, a significant concentration of production in master's dissertations and a downward trend in publications in recent years were identified, related to budgetary constraints that directly impacted postgraduate studies in Brazil. The data obtained reinforce the relevance of biotechnology as an ally in facing the challenges of contemporary agriculture and the need for public policies that ensure the continuity of scientific production in this strategic field.

**Keywords:** Biotechnology; Genetic transformation; Molecular tools; Plant disease control; Plant genetic improvement.

## LISTA DE ILUSTRAÇÕES

- Figura 1.** Distribuição percentual das categorias de documentos indexados no Google Acadêmico relacionados ao uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos.....15
- Figura 2.** Evolução do número de estudos ao longo dos anos indexados no Google Acadêmico sobre o uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos.....17
- Figura 3.** Frequência de termos relacionados a biotecnologia no melhoramento de plantas em trabalhos indexados no Google Acadêmico sobre o uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos. ....18
- Figura 4.** Distribuição de publicações sobre grupos patogênicos em trabalhos acadêmicos que abordam o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....19
- Figura 5.** Agrupamento de termos recorrentes em trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....20
- Figura 6.** Frequência dos principais termos encontrados em documentos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....22
- Figura 7.** Rede de co-ocorrência entre os principais termos presentes em trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....23
- Figura 8.** Evolução temporal da frequência de termos utilizados em resumos de trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....24
- Figura 9.** Orientadores com maior número de trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos.....25

## SUMÁRIO

1 INTRODUÇÃO .....	12
2 MATERIAL E MÉTODOS .....	15
3 RESULTADOS E DISCUSSÃO .....	20
4 CONSIDERAÇÕES FINAIS.....	31
REFERÊNCIAS.....	33
APÊNDICE.....	37

## 1 INTRODUÇÃO

A bibliometria configura-se como uma abordagem quantitativa voltada para a mensuração e interpretação da produção científica, permitindo o mapeamento de tendências e a identificação de padrões de comunicação entre autores, instituições e áreas do conhecimento. Seu potencial analítico está na capacidade de transformar registros bibliográficos em indicadores que revelam tanto a evolução interna de uma disciplina quanto sua interação com outros campos. Para Guedes e Borschiver (2005), a bibliometria constitui um recurso estratégico não apenas para compreender a dinâmica da ciência, mas também para subsidiar processos de gestão, avaliação e planejamento em pesquisa, assumindo relevância crescente em estudos exploratórios. Essas práticas tiveram origem na década de 1960, a partir de iniciativas da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) e da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que buscaram desenvolver metodologias destinadas à avaliação da atividade científica e tecnológica (Spinak, 1998; Figueira et al., 1999, citados por Garrido & Rodrigues, 2011). Com base nas teorias de Lotka e Zipf, apresentadas por Santos e Kobashi (2009), Price ampliou as fronteiras dos estudos quantitativos ao direcionar sua atenção à dinâmica da atividade científica, contemplando tanto os produtos quanto os agentes envolvidos na produção do conhecimento.

Nessa perspectiva, compreender o progresso de um domínio específico do saber revela-se uma tarefa complexa. Durante séculos, a produção e a consolidação do pensamento científico ocorreram de maneira gradual. Entretanto, o avanço das indústrias da informação e do conhecimento, associado às transformações nas disciplinas científicas e ao desenvolvimento das tecnologias de informação e comunicação (Le Coadic, 2003; Targino, 2000), intensificou a produção e a disseminação de pesquisas. Como consequência, a análise do estado atual de qualquer campo científico tornou-se mais desafiadora, visto que a nova dinâmica da produção de conhecimento impõe aos pesquisadores um obstáculo semelhante ao enfrentado pelas organizações: administrar volumes de informação superiores à capacidade humana de processamento.

Esse movimento internacional de expansão da ciência repercutiu também no Brasil, cuja trajetória científica e de políticas de fomento à pesquisa pode ser considerada recente (Borges, 2016). Até o início do século XX, o ensino superior brasileiro era formado, predominantemente, por escolas profissionais, academias militares e instituições isoladas (Barreto; Filgueiras, 2007). O modelo que fundamenta as universidades nacionais surgiu apenas na década de 1920, com o

propósito de institucionalizar um sistema até então fragmentado e desprovido de regulamentação (Guimarães, 2002). Até a década de 1940, a pesquisa científica concentrava-se em poucos centros públicos voltados às áreas da saúde, da agricultura e da tecnologia industrial. Apenas no período pós-guerra ocorreram ações mais sistemáticas de apoio à ciência, inspiradas no modelo linear de inovação norte-americano (Furtado, 2005; Ottoboni, 2011).

Nesse contexto, Oliveira et al. (1992) defendem que a avaliação da produtividade científica constitui elemento central para o estabelecimento e acompanhamento de políticas nacionais de ensino e pesquisa, uma vez que possibilita diagnosticar as reais potencialidades de grupos e instituições. Torna-se, assim, necessária a adoção de técnicas de avaliação específicas, de caráter quantitativo, qualitativo ou combinado. Entre os métodos de natureza quantitativa, destacam-se a bibliometria, a cienciometria, a informetria e, mais recentemente, a webometria. Apesar de apresentarem funções semelhantes, cada uma delas avalia a disseminação do conhecimento científico e o fluxo da informação sob diferentes perspectivas, embora haja dificuldades em delimitar com precisão suas fronteiras (Vanti, 2002).

A bibliometria e a cienciometria, em particular, oferecem um conjunto consistente de métricas aplicáveis à análise da comunicação científica (Camargo; Barbosa, 2018), com ênfase recorrente na avaliação do desempenho de indivíduos ou grupos de pesquisadores. No Brasil, diversas áreas vêm registrando crescimento expressivo na produção de estudos com enfoque cienciométrico (Araújo; Alvarenga, 2011).

É a partir desse avanço metodológico que a cienciometria se revela instrumento estratégico para mapear o desenvolvimento de áreas específicas do conhecimento, como ocorre com a biotecnologia aplicada ao melhoramento genético de plantas. Dentre os setores mais impactados pelas inovações científicas e tecnológicas, destaca-se o melhoramento genético vegetal, cuja transição para métodos biotecnológicos configura uma mudança significativa, ainda carente de investigações aprofundadas sobre sua aplicação na agricultura contemporânea. Nesse âmbito, a biotecnologia tem desempenhado papel crucial no desenvolvimento de plantas resistentes a patógenos, consolidando-se como ferramenta estratégica para o manejo sustentável de doenças em diferentes culturas. Técnicas como a transgenia e a edição gênica vêm possibilitando a criação de variedades vegetais mais tolerantes a fungos, bactérias e vírus, contribuindo para a redução do uso de defensivos químicos e para a adoção de práticas agrícolas ambientalmente mais sustentáveis (Wang et al., 2019).

Considerando esse panorama, a presente pesquisa adota abordagem exploratória com o intuito de identificar trabalhos acadêmicos publicados entre 2013 e 2023 relacionados ao

emprego da biotecnologia no melhoramento genético de plantas com ênfase em doenças. Busca-se, ainda, oferecer uma visão panorâmica do avanço científico brasileiro nesse campo, evidenciando as principais tendências, autores, instituições e contribuições do período.

A opção por essa abordagem mostra-se especialmente pertinente para a área de Agronomia, por fornecer uma perspectiva abrangente acerca das pesquisas em biotecnologia aplicada ao melhoramento genético vegetal, permitindo identificar tendências emergentes, lacunas investigativas e padrões de colaboração (Judijanto & Auliani, 2024). Esse mapeamento sistemático contribui para a formulação de agendas de pesquisa mais coerentes e orientadas aos desafios agrícolas contemporâneos, como o manejo de doenças e a sustentabilidade dos sistemas produtivos (Das et al., 2023). Por fim, tal diagnóstico bibliográfico constitui base para a proposição de novos estudos que acelerem a incorporação de ferramentas biotecnológicas na prática agrônoma, favorecendo tanto o aumento da produtividade quanto a redução da dependência de agroquímicos.

## 2 MATERIAL E MÉTODOS

Para a construção desta pesquisa, foram consultadas diferentes fontes acadêmicas, como trabalhos de conclusão de curso, dissertações e teses que se encontram disponíveis em bases de dados institucionais indexadas no Google Acadêmico, conhecido como Google Scholar, é uma ferramenta de busca voltada para a literatura científica, incluindo artigos, livros, teses e trabalhos apresentados em conferências. Ele permite que pesquisadores localizem rapidamente publicações relevantes, acompanhem o número de citações de determinados trabalhos e identifiquem padrões de colaboração entre autores e instituições. Além disso, disponibiliza métricas básicas, como índice h e contagem de citações, que podem ser utilizadas em estudos bibliométricos e análises exploratórias da ciência. Apesar de não substituir bases de dados tradicionais em termos de rigor metodológico, o Google Acadêmico oferece uma cobertura ampla e interdisciplinar, sendo especialmente útil para mapeamentos preliminares e identificação de tendências emergentes na produção científica (Jacsó, 2005; Meho & Yang, 2007). A seleção dos termos e dos dados foi realizada criteriosamente, visando alcançar o maior número possível de estudos relevantes para o tema.

Foram definidas palavras-chave principais e palavras compostas secundárias com o objetivo de analisar relações conceituais entre os termos. As palavras-chave principais selecionadas foram: Biotecnologia, Ferramenta Molecular e Melhoramento Genético. As palavras compostas secundárias incluíram: Interação planta-patógeno, Resistência a doenças, Transformação gênica, Fitopatógenos, Doenças de plantas e Mecanismos de defesa a fitopatógenos.

A metodologia adotada consistiu na formação de pares de palavras, sendo cada palavra-chave principal combinada com todas as palavras compostas secundárias. Inicialmente, a primeira palavra principal, Biotecnologia, foi associada a todas as palavras secundárias. Em seguida, o mesmo procedimento foi aplicado à segunda palavra principal, Ferramenta Molecular, e posteriormente à terceira, Melhoramento Genético.

Para a construção desses pares, utilizou-se o operador booleano AND, o que possibilitou a sistematização e a precisão na identificação das relações entre os termos. Entretanto, reconhece-se que o uso exclusivo do Google Scholar como base de dados apresenta limitações. Embora sua ampla cobertura e facilidade de acesso sejam vantajosas, essa plataforma não oferece um controle rigoroso de indexação e padronização dos metadados, o que pode resultar em duplicações, ausência de filtros específicos ou inclusão de documentos fora do escopo

acadêmico formal. A ausência de uma taxonomia detalhada na base de dados entendida como uma estrutura organizada e sistemática de categorias que permite classificar os artigos de acordo com seus temas e conteúdos pode comprometer a precisão na identificação de estudos altamente relevantes. Para minimizar essas limitações, realizou-se uma análise manual da pertinência de cada estudo, avaliando individualmente a relevância em relação aos objetivos da pesquisa.

Após essa etapa, os estudos identificados foram submetidos a uma análise de relevância em relação ao tema, considerando especificamente aqueles com ênfase em biotecnologia aplicada a plantas, com foco em doenças vegetais. Todos os materiais selecionados foram organizados no Mendeley®, onde foram salvos em formato PDF para facilitar uma análise mais detalhada posteriormente. Com o objetivo de manter a organização e facilitar o rastreamento, cada documento foi numerado de forma sequencial, utilizando um código iniciado em (DE0001). Essa identificação acompanhou todas as etapas seguintes da análise.

De cada artigo selecionado, foram extraídas as seguintes informações: (1) título do periódico; (2) autor; (3) ano de publicação; (4) palavras-chave do periódico; (5) arquivo PDF; e (6) link de acesso. Para padronização, os dados foram editados, mantendo a primeira letra de cada palavra em maiúscula, o que contribuiu para uma melhor visualização e organização dos registros. Na sequência, essas informações foram transferidas manualmente para uma planilha no Excel, organizada por categorias. Os campos dessa planilha incluíram: (1) Identificador (ID); (2) tipo de publicação; (3) autor; (4) primeiro autor; (5) código do autor; (6) instituição do primeiro autor; (7) instituição do autor correspondente; (8) país de origem do primeiro autor; (9) título do trabalho; (10) ano; (11) referência; (12) universidade; (13) código da universidade; (14) ferramenta molecular utilizada; (15) espécie-alvo; (16) nome científico da espécie-alvo; (17) resultado obtido ou esperado; (18) gene molecular; (19) tipo de patógeno; e (20) espécie ou família do patógeno.

Foram definidos critérios de seleção para selecionar trabalhos diretamente relacionados à aplicação da biotecnologia no melhoramento genético de plantas com foco na resistência a doenças, com o objetivo de refinar a relevância temática dos documentos analisados. Inicialmente, foram descartados os registros que apresentavam inconsistências ou ausência de informações fundamentais, como título ou ano de publicação, uma vez que esses campos são essenciais para a estruturação e análise cronológica dos dados. Adicionalmente, durante a etapa de triagem manual, foram excluídos artigos que, embora aparecessem nos resultados de busca por conter os termos-chave, não apresentavam efetiva relação com a temática central da pesquisa. Isso incluiu trabalhos que abordavam biotecnologia de maneira genérica, sem

conexão direta com o melhoramento genético vegetal ou com aspectos fitopatológicos, bem como estudos focados em biotecnologia humana, animal ou ambiental. Também foram eliminados documentos duplicados e resumos excessivamente curtos ou vagos, que impediriam a avaliação adequada de sua pertinência.

Para a análise textual exploratória, foram utilizadas ferramentas computacionais que permitiram explorar a frequência, co-ocorrência e evolução de termos nos resumos dos trabalhos selecionados. Todas as etapas analíticas foram conduzidas no ambiente estatístico R, versão 4.3.2 (R Core Team, 2024). Para a análise bibliométrica inicial e importação dos metadados em formato RIS, utilizou-se o pacote *Synthesizr* (westgate; grames, 2020). Na sequência, foi empregada a ferramenta *Voyant Tools*, versão 2.4 (Sinclair; Rockwell, 2016), para a exploração visual interativa do corpus textual, incluindo nuvem de palavras, frequência de termos, co-ocorrência e evolução temporal. Para a análise de redes de co-ocorrência e clusters temáticos, utilizou-se também o software *VOSviewer*, versão 1.6.20 (van Eck; Waltman, 2010), específico para visualização de mapas científicos baseados em dados bibliográficos.

Complementarmente, foram utilizados os seguintes pacotes no R: *tidytext* (Silge; Robinson, 2016), para tokenização e organização dos dados textuais; *quanteda* (v.3.0.0; Benoit et al., 2018), para construção da matriz documento-termo; *dplyr* (Wickham et al., 2024), para manipulação dos dados; *igraph* (Csardi; Nepusz, 2006), para criação da rede semântica de termos; e *ggraph* (Pedersen, 2024), para visualização gráfica das relações de co-ocorrência. A inclusão dessas ferramentas visou garantir a reprodutibilidade e a confiabilidade dos resultados obtidos, alinhando-se às boas práticas da pesquisa científica computacional.

A análise dos dados desta pesquisa foi conduzida em etapas bem definidas, utilizando diversos pacotes especializados. Inicialmente, os registros bibliográficos foram obtidos por meio de exportação da base Google Acadêmico, em formato *RIS (Research Information Systems)*, contendo metadados nomeados como *Title*, ano de defesa *Year*, resumo *Abstract* e autores *Author*. Esses dados foram importados utilizando o pacote *Synthesizr*, que converteu o arquivo *RIS (Research Information Systems)* em um objeto do tipo *data.frame*.

Com o conjunto de dados carregado, procedeu-se à limpeza inicial. Foram excluídos os registros com valores ausentes nos campos *Title* ou *Year*, bem como removidas duplicatas exatas com base na combinação desses dois campos. O resultado desse processo foi um banco refinado de dados, pronto para as análises subsequentes. Na etapa descritiva, buscou-se identificar padrões na evolução temporal da produção científica. Para isso, calculou-se o

número de trabalhos publicados por ano. Os resultados foram visualizados por meio de um gráfico de barras construído com o pacote *ggplot2* (Wickham, 2024), o que permitiu observar tendências de crescimento ou queda ao longo do tempo. Em complemento, realizou-se uma análise para identificar os orientadores mais frequentes. A partir disso, os dez orientadores mais recorrentes foram identificados e os resultados foram exibidos em um gráfico de barras horizontais. Na sequência, foi realizada a mineração de texto sobre os resumos dos trabalhos. Em seguida foi construído um corpus textual a partir do campo *Abstract*.

No intuito de garantir a consistência e a qualidade dos dados textuais analisados, foi realizado um cuidadoso processo de processamento e pré-processamento do *corpus*. Em um primeiro momento, os dados foram padronizados com a conversão de todos os termos para letras minúsculas, medida essencial para evitar duplicidade de palavras que, embora semanticamente idênticas, poderiam ser tratadas como distintas em função da capitalização. A organização visual com a primeira letra em maiúscula foi mantida apenas na planilha de controle, sendo corrigida antes das etapas de análise computacional. Posteriormente, foram eliminados caracteres não textuais como pontuações, números e símbolos, além das chamadas *stopwords*, que são palavras de uso muito comum na língua portuguesa como “de”, “a”, “o”, “em” que não agregam valor analítico ao contexto.

A partir dessa base limpa, foi construída a Matriz Documento-Termo (DFM), que foi posteriormente ajustada para manter apenas termos com frequência igual ou superior a cinco ocorrências. Após a construção da DFM, foram identificados os 30 termos mais frequentes, e os resultados foram representados graficamente em um gráfico de barras utilizando o pacote *quanteda*.

Com os termos mais utilizados identificados, partiu-se para a construção de uma rede de co-ocorrência entre eles. Selecionaram-se os 10 termos mais frequentes para essa análise. A matriz de co-ocorrência foi gerada a partir da DFM, e dela foi extraído o subgrafo correspondente. A estrutura foi representada por meio de um objeto *igraph*, no qual os nós representam os termos e as arestas indicam o peso da co-ocorrência. A visualização do grafo foi realizada com o pacote *ggraph*, empregando o layout de Fruchterman–Reingold. As arestas foram ajustadas com espessura proporcional ao peso da relação entre os termos. Buscando compreender a dinâmica temporal da linguagem utilizada nas produções científicas analisadas, foi realizada uma análise temporal dos termos mais frequentes. Os cinco termos mais recorrentes foram identificados e analisados ao longo dos anos. Para isso, os resumos dos trabalhos foram convertidos em palavras individuais com o auxílio do pacote *tidytext*,

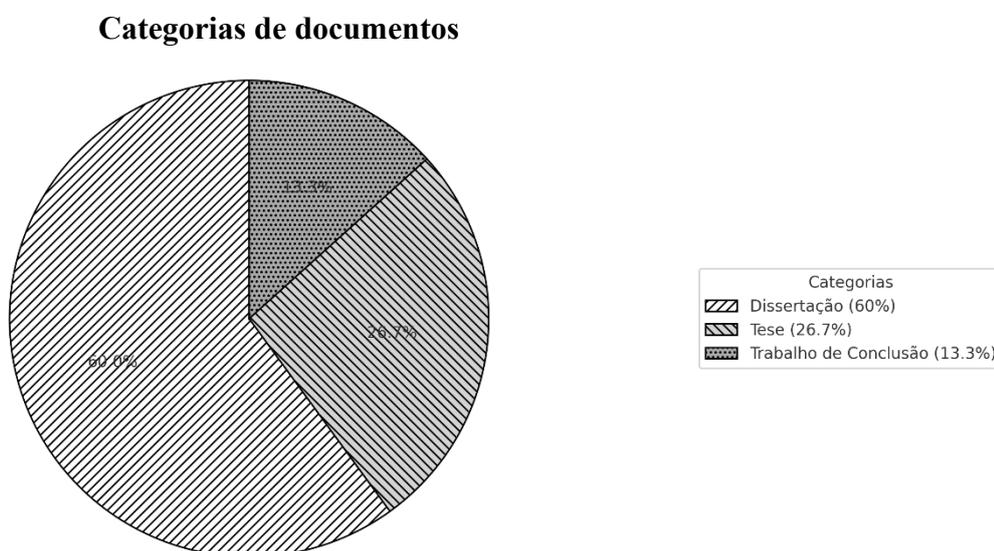
permitindo a estruturação dos dados textuais. Em seguida, as frequências de ocorrência de cada termo foram calculadas com o suporte do pacote *dplyr*, possibilitando o monitoramento do uso desses termos em diferentes períodos. A representação visual dessa evolução temporal foi realizada por meio de gráficos de linha, elaborados para evidenciar tendências e mudanças no vocabulário ao longo do tempo.

Na etapa final da análise, foi aplicado o agrupamento hierárquico dos resumos, com o objetivo de identificar padrões de similaridade semântica entre os documentos. A DFM foi convertida em vetores numéricos e, com base nela, foram calculadas as distâncias euclidianas entre os resumos. O agrupamento foi realizado com o método de ligação de Ward, e os resultados foram apresentados em forma de dendrograma, o que resultou na formação de quatro clusters principais, cada um reunindo documentos com características lexicais semelhantes. Para caracterização desses agrupamentos, foi reconstruída uma DFM restrita aos resumos de cada cluster, a partir da qual foram extraídos os dez termos mais frequentes. Esses termos foram apresentados em gráficos de barras posicionados junto ao dendrograma, permitindo uma análise visual e interpretativa do conteúdo predominante em cada grupo.

### 3 RESULTADOS E DISCUSSÃO

Foram avaliados um total de 74 trabalhos, sendo observado que as dissertações representam a maior parcela dos registros. A Figura 1, construída com o pacote ggplot2, ilustra essa distribuição, indicando que as dissertações correspondem a 60% do total analisado. Esse dado indica que, mesmo diante das dificuldades enfrentadas ao longo dos anos, os programas de pós-graduação têm mantido certa regularidade na produção científica, sobretudo quando comparados aos trabalhos de conclusão de curso (TCCs). A predominância de dissertações pode ser explicada pelo número expressivo de programas de mestrado, os quais exigem, como requisito obrigatório, a elaboração desse tipo de trabalho acadêmico. As teses, por sua vez, correspondem a 26,7% dos documentos identificados, o que evidencia uma produção significativa de pesquisas em nível de doutorado.

Em contrapartida, os Trabalhos de Conclusão compõem apenas 13,3% dos registros, o que sugere uma menor frequência de localização ou publicação desses trabalhos. Essa baixa representatividade pode estar relacionada à limitada maturidade acadêmica dos estudantes de graduação, bem como à escassez de tempo disponível para o desenvolvimento de pesquisas com maior complexidade, como é exigido nos níveis de pós-graduação.



**Figura 1.** Distribuição percentual das categorias de documentos indexados no Google Acadêmico relacionados ao uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

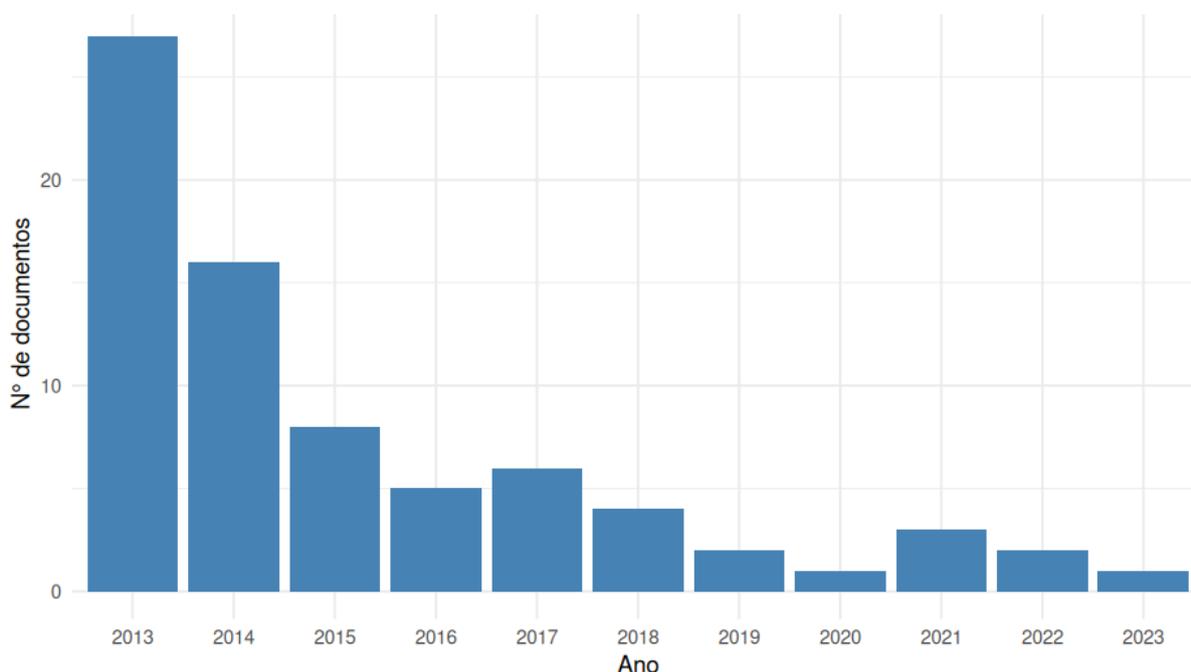
Entre os anos de 2013 e 2023, a área de ciência, tecnologia e inovação no Brasil foi fortemente impactada por restrições orçamentárias que afetaram diretamente o desenvolvimento da pesquisa acadêmica. Um dos setores mais comprometidos foi o da pós-

graduação, especialmente nos programas vinculados à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Os cortes nos principais mecanismos de fomento científico resultaram não apenas em menor capacidade operacional das instituições, mas também na descontinuidade de pesquisas relevantes em diversas áreas do conhecimento. A tendência de queda nos investimentos consolidou-se a partir de 2014 e foi agravada pela promulgação da Emenda Constitucional nº 95/2016, que impôs o teto de gastos públicos, restringindo o crescimento real do orçamento federal por duas décadas. Como evidenciam Ribeiro e Oliveira (2024, p. 36). Além do corte direto de recursos, houve uma estagnação nos valores das bolsas de mestrado e doutorado, que permaneceram congeladas por quase uma década. Segundo a Agência Brasil (2023), o primeiro reajuste significativo ocorreu apenas em 2023, após nove anos sem correção e, ainda assim, não compensou a defasagem inflacionária acumulada. Esse conjunto de fatores compromete a atratividade da carreira acadêmica, provocando evasão de pesquisadores, descontinuidade de linhas de pesquisa e uma retração visível na produção científica nacional.

A distribuição dos documentos, evidenciando a predominância de dissertações, seguidas por teses e, em menor proporção, trabalhos de conclusão de curso (TCCs), permite inferir os tipos de produções acadêmicas mais frequentemente digitalizadas e disponibilizadas em repositórios eletrônicos. Tal panorama também reflete o volume da produção científica nos distintos níveis da educação superior.

Na figura 2 observa-se a distribuição da produção acadêmica ao longo do período, considerando dissertações, teses e trabalhos de conclusão de curso. Nota-se uma concentração expressiva de documentos nos anos iniciais, especialmente em 2013, com o maior número de registros. A partir de 2014, observa-se uma redução gradual, com quedas mais acentuadas a partir de 2015, resultando em uma produção visivelmente menor nos anos subsequentes, especialmente entre 2019 e 2023.

### Distribuição temporal dos estudos

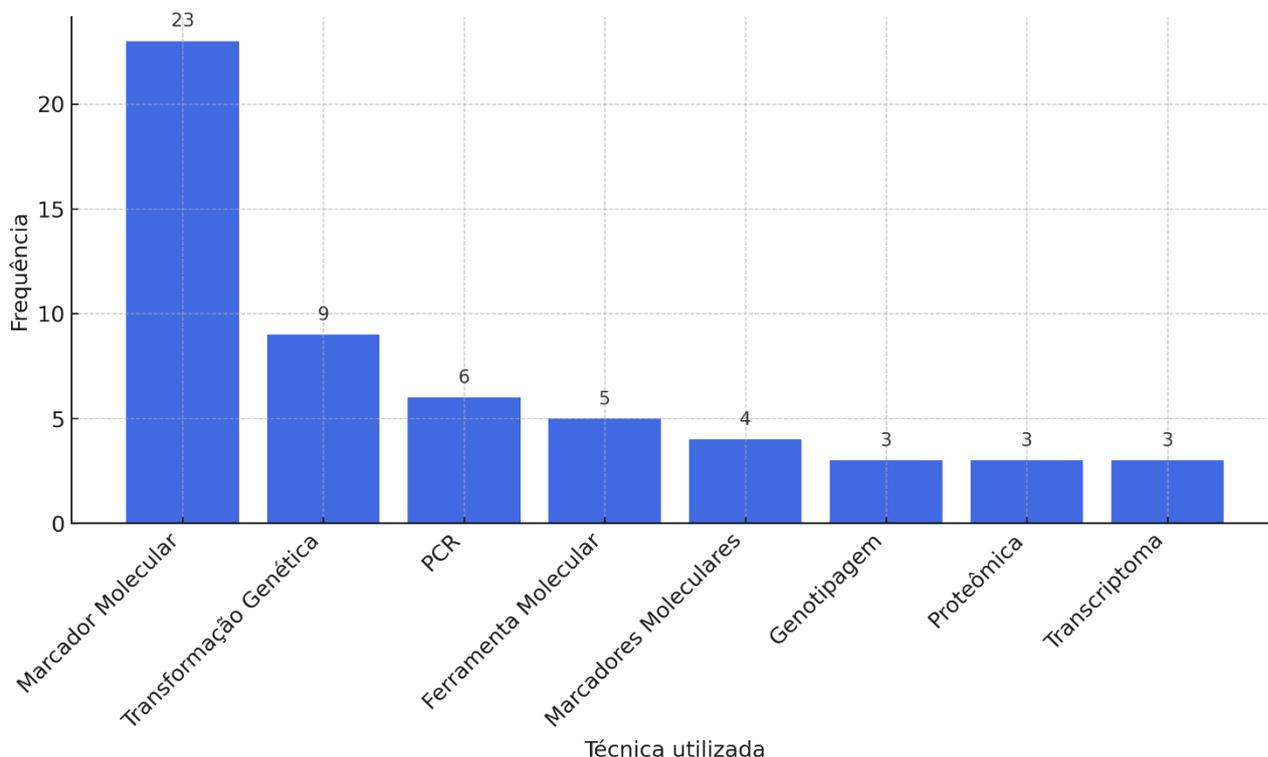


**Figura 2.** Evolução do número de estudos indexados no Google Acadêmico sobre o uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

Considera-se que essa tendência descendente está relacionada às restrições orçamentárias enfrentadas pelas instituições públicas de ensino e pesquisa no Brasil ao longo da última década. A partir de 2014, houve sucessivos cortes nos recursos destinados à ciência, tecnologia e inovação, comprometendo diretamente o financiamento de bolsas, projetos e infraestrutura científica. Além disso, o contexto político e econômico do país contribuiu para a instabilidade do setor, afetando o planejamento e a execução de estudos que exigem suporte técnico, financeiro e infraestrutura laboratorial, características comuns às pesquisas em melhoramento genético vegetal. Essa conjuntura ajuda a compreender a acentuada diminuição na produção ao longo dos anos analisados.

A partir da análise do conteúdo dos documentos, conforme mostrado anteriormente, foi possível identificar os tipos de ferramentas moleculares mais utilizadas nos experimentos descritos nos trabalhos indexados no período de 2013 a 2023 (Figura 3). A figura apresenta a distribuição da frequência de uso dessas técnicas, fornecendo uma visão quantitativa sobre sua recorrência nas pesquisas analisadas.

### Contagem de termos utilizados nos documentos indexados no Google Acadêmico

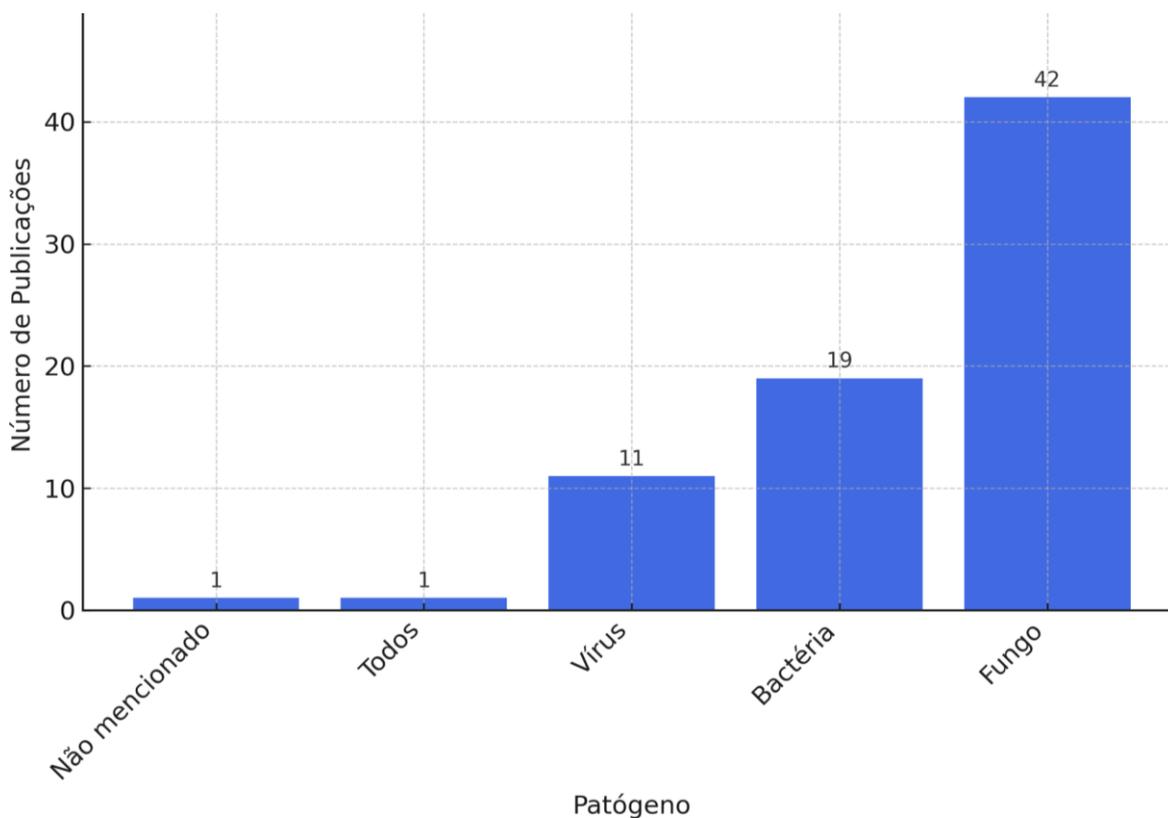


**Figura 3.** Frequência de termos relacionados a biotecnologia no melhoramento de plantas em trabalhos indexados no Google Acadêmico sobre o uso da biotecnologia no melhoramento de plantas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

A distribuição da frequência de uso dos termos representa uma visão quantitativa sobre quais técnicas são mais recorrentes nos estudos analisados. Conforme mostrado pela altura da barra correspondente, é o marcador molecular, que aparece em mais de 23 ocorrências, destacando-se como a abordagem mais utilizada. Isso sugere que os marcadores moleculares desempenham um papel fundamental na análise genética e molecular dos experimentos avaliados, por sua aplicabilidade em estudos de diversidade genética, mapeamento de genes e melhoramento genético. Outras técnicas, como transformação genética e Reação em Cadeia da Polimerase (PCR), também aparecem com frequência significativa, embora menor do que os marcadores moleculares, ferramentas amplamente utilizadas em manipulação genética e amplificação de sequências de DNA. Distribuição de publicações sobre grupos patogênicos em trabalhos acadêmicos que abordam o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos

A análise dos dados também permitiu identificar os patógenos mais frequentemente abordados nas pesquisas (Figura 4). Entre as categorias analisadas, os fungos destacam-se como o grupo patogênico mais recorrente, sendo objeto de estudo em 42 publicações.

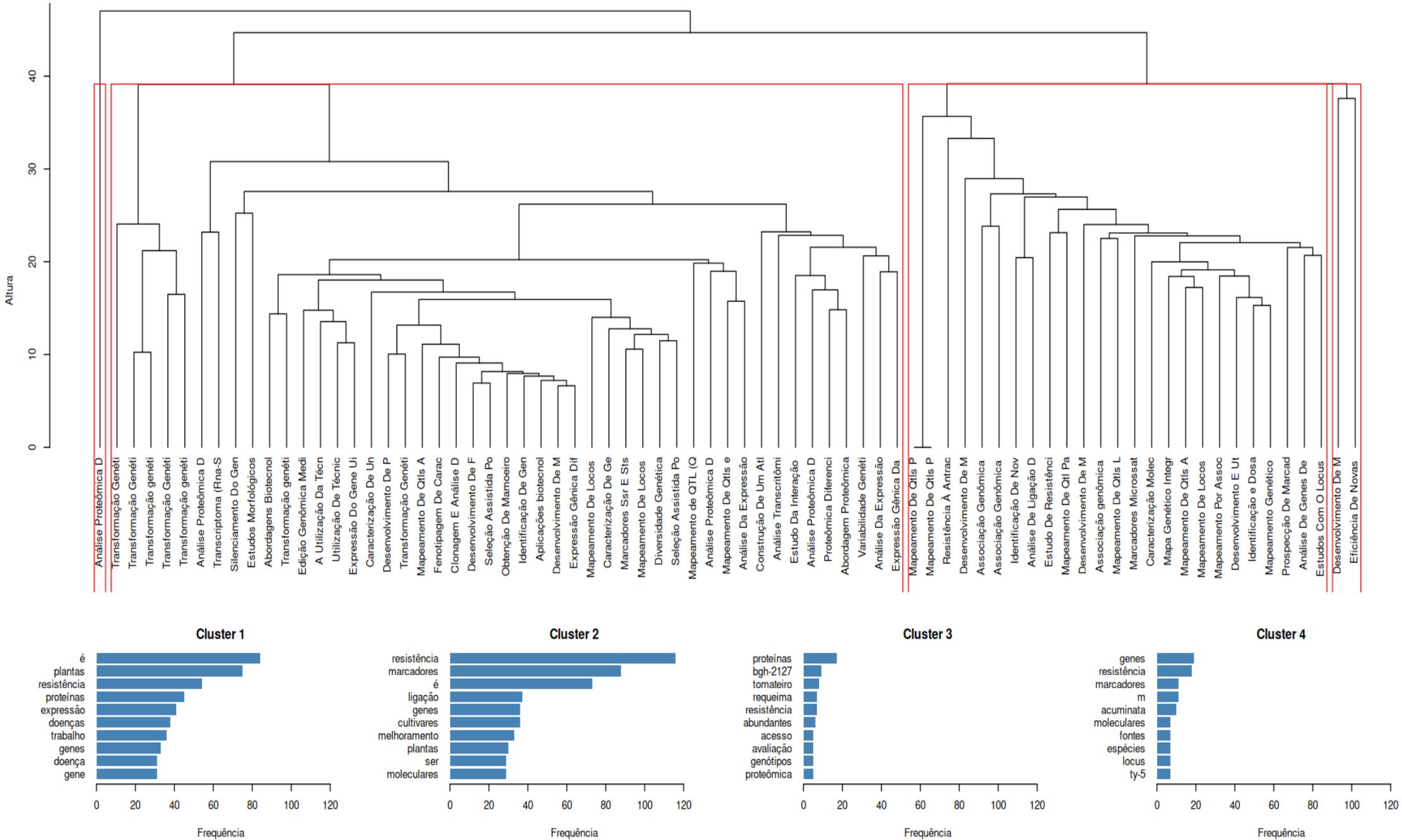
### Panorama das publicações sobre os grupos patogênicos



**Figura 4.** Distribuição de publicações sobre grupos patogênicos em trabalhos acadêmicos que abordam o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

Os dados indicam uma ênfase das pesquisas na resistência genética a fungos fitopatogênicos, os quais representam uma das principais causas de perdas na agricultura, especialmente em culturas de alto valor econômico.

Para a análise de agrupamento dos resumos (Figura 5), foi aplicado o método de agrupamento hierárquico, com a medida de distância euclidiana calculada sobre os vetores numéricos derivados da DFM. A técnica de ligação adotada foi o método de Ward (ward. D2), por sua eficácia em minimizar a variância intragrupo e produzir agrupamentos bem definidos. O dendrograma gerado refletiu a proximidade lexical entre os documentos e permitiu a extração de quatro clusters principais, definidos por meio da função *cutree()*. Para caracterizar semanticamente cada agrupamento, foi reconstruída uma DFM específica para os resumos de cada cluster, a partir da qual foram extraídos os dez termos mais frequentes, representados graficamente em barras dispostas junto ao dendrograma.



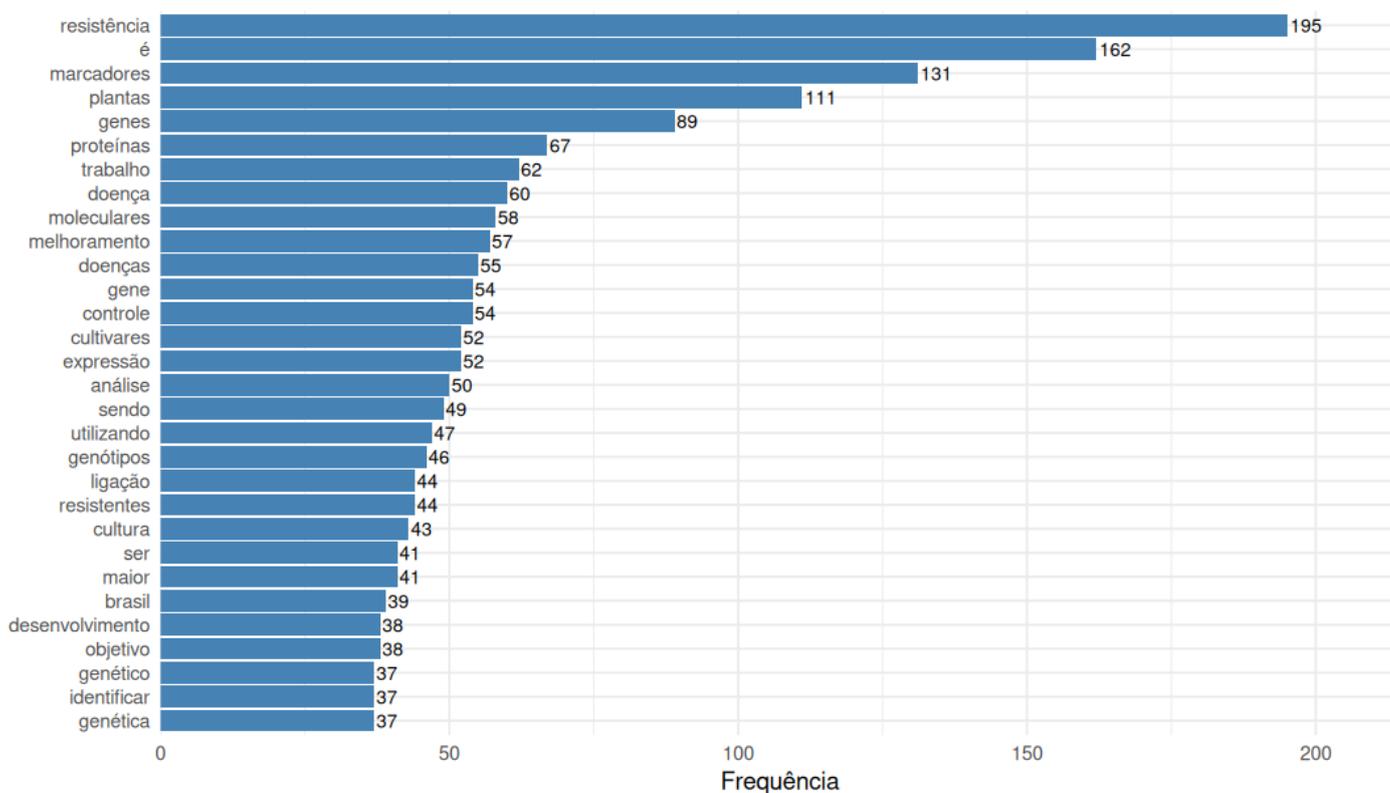
**Figura 5.** Agrupamento de termos recorrentes em trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

O resultado dessa análise permite a visualização da estrutura de agrupamentos formados em quatro principais (clusters), os quais revelam a organização conceitual predominante nos textos avaliados.

O Cluster 1 concentrou termos como “*plantas*”, “*resistência*”, “*proteínas*”, “*expressão*” e “*doenças*”, refletindo um conjunto de publicações voltadas à expressão gênica e aos mecanismos de defesa das plantas contra patógenos. No Cluster 2, os termos mais frequentes foram “*resistência*”, “*marcadores*”, “*genes*”, “*cultivares*” e “*melhoramento*”, indicando uma ênfase nos estudos relacionados ao uso de marcadores moleculares no processo de seleção de genótipos resistentes. O Cluster 3 apresenta palavras como “*proteínas*”, “*tomateiro*”, “*requeima*”, “*proteômica*” e “*acesso*”, apontando para pesquisas mais específicas, direcionadas à análise proteômica de culturas agrícolas específicas, com destaque para o tomateiro. Por fim, o Cluster 4 reuniu termos como “*genes*”, “*locus*”, “*espécies*”, “*moleculares*” e “*ty-5*”, representando estudos voltados à caracterização genética de espécies vegetais e à identificação de loci associados à resistência. Com o dendograma é possível identificar agrupamentos temáticos coesos e os principais focos da pesquisa, demonstrando-se eficiente na organização dos dados textuais para a compreensão das tendências conceituais e dos eixos temáticos predominantes nos estudos.

Com o intuito de identificar os conceitos centrais da literatura analisada (Figura 6), foi realizada uma análise de frequência textual. Essa filtragem tem por base quantificar os termos mais recorrentes, permitindo observar os eixos temáticos predominantes e os principais focos de investigação presentes na produção científica relacionada. A Figura apresenta os 30 termos mais frequentes. Os termos com maiores ocorrências “*resistência*”, com 195 menções, seguido de “*marcadores*” (131), “*plantas*” (111) e “*genes*” (89). Embora o termo “*é*” apareça com alta frequência, de acordo com Manning, Raghavan e Schütze (2008), as *stopwords* correspondem a palavras funcionais que ocorrem com alta frequência em um corpus textual, mas não possuem relevância semântica para a análise de conteúdo, sendo tratadas como ruído e geralmente removidas em processos de mineração de texto.

### Frequência de termos em publicações sobre biotecnologia no melhoramento de plantas com foco em doenças na última década

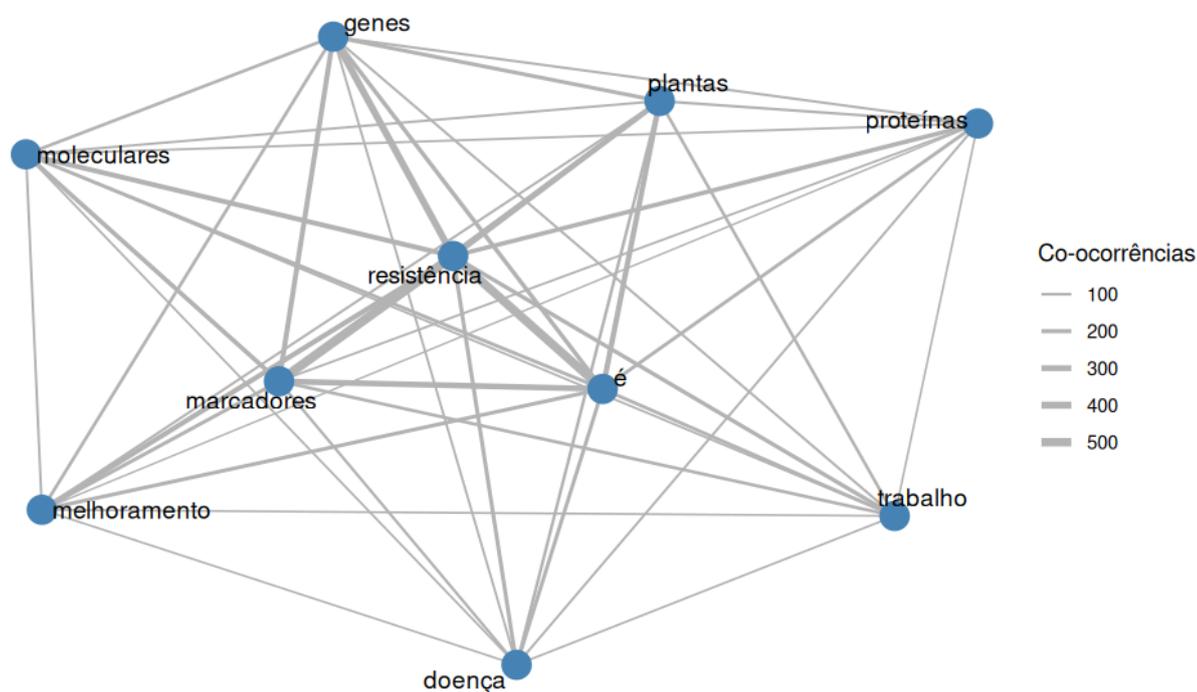


**Figura 6.** Frequência dos principais termos encontrados em documentos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

Outros termos de destaque como “plantas”, “genes”, “proteínas”, “doença” e “melhoramento” reforçam a centralidade de temas ligados ao melhoramento genético, à biotecnologia molecular e à interação planta-patógeno no conjunto de publicações analisadas. A presença dos termos “trabalho”, “controle”, “expressão”, “genótipos” e “cultivares” também indica a recorrência de estudos voltados à caracterização genética, funcional e agrônômica das espécies.

Para a construção da rede de co-ocorrência de termos (Figura 7), foram selecionados os dez vocábulos mais frequentes a partir da Matriz Documento-Termo (DFM), previamente processada com o pacote *quanteda*. A partir disso, foi gerada uma matriz de co-ocorrência por meio da função *fcm()*, e a estrutura resultante foi convertida em objeto da classe *igraph*, no qual os nós representam os termos e as arestas indicam o peso da relação entre eles. A visualização do grafo foi realizada com o pacote *ggraph*, utilizando o layout de Fruchterman–Reingold, que permite uma representação espacial balanceada das conexões, com espessura das arestas proporcional à força de co-ocorrência.

### Rede semântica de termos recorrentes em publicações sobre o uso da biotecnologia aplicada ao melhoramento vegetal

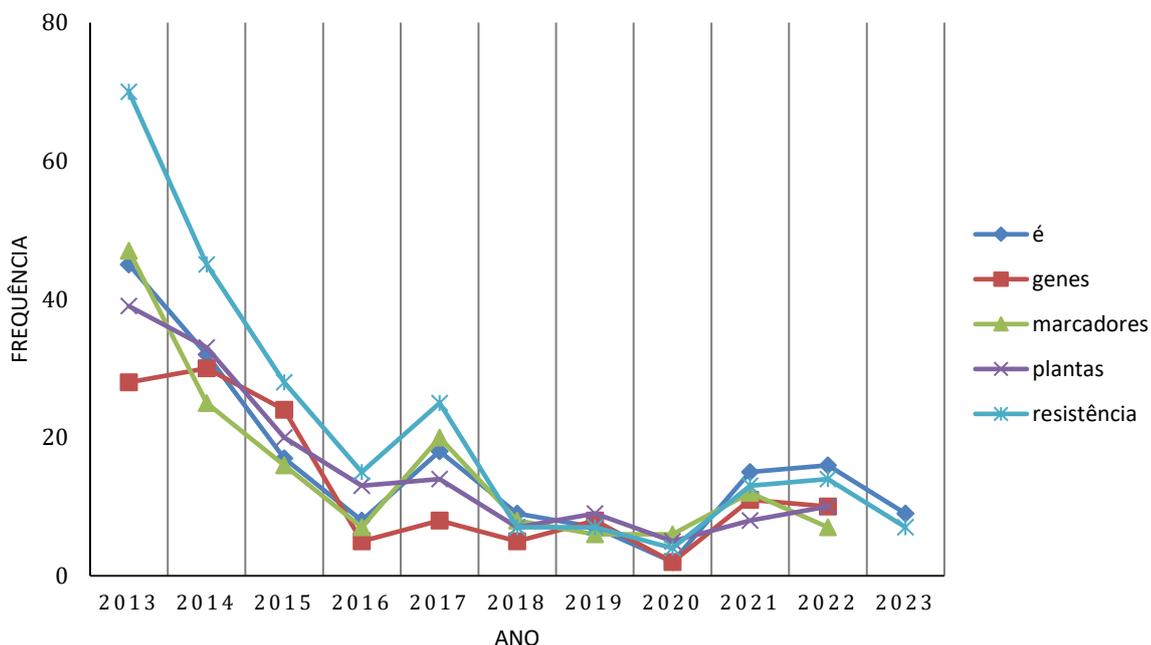


**Figura 7.** Rede de co-ocorrência entre os principais termos presentes em trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

Observa-se que o termo “resistência” ocupa posição central na rede, com conexões intensas com “marcadores”, “genes”, “plantas”, “doença” e “moleculares”. Essa configuração confirma a relevância da resistência genética como tema principal da produção científica analisada. Outros termos, como “melhoramento”, “proteínas” e “trabalho”, também apresentam conexões significativas, indicando a integração de abordagens moleculares com estratégias de melhoramento genético. A presença do termo “é”, embora com alta conectividade, representa um elemento linguístico sem carga temática.

Com o objetivo de analisar a dinâmica temporal desta produção científica ao longo da última década (Figura 8), analisamos a frequência anual dos cinco termos mais recorrentes nos resumos dos documentos. Essa abordagem permite observar o uso de conceitos ao longo do tempo, refletindo mudanças de foco temático, avanços metodológicos e contextos externos que influenciaram a atividade científica.

### Distribuição temporal de frequência de termos



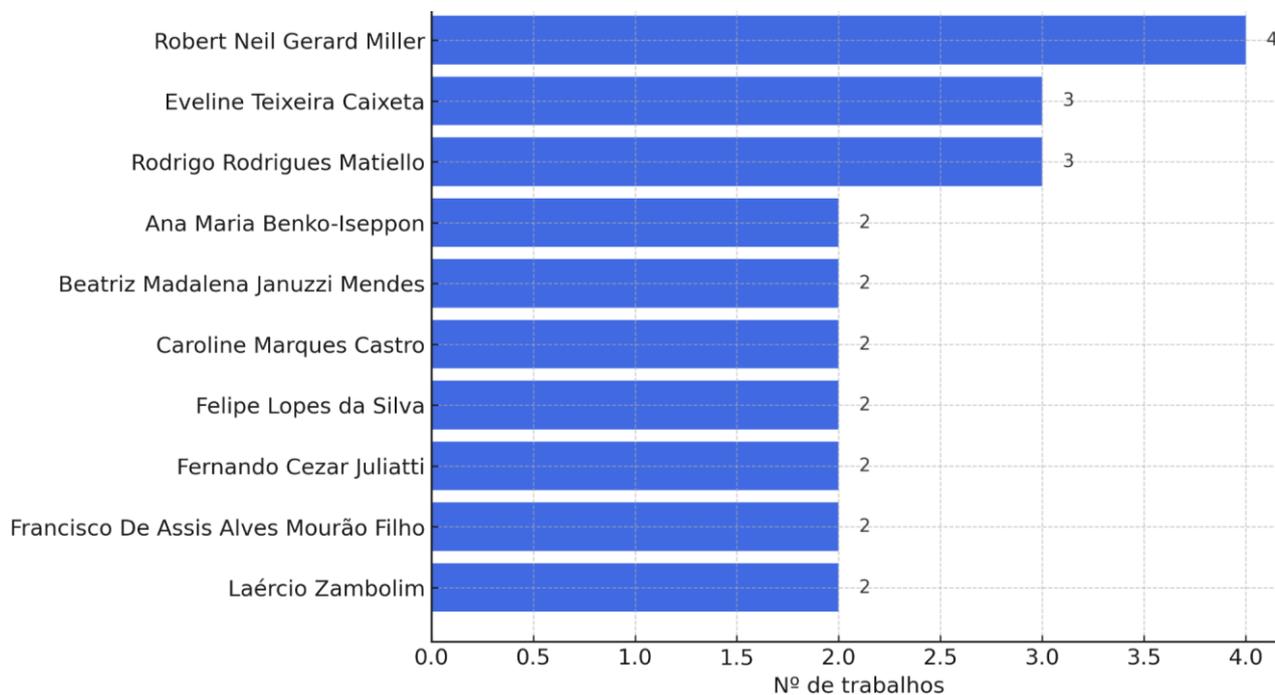
**Figura 8.** Evolução temporal da frequência de termos utilizados em resumos de trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

A evolução anual da frequência dos termos “resistência”, “plantas”, “marcadores”, “genes” entre os anos, observa-se que o termo “resistência” apresentou a maior frequência no início da série, com pico em 2013, seguido de uma redução significativa nos anos posteriores. Um comportamento semelhante é verificado para os demais termos, especialmente “plantas” e “marcadores”, cuja frequência diminui gradualmente até estabilizar em patamares mais baixos a partir de 2019.

Os dados apresentados permitiram uma interpretação quantitativa e visual da ocorrência, associação e evolução dos termos mais frequentes na literatura científica analisada. A análise de frequência destacou os principais conceitos abordados nos resumos, enquanto a rede de co-ocorrência evidenciou as relações semânticas entre os termos, apontando núcleos temáticos consolidados. O dendrograma complementou essas informações ao organizar os termos em agrupamentos hierárquicos, reforçando a estrutura conceitual da área. A evolução temporal demonstrou a variação no uso desses termos ao longo dos anos, revelando padrões de queda na produção e na ênfase de determinados temas.

A análise final (Figura 9) tem como objetivo identificar os principais autores que mais contribuíram com publicações relacionadas ao tema central deste estudo.

### Análise Quantitativa da Produção Acadêmica por Orientador



**Figura 9.** Orientadores com maior número de trabalhos acadêmicos sobre o uso da biotecnologia no melhoramento de plantas cultivadas com foco em doenças nos últimos 10 anos. **Fonte:** Elaborado pelo autor (2025).

A análise da autoria dos trabalhos também incluiu a identificação dos principais orientadores responsáveis pelas produções acadêmicas presentes no conjunto de documentos avaliados. Essa etapa tem por finalidade compreender quem são os pesquisadores que mais contribuíram, como orientadores, para o desenvolvimento de estudos relacionados ao tema abordado neste trabalho.

Observa-se que Robert Neil Gerard Miller se destaca com quatro trabalhos, o que sugere sua significativa atuação e recorrência na área de estudo. Em seguida, aparecem Rodrigo Rodrigues Matiello e Eveline Teixeira Caixeta, ambos com três publicações, evidenciando também sua relevância no cenário científico avaliado. A lista segue com um conjunto de autores que possuem duas publicações cada, entre eles Laércio Zambolim, Francisco De Assis Alves Mourão Filho e Ana Maria Benko-Iseppon, o que demonstra uma diversidade de pesquisadores contribuindo com investigações relevantes.

#### 4 CONSIDERAÇÕES FINAIS

A análise exploratória realizada permitiu identificar tendências, ferramentas e abordagens recorrentes na literatura acadêmica relacionada à biotecnologia aplicada ao melhoramento genético de plantas com foco na resistência a doenças. Observou-se uma predominância de dissertações (60%), o que evidencia a maior representatividade dos programas de mestrado na produção científica sobre o tema. Essa distribuição sugere que, mesmo diante das adversidades enfrentadas pelo sistema de pós-graduação brasileiro, especialmente nos últimos anos, ainda se mantém uma produção significativa em nível de mestrado, refletindo o esforço contínuo das instituições acadêmicas.

A produção acadêmica no Brasil apresentou queda a partir de 2014, agravada entre 2019 e 2023, devido a cortes no financiamento da ciência e tecnologia. Essa retração está ligada à Emenda Constitucional nº 95/2016, que limitou os gastos públicos e reduziu os recursos para agências como CAPES e CNPq. Como consequência, houve evasão de talentos, desmotivação de jovens cientistas e impacto negativo na capacidade de inovação do país, afetando setores estratégicos como o agronegócio. A análise das ferramentas moleculares utilizadas demonstrou que os marcadores moleculares continuam sendo amplamente empregados, seguidos por técnicas como a PCR e a transformação genética. Além disso, o destaque dado aos fungos como principais patógenos reforça a relevância da resistência genética como estratégia de controle fitossanitário eficiente e sustentável.

As abordagens computacionais aplicadas como análise de co-ocorrência, agrupamento hierárquico e análise temporal de termos possibilitaram uma leitura abrangente do panorama temático e evolutivo da produção científica. Dessa forma, os resultados deste estudo não apenas contribuem para a compreensão das dinâmicas da produção acadêmica na interface entre biotecnologia e resistência genética, como também oferecem subsídios relevantes para a formulação de políticas públicas, a definição de prioridades de financiamento e o fortalecimento da ciência nacional em áreas estratégicas, como a Agronomia.

É importante, contudo, reconhecer as limitações deste trabalho. A principal delas reside na adoção exclusiva do Google Acadêmico como fonte de coleta dos documentos. Embora essa plataforma ofereça ampla cobertura, sua limitação em filtros avançados e a ausência de padronização na indexação podem ter ocasionado a omissão de estudos relevantes presentes em bases de dados especializadas, como Scopus, Web of Science e repositórios institucionais. Além disso, o recorte metodológico privilegiou documentos acadêmicos formais como

dissertações, teses e TCCs o que restringe a abrangência da análise, não refletindo o volume total de pesquisas desenvolvidas na área.

Com base nos resultados obtidos, este estudo abre caminho para diversas possibilidades de aprofundamento. Pesquisas futuras podem ampliar o recorte temporal, considerando um intervalo de 20 a 50 anos, a fim de construir uma linha evolutiva mais completa da aplicação da biotecnologia ao melhoramento genético de plantas. Outra vertente relevante seria a expansão da análise para uma perspectiva internacional, permitindo comparar a produção científica brasileira com a de outros países. A construção de uma árvore cronológica destacando os marcos teóricos, metodológicos e tecnológicos também enriqueceria o entendimento do campo, ao evidenciar a adoção e o abandono de técnicas ao longo do tempo. Além disso, estudos que mapeiam os principais pesquisadores atuantes na área e analisam os impactos da migração de cientistas brasileiros para o exterior, frequentemente motivada pela desvalorização da ciência no país, podem contribuir para um diagnóstico mais profundo sobre os desafios e oportunidades da biotecnologia no contexto nacional.

## REFERÊNCIAS

AGÊNCIA BRASIL. **Bolsas de pesquisa terão reajuste de 25% a 200%**. 2023. Disponível em: <https://agenciabrasil.ebc.com.br/educacao/noticia/2023-02/bolsas-de-pesquisa-terao-reajuste-de-25-200>. Acesso em: 12 maio 2025.

AGÊNCIA GOV. **Execução orçamentária da CAPES é a maior dos últimos 7 anos**. 2024. Disponível em: <https://agenciagov.ebc.com.br/noticias/202404/execucao-orcamentaria-da-capes-e-a-maior-dos-ultimos-7-anos>. Acesso em: 12 maio 2025.

AGÊNCIA GOV. **Corte de orçamento da Capes prejudica pós-graduação no Brasil**. Brasília: Agência Gov, 2024. Disponível em: <https://www.gov.br>. Acesso em: 25 jun. 2025.

ALVARADO, D. A.; ARIAS, H. D. **Introducción a la cienciometría: fundamentos y aplicaciones**. Bogotá: Universidad Nacional de Colombia, 2017.

ARAÚJO, R. F.; ALVARENGA, L. **A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007**. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 16, n. 31, p. 51–70, 2011. Disponível em: <https://doi.org/10.5007/1518-2924.2011v16n31p51>. Acesso em: 6 nov. 2024.

ARIA, Massimo; CUCCURULLO, Corrado. **Bibliometrix: an R-tool for comprehensive science mapping analysis**. *Journal of Informetrics*, v. 11, n. 4, p. 959–975, 2017. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>.

BARRETO, A. L.; FILGUEIRAS, C. A. L. **Origens da universidade brasileira**. Química Nova, São Paulo, v. 30, n. 7, p. 1780–1790, 2007.

BENOIT, K. et al. **quanteda: quantitative analysis of textual data**. R package, version 3.3.1, 2024. Disponível em: <https://quanteda.io>. Acesso em: 27 maio 2025.

BORGES, M. N. **Ciência, tecnologia e inovação para o desenvolvimento do Brasil**. Scientia Plena, São Cristóvão, v. 12, n. 8, p. 1–11, 2016.

BRASIL. Constituição (1988). **Emenda Constitucional nº 95, de 15 de dezembro de 2016. Altera o Ato das Disposições Constitucionais Transitórias para instituir o Novo Regime Fiscal**. Diário Oficial da União: seção 1, Brasília, DF, 16 dez. 2016.

CABRAL, Thiago Luiz de Oliveira et al. **A CAPES e suas sete décadas: trajetória da pós-graduação stricto sensu no Brasil**. Revista Brasileira de Pós-Graduação, Brasília, v. 16, n. 36, p. 1-20, out. 2020. Disponível em: <https://rbpg.capes.gov.br/index.php/rbpg/article/view/534>. Acesso em: 28 jun. 2025.

CAMARGO, L. S. de; BARBOSA, R. R. **Bibliometria, cienciometria e um possível caminho para a construção de indicadores e mapas da produção científica**. Ponto de Acesso, Salvador, v. 12, n. 3, p. 109–125, dez. 2018. Disponível em: <https://www.pontodeacesso.ici.ufba.br>.

CARRER, H.; BARBOSA, A. L. RAMIRO, D. A. **Biotecnologia na agricultura**. Estudos Avançados, v. 24, n. 70, p. 149–164, 2010. Acesso em: 6 nov. 2024.

CARVALHO, O. M. F. W.; HENNING, E.; KONRATH, A. C.; ALVES, C. C.; SAMOBYL, R. W. **Uma visão geral do RStudio aplicado ao ensino de controle estatístico do processo**. In: CONGRESSO BRASILEIRO DE EDUCAÇÃO EM ENGENHARIA – COBENGE, 2016, [s.l.]. Anais [...]. [S.l.]: ABENGE, 2016. Disponível em: <https://www.abenge.org.br/cobenge/legado/arquivos/7/artigos/104106.pdf>. Acesso em: 6 nov. 2024.

CSÁRDI, Gábor; NEPUSZ, Tamás. **The igraph software package for complex network research**. *InterJournal*, Complex Systems, v. 1695, 2006. Disponível em: <https://igraph.org>. Acesso em: 25 jun. 2025.

DAS, P. R. et al. **Biotecnologia para sustentabilidade no manejo de doenças de plantas: um enfoque bibliométrico**. *Journal of Agricultural Research & Development*, v. 8, n. 1, p. 110–125, 2023.

FAPESP. **Produção científica brasileira cai pelo segundo ano consecutivo**. Agência FAPESP, 2024. Disponível em: <https://agencia.fapesp.br/producao-cientifica-brasileira-cai-pelo-segundo-ano-consecutivo/52363>. Acesso em: 12 maio 2025.

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo. **Boletim de Indicadores de Ciência, Tecnologia e Inovação**. São Paulo: FAPESP, 2024. Disponível em: <https://www.fapesp.br>. Acesso em: 25 jun. 2025.

FURTADO, A. T. **Novos arranjos produtivos, Estado e gestão da pesquisa pública**. *Ciência e Cultura*, São Paulo, v. 57, n. 1, p. 41–45, jan./mar. 2005.

GUEDES, Vânia L. da Rocha; BORSCHIVER, Suzana. **Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica**. *Encontro Nacional de Ciência da Informação*, v. 6, n. 1, p. 1-18, 2005.

GUIMARÃES, R. **Pesquisa no Brasil: a reforma tardia**. São Paulo em Perspectiva, São Paulo, v. 16, n. 4, p. 41–47, out./dez. 2002.

HAIR JR., J. F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

JUDIANTO, A.; AULIANI, A. **Mapeamento de tendências em biotecnologia agrícola: uma revisão bibliométrica**. *Revista Brasileira de Biotecnologia & Agronomia*, v. 15, n. 2, p. 45–62, 2024.

LE COADIC, Y.-F. **A ciência da informação**. 2. ed. Brasília, DF: Briquet de Lemos/Livros, 2003.

LEITE, F. A. et al. **Mineração de textos: conceitos, técnicas, ferramentas e aplicações**. *Revista de Administração Pública*, v. 51, n. 3, p. 389–410, 2017.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2008.

MORATTO, H. B.; MORAIS, A. L. de; NUNES, R. L. **Análise de co-ocorrência de palavras: contribuições metodológicas e aplicação ao estudo da inovação.** Revista Digital de Biblioteconomia e Ciência da Informação, v. 16, n. 2, 2018.

MOURA, Egberto Gaspar de; CAMARGO JUNIOR, Kenneth Rochel de. **A crise no financiamento da pesquisa e pós-graduação no Brasil.** Cadernos de Saúde Pública, Rio de Janeiro, v. 33, n. 4, e00052917, 2017. Disponível em: <https://doi.org/10.1590/0102-311X00052917>. Acesso em: 28 jun. 2025.

OLIVEIRA, A. C.; DÓREA, J. G.; DOMENE, S. M. A. **Bibliometria na avaliação da produção científica da área de nutrição registrada no Cibran: período de 1984–1989.** Ciência da Informação, Brasília, v. 21, n. 3, p. 239–242, set./dez. 1992.

OTTOBONI, C. **Capacidade para inovar de indústrias eletroeletrônicas: estudo de múltiplos casos no Vale da Eletrônica em Minas Gerais.** 2011. 386 f. Tese (Doutorado em Administração) – Universidade Federal de Lavras, Lavras, 2011.

PEDERSEN, T. L. **ggraph: an implementation of grammar of graphics for graphs and networks.** R package, version 2.1.0, 2024. Disponível em: <https://ggraph.data-imaginist.com>. Acesso em: 27 maio 2025.

R CORE TEAM. **R: A language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing, 2024. Disponível em: <https://www.r-project.org/>. Acesso em: 27 maio 2025.

RAFAEL, R. M. R. et al. **A análise de conteúdo e a análise textual quantitativa: aproximações metodológicas.** Revista Enfermagem UERJ, v. 24, n. 1, 2016.

RAZERA, J. C. C. **Contribuições da cienciométrica para a área brasileira de Educação em Ciências.** Ciência & Educação, v. 22, n. 3, p. 557–560, 2016. Disponível em: <https://doi.org/10.1590/1516-731320160030001>. Acesso em: 6 nov. 2024.

RIBEIRO, D. B.; OLIVEIRA, E. F. dos A. **A distribuição de bolsas da CAPES em tempos de cortes orçamentários.** Temporalis, Brasília (DF), v. 24, n. 47, p. 35–50, jan./jun. 2024. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/10089399.pdf>. Acesso em: 12 maio 2025.

SANTOS, R. N. M.; KOBASHI, N. Y. **Bibliometria, cientometria, informetria: conceitos e aplicações.** Tendências da Pesquisa Brasileira em Ciência da Informação, Brasília, v. 2, n. 1, 2009. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/21/43>. Acesso em: 6 nov. 2024.

SBPC; ABC. **Carta conjunta da SBPC e ABC sobre a situação da ciência no Brasil.** São Paulo: SBPC, 2021. Disponível em: <https://www.sbpcnet.org.br/site/noticias/carta-aberta-da-sbpc-e-abc-2021/>. Acesso em: 28 jun. 2025.

SCHWARTZMAN, Simon. **Pesquisa e pós-graduação no Brasil: duas faces da mesma moeda?** Estudos Avançados, São Paulo, v. 36, n. 104, p. 225–254, 2022. Disponível em: <https://doi.org/10.1590/s0103-4014.2022.36104.011>. Acesso em: 28 jun. 2025.

SILGE, J.; ROBINSON, D. **Tidytext: text mining using 'dplyr', 'ggplot2', and other tidy tools**. R package, version 0.4.2, 2024. Disponível em: <https://www.tidytextmining.com>. Acesso em: 27 maio 2025.

SINCLAIR, Stéfan; ROCKWELL, Geoffrey. **Voyant Tools: web-based reading and analysis environment for digital texts**. Version 2.4, 2016. Disponível em: <https://voyant-tools.org/>. Acesso em: 25 jun. 2025.

SPINAK, E. **Indicadores cienciométricos**. Ciência da Informação, Brasília, v. 27, n. 2, p. 141–148, maio/ago. 1998.

TARGINO, Maria das Graças. **Comunicação científica: uma revisão de seus elementos básicos**. Informação & Sociedade: Estudos, João Pessoa, v. 10, n. 2, p. 37–85, 2000.

VAN ECK, Nees Jan; WALTMAN, Ludo. **Software survey: VOSviewer, a computer program for bibliometric mapping**. *Scientometrics*, v. 84, n. 2, p. 523–538, 2010. DOI: <https://doi.org/10.1007/s11192-009-0146-3>.

VANTI, N. A. **Bibliometria e cientometria: contribuições para a avaliação da produção científica**. Ciência da Informação, v. 31, n. 2, p. 152–162, 2002.

VANTI, N. A. P. **Da Bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento**. Ciência da Informação, v. 31, n. 2, p. 369–379, 2002. Disponível em: <https://doi.org/10.1590/s0100-19652002000200016>. Acesso em: 6 nov. 2024.

WANG, F.; WANG, C.; LIU, P.; LEI, C.; HAO, W.; GAO, Y.; ZHAO, K. **Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene**. *BMC Plant Biology*, v. 19, n. 1, p. 1–8, 2019. Disponível em: <https://doi.org/10.1186/s12870-019-1890-7>. Acesso em: 6 nov. 2024.

WICKHAM, H. **ggplot2: elegant graphics for data analysis**. 2. ed. Nova York: Springer, 2024. Disponível em: <https://ggplot2.tidyverse.org>. Acesso em: 27 maio 2025.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. **dplyr: a grammar of data manipulation**. R package, version 1.1.4, 2024. Disponível em: <https://dplyr.tidyverse.org>. Acesso em: 27 maio 2025.

## APÊNDICE

Análise de dados

#

=====

====

# 0. Instalar e carregar pacotes necessários -----

```
pkgs <- c(
  "synthesisr",    # leitura de .RIS em data.frame
  "dplyr",         # manipulação de dados
  "tidyr",         # separate_rows()
  "stringr",      # operações em strings
  "ggplot2",      # gráficos
  "quanteda",     # corpus, tokens, DFM
  "quanteda.textstats", # cálculo de distâncias
  "tidytext",     # unnest_tokens()
  "igraph",       # grafos
  "ggraph",       # visualização de redes
  "tidygraph"     # interface tidy para igraph
)

new_pkgs <- setdiff(pkgs, installed.packages()[, "Package"])
if (length(new_pkgs)) install.packages(new_pkgs)
lapply(pkgs, library, character.only = TRUE)
```

# 1. Importar e preparar os dados -----

```

# 1.1 Importa o arquivo RIS para um data.frame
df_raw <- read_refs("dados.ris")

# 1.2 Ajuste de nomes de colunas, se necessário
# Assumimos que existem: title, year, abstract, author ("Orientando and Orientador")
df <- df_raw

# 1.3 Limpeza básica e remoção de duplicatas
df_clean <- df %>%
  filter(!is.na(title), !is.na(year)) %>% # mantêm só registros com título e ano
  distinct(title, year, .keep_all = TRUE) # elimina duplicatas exatas por título+ano

# 2. Estatísticas descritivas -----
# 2.1 Produção anual de trabalhos
df_clean %>%
  count(year) %>%
  ggplot(aes(x = year, y = n)) +
  geom_col(fill = "steelblue") +
  labs(title = "Produção por ano",
        x = "Ano", y = "Número de documentos") +
  theme_minimal()

# 2.2 Top 10 orientadores (segundo autor)
advisor_df <- df_clean %>%
  mutate(
    # separa "Orientando and Orientador" e extrai o segundo

```

```

auth_list = str_split(author, " and "),
advisor = sapply(auth_list, function(x) if (length(x)>=2) str_trim(x[2]) else NA)
) %>%
filter(!is.na(advisor), advisor != "")

```

```

advisor_df %>%
count(advisor, sort = TRUE) %>%
slice_head(n = 10) %>%
ggplot(aes(x = reorder(advisor, n), y = n)) +
geom_col(fill = "darkgreen") +
coord_flip() +
geom_text(aes(label = n), hjust = -0.1, size = 3) +
labs(title = "Top 10 orientadores",
      x = NULL, y = "Número de trabalhos") +
theme_minimal()

```

# 3. Mineração de texto em resumos (abstracts) -----

# 3.1 Criar corpus e tokens, removendo pontuação, números e stopwords pt

```

corp <- corpus(df_clean, text_field = "abstract")
toks <- tokens(corp, remove_punct = TRUE, remove_numbers = TRUE) %>%
tokens_remove(pattern = stopwords("pt"))

```

# 3.2 Montar DFM e aparar termos muito raros

```

dfm_mat <- dfm(toks) %>%
dfm_trim(min_termfreq = 5)

```

```
# 3.3 Gráfico dos 30 termos mais frequentes
```

```
freq_vec <- topfeatures(dfm_mat, 30)
```

```
freq_df <- data.frame(term = names(freq_vec), freq = as.numeric(freq_vec))
```

```
ggplot(freq_df %>% arrange(freq),
```

```
  aes(x = freq, y = factor(term, levels = term))) +
```

```
  geom_col(fill = "steelblue") +
```

```
  geom_text(aes(label = freq), hjust = -0.1, size = 3) +
```

```
  scale_x_continuous(expand = expansion(mult = c(0, 0.1))) +
```

```
  labs(title = "Top 30 termos mais frequentes em resumos",
```

```
        x = "Frequência", y = NULL) +
```

```
  theme_minimal()
```

```
# 4. Rede de co-ocorrência de termos (Top 10) -----
```

```
# 4.1 Seleciona os 10 termos mais frequentes
```

```
top10 <- names(topfeatures(dfm_mat, 10))
```

```
# 4.2 Calcula matriz de co-ocorrência e converte para igraph
```

```
fcmat <- fcm(dfm_mat)
```

```
fcm_top10 <- fcm_select(fcmat, pattern = top10)
```

```
mat10 <- as.matrix(fcm_top10)
```

```
g_terms <- graph_from_adjacency_matrix(mat10, mode = "undirected", weighted =
TRUE, diag = FALSE)
```

```
# 4.3 Plota a rede com legenda de largura
```

```
set.seed(2025)
```

```

ggraph(g_terms, layout = "fr") +
  geom_edge_link(aes(width = weight), color = "grey70") +
  geom_node_point(size = 6, color = "steelblue") +
  geom_node_text(aes(label = name), repel = TRUE, size = 4) +
  scale_edge_width(name = "Co-ocorrências", range = c(0.4, 2)) +
  guides(edge_width = guide_legend()) +
  labs(title = "Rede de co-ocorrência: Top 10 termos") +
  theme_void()

```

# 5. Evolução anual dos 5 termos mais frequentes -----

# 5.1 Definir top 5 e preparar tidytext

```
top5 <- names(topfeatures(dfm_mat, 5))
```

```

term_trends <- df_clean %>%
  mutate(year = as.numeric(year),
         text = tolower(abstract)) %>%
  select(year, text) %>%
  unnest_tokens(word, text) %>%
  filter(word %in% top5) %>%
  count(year, word)

```

# 5.2 Plotar linhas de tendência

```

ggplot(term_trends, aes(x = year, y = n, color = word)) +
  geom_line(size = 1) +
  geom_point() +
  scale_x_continuous(breaks = sort(unique(term_trends$year))) +

```

```
labs(title = "Evolução anual dos 5 termos mais frequentes",
      x = "Ano", y = "Frequência", color = "Termo") +
theme_minimal()
```

```
# 6. Clustering de abstracts e visualização combinada -----
```

```
# 6.1 Distância e agrupamento hierárquico (Ward.D2)
```

```
dist_mat <- textstat_dist(dfm_mat, method = "euclidean")
```

```
hc <- hclust(as.dist(dist_mat), method = "ward.D2")
```

```
k <- 4
```

```
clusters <- cutree(hc, k)
```

```
# 6.2 Extrair Top-10 termos por grupo
```

```
cluster_terms <- lapply(seq_len(k), function(grp) {
```

```
  docs <- df_clean$abstract[clusters == grp]
```

```
  toks_grp <- tokens(docs, remove_punct = TRUE, remove_numbers = TRUE) %>%
```

```
    tokens_remove(pattern = stopwords("pt"))
```

```
  topfeatures(dfm(dfm_trim(dfm(toks_grp), min_termfreq = 1)), 10)
```

```
})
```

```
# 6.3 Layout e plotagem: dendrograma + barras absolutas
```

```
layout(matrix(c(rep(1, k), 2:(k+1)), nrow = 2, byrow = TRUE),
```

```
        heights = c(3, 1))
```

```
# Dendrograma com rótulos curtos e retângulos vermelhos
```

```
par(mar = c(2, 4, 2, 1))
```

```
labels20 <- substr(df_clean$title, 1, 20)
```

```
plot(hc, labels = labels20, hang = -1,
     main = "Dendrograma de similaridade (abstracts)", xlab = "", ylab = "Altura")
rect.hclust(hc, k = k, border = "red")

# Barras de frequência absolutas, escala ajustada ao maior valor
max_val <- max(unlist(cluster_terms))
for (i in seq_len(k)) {
  par(mar = c(4, 8, 2, 1))
  barplot(rev(cluster_terms[[i]]),
          horiz = TRUE,
          las = 1,
          main = paste("Grupo", i),
          xlab = "Frequência",
          xlim = c(0, max_val * 1.1),
          col = "steelblue",
          border = NA)
}
```