

A REGULAÇÃO DA EXPRESSÃO GÊNICA PELOS miRNAs EM PLANTAS: BANCOS DE DADOS E MÉTODOS IN SILICO

THE REGULATION OF GENE EXPRESSION BY miRNAs IN PLANTS: DATABASES AND IN SILICO METHODS

João Victor da Paixão Reis Soares Dias¹

Cláudia Araújo Bastos²

Resumo: Os miRNAs são um tipo de molécula reguladora que permite (ou não) a tradução do mRNA (RNA mensageiro), impedindo a codificação de determinada proteína. Com tamanho entre 18 e 25 nucleotídeos, atuam como importantes reguladores da expressão gênica em plantas, animais e vírus. Em vegetais, os miRNAs desempenham papéis cruciais em processos como morfogênese, indução floral, desenvolvimento reprodutivo e resposta a estresses bióticos e abióticos. A identificação e a caracterização dessas moléculas por métodos experimentais apresentam elevados custos e demandam tempo considerável, o que torna as abordagens computacionais ferramentas essenciais para a redução do espaço de hipóteses a serem validadas experimentalmente. O presente trabalho teve como objetivo levantar os principais bancos de dados públicos e os métodos computacionais para a predição de miRNAs em plantas. Para tanto, realizou-se uma revisão integrativa da literatura científica publicada entre 2020 e 2025, com busca em bases de dados de artigos científicos, utilizando descritores em português e em inglês relacionados a miRNA, regulação gênica e bioinformática. Como resultados, foram identificados e caracterizados os principais repositórios de dados de miRNAs, bem como os principais métodos computacionais para a predição de miRNAs em plantas, categorizados em abordagens comparativas, não comparativas e baseadas em dados de sequenciamento de nova geração.

1 Bolsista PROBIC de Iniciação Científica do Curso de Engenharia da Computação da Universidade Estadual de Feira de Santana

2 Profa Adjunta da Universidade Estadual de Feira de Santana, Doutora em Ciência-Botânica

Adicionalmente, foi desenvolvido um tutorial descritivo para a utilização da ferramenta TargetFinder na plataforma Galaxy. Os desafios inerentes à execução de ferramentas de bioinformática limitam a realização desse tipo de pesquisa, o que indica que parcerias entre pesquisadores da área das Ciências Biológicas e afins e profissionais da tecnologia da informação podem ser excelentes alternativas para o desenvolvimento desse tipo de pesquisa, cujos resultados podem ser bastante promissores no entendimento dos mecanismos de expressão gênica regulados por moléculas de miRNA, utilizando-se modelos biológicos como plantas e outros organismos vivos.

Palavras-chave: bioinformática; engenharia genética; epigenética; mecanismo molecular; regulação gênica.

Abstract: MicroRNAs (miRNAs) are a type of regulatory molecule that allows (or prevents) the translation of mRNA (messenger RNA), thereby blocking the encoding of a given protein. Between 18 and 25 nucleotides in length, they act as important regulators of gene expression in plants, animals, and viruses. In plants, miRNAs play crucial roles in processes such as morphogenesis, floral induction, reproductive development, and responses to biotic and abiotic stresses. The identification and characterization of these molecules through experimental methods are costly and time-consuming, making computational approaches essential tools for reducing the hypothesis space to be experimentally validated. This study aimed to survey the main public databases and computational methods for predicting miRNAs in plants. To this end, an integrative review was conducted of the scientific literature published between 2020 and 2025, searching scientific article databases using descriptors in both Portuguese and English related to miRNA, gene regulation, and bioinformatics. The main miRNA data repositories were identified and characterized, as well as the main computational methods for predicting miRNAs in plants, categorized as comparative, non-comparative, and based on next-generation sequencing data. Additionally, a descriptive tutorial was developed for using the TargetFinder tool on the Galaxy platform. The inherent challenges in running bioinformatics tools

limit the conduct of this type of research, which indicates that partnerships between researchers in the Biological Sciences and related fields and information technology professionals can be excellent alternatives for developing this type of research, whose results can be highly promising in understanding the mechanisms of gene expression regulated by miRNA molecules, with the use of biological models such as plants and other living organisms.

Keywords: bioinformatics; epigenetics; gene regulation; genetic engineering; molecular mechanism.

INTRODUÇÃO

Os microRNAs (miRNAs) são importantes elementos reguladores da expressão gênica, constituindo uma das mais intrigantes descobertas biotecnológicas. Foram descritos na década de 1990 e, desde 2000, são relatados os seus papéis como reguladores biológicos, o que demonstra a participação dessas moléculas em diferentes processos biológicos. Desde então, essa classe de pequenos RNAs foi descrita também em plantas, nas quais vários avanços foram feitos visando ao melhor entendimento da sua ação e biogênese, bem como dos seus efeitos biológicos. Correspondem a pequenos RNAs (18 a 25 nucleotídeos), importantes para a regulação transcricional e pós-transcricional em animais, plantas e vírus. Os miRNAs são capazes de se ligar complementarmente às suas sequências de mRNA-alvo, levando à repressão traducional ou à clivagem do RNA. Eles se originam de precursores de fita simples (pré-miRNAs) capazes de formar estruturas em grampo (Pereira, 2015; Silva et al., 2015).

Os miRNAs vegetais desempenham papéis cruciais em diferentes processos, desde a morfogênese de folhas e raízes até a indução floral, formação de órgãos, reprodução e resposta ao estresse. Estão entre os mecanismos mais especializados, juntamente com o controle epigenético, que evoluíram para regular também a expressão dos genes MADS-box. Esses genes pertencem a uma classe de genes homeóticos que são responsáveis pela especificação do órgão floral e também estão envolvidos na iniciação e no desenvolvimento do meristema floral, o que sugere que tais genes

poderiam funcionar no topo da hierarquia regulatória envolvida no desenvolvimento da flor (Cheng et al., 2022).

Desde o reconhecimento dos miRNAs como uma classe distinta de reguladores fundamentais para diversos processos celulares em vários organismos, a caracterização desses pequenos RNAs passou a ser uma atividade essencial no estudo de genomas. No entanto, os procedimentos experimentais para a identificação dos loci de miRNAs ainda apresentam grandes limitações no que concerne ao custo e à morosidade. Por esse motivo, diversos métodos computacionais têm sido propostos para que o número de hipóteses a serem verificadas experimentalmente seja significativamente reduzido.

A identificação dos alvos dos miRNAs pode ser feita a partir da predição computacional dos alvos *in silico* ou por experimentação molecular, sendo que a compreensão da função dos miRNAs depende, em grande parte, da identificação de seus alvos. Na predição computacional, usa-se a sequência do miRNA para buscar possíveis alvos em bancos de genomas ou transcriptomas disponíveis para o organismo em estudo. Desde a descoberta dos primeiros miRNAs, foram desenvolvidos diversos programas de predição de alvos, entre os quais se destacam: TargetScan (Shaker et al., 2020), PicTar (Krek et al., 2005), miRanda (Betel et al., 2008), RNA22 (Miranda et al., 2006), DIANA-microT (Tastsoglou et al., 2023) e psRNATarget (Dai et al., 2018). Em geral, todos permitem predizer alvos de um miRNA específico em bancos de dados genômicos ou transcriptômicos e também verificar se um determinado transcrito pode ou não ser regulado por miRNAs já conhecidos. Como as regras de interação miRNA-alvo podem variar de um organismo para outro, muitos programas só podem ser usados para determinadas espécies.

A caracterização funcional de miRNAs tem fornecido evidências para o entendimento de importantes processos biológicos e lançado desafios em diversas áreas de pesquisa. O crescente número de genomas completos sequenciados e a necessidade de identificar seus miRNAs e precursores têm exigido o desenvolvimento de ferramentas de bioinformática cada vez mais robustas.

Diversos genes de miRNAs são conservados entre espécies filogeneticamente próximas. Contudo, outros miRNAs são confirmados como sendo espécie-específicos. Diante disso, a

caracterização dos miRNAs de um determinado organismo é dependente de uma busca inicial daqueles já identificados e, em seguida, da identificação de novos miRNAs particulares da espécie em estudo. Dessa forma, é imprescindível conhecer os repositórios de miRNAs previamente identificados.

Os métodos computacionais para a predição de miRNAs em genomas são organizados em comparativos e não comparativos, conforme utilizem ou não, como principal linha de solução, a conservação de elementos funcionais de genomas de espécies filogeneticamente próximas. Há, ainda, métodos baseados em dados de sequenciamento de nova geração (Gomes et al., 2013).

Por outro lado, os avanços em bioinformática ocorrem em velocidade considerável, sendo necessária a atualização constante por parte dos pesquisadores. Nessa perspectiva, a presente proposta visa ao levantamento dos principais bancos de dados públicos para miRNAs de plantas, bem como dos principais métodos computacionais para a predição de miRNAs em plantas.

METODOLOGIA

Foi realizada uma revisão integrativa, de acordo com Souza et al. (2010), a partir da seguinte questão norteadora: “Quais os principais bancos de dados públicos de miRNAs e quais os métodos computacionais utilizados para a predição de miRNAs em vegetais?”

Os principais bancos de dados públicos de miRNAs disponíveis online foram pesquisados, bem como os principais métodos computacionais para a predição de miRNAs em plantas, a partir de buscas na internet e da análise metodológica dos artigos científicos disponíveis sobre o tema nos últimos cinco anos, o que embasou a elaboração de um tutorial descritivo para o uso de um dos métodos computacionais selecionados entre os mais citados.

As bases de dados utilizadas para o levantamento dos artigos corresponderam aos principais repositórios de artigos publicados nas áreas de Ciências Biológicas, Genética e Biologia Molecular de plantas disponíveis online: arXiv, ScienceDirect, Microsoft Academic Search e SciELO. Foram utilizados descritores em português: miRNA, regulação gênica, expressão gênica, genes-alvo e

bioinformática; e em inglês: miRNA, gene regulation, gene expression, target genes e bioinformatics. Os operadores booleanos “e”, “ou”, “and” e “or” foram utilizados para combinar os termos de busca, a fim de ampliar ou limitar a recuperação de dados do levantamento. Como critérios de inclusão e exclusão, os artigos foram selecionados pelo título, que deveria conter os descritores propostos, e, em seguida, pela leitura do resumo, com o propósito de abarcar todos os trabalhos pertinentes à pesquisa que estivessem disponíveis online. Foi também utilizado o recorte temporal entre os anos de 2020 e 2025. Ainda entre os critérios de exclusão, foram desconsideradas todas as publicações que estivessem fora do recorte temporal proposto ou não apresentassem análise computacional dos dados. Também foram rejeitadas publicações de revisão bibliográfica, pois este estudo buscou identificar métodos computacionais aplicados nos estudos com miRNA. Para a sistematização dessa revisão, foi utilizado o protocolo PRISMA (Page et al., 2021; Moher et al., 2009), que delimitou a análise dos dados em quatro etapas: identificação, triagem, elegibilidade e inclusão.

Para a sistematização dos dados, foram elaboradas tabelas com informações fundamentais para o desenvolvimento da revisão. Na primeira, foram registrados os bancos de dados, a descrição, a URL, o grupo de estudo e o tipo de dado. Na segunda tabela, foram inseridos os métodos computacionais, a característica do método e a URL para acesso ao software ou a referência do artigo que o utilizou, ainda que este não esteja mais disponível.

RESULTADOS

Principais Bancos de Dados Públicos de miRNAs em Plantas

A revisão da literatura apontou cinco artigos sobre ferramentas de bioinformática para análise de miRNA em plantas. Com base nesses trabalhos e nos repositórios online encontrados, foram identificados e caracterizados os principais bancos de dados públicos para estudos de miRNAs em plantas. Os resultados estão sintetizados na Tabela 1.

Tabela 1. Principais bancos de dados públicos para miRNAs em plantas

Banco de dados	Descrição	URL	Grupo de estudo	Tipo de dado
miRBase	Repositório central de sequências de miRNA	https://www.mirbase.org/	Múltiplos	Sequências maduras e hairpin
NCBI SRA	Arquivo de sequenciamento de alto rendimento	https://www.ncbi.nlm.nih.gov/sra	Múltiplos	Dados brutos (FASTQ/SRA)
PMRD (Plant miRNA Database)	Combina dados disponíveis em bancos de dados públicos, com base na literatura recente. Atualizado a cada seis a oito meses	https://bioinformatics.cau.edu.cn/PMRD/	Plantas	Gene, genoma e anotações
PNRD (Plant Non-coding RNA Database)	Apresenta ferramentas de análise e busca por função dos RNAs	https://ngdc.cncb.ac.cn/databasecommons/database/id/138	Plantas	RNAs não codificantes (ncRNA)
P m i R E N (Plant miRNA ENcyclopedia)	Contém ampla gama de espécies, com intervalo de variedade que vai de algas verdes a angiospermas	https://pmiren.com/	Plantas	Loci e miRNA-alvo
miRNEST 2.0	Base de dados abrangente; fornece sequências de microRNA públicas, incluindo dados de expressão gênica e outras informações sobre os alvos	http://rhesus.amu.edu.pl/mirnest/copy/home.php	Múltiplos	Genes-alvo, precursores, informações básicas

Fonte: elaborado pelos autores (2026).

Entre os bancos de dados listados na Tabela 1, o miRBase (Kozomara & Griffiths-Jones, 2011) constitui o repositório central de sequências de miRNA, contendo informações sobre miRNAs maduros e precursores (hairpin) de diversas espécies. Possui divisão entre as versões de 2020 a 2022 e os lançamentos anteriores e reúne, em todas elas, ampla representação de modelos biológicos vegetais, a exemplo de *Arabidopsis thaliana* (L.) Heynh.

O Sequence Read Archive (SRA – Leinonen et al., 2011) armazena leituras curtas provenientes de sequenciamento de alto rendimento. Trata-se de um banco de dados que também funciona como repositório público para dados bioinformáticos. Em contraste com o miRBase, que se concentra nos microRNAs, o SRA volta-se ao armazenamento de dados massivos de cada espécie, de forma a servir como um referencial, tanto para comparações quanto para predições, complementando o miRBase. Nesse repositório, foram encontrados dados de sequenciamento para diferentes espécies de plantas, das quais foram selecionadas: *Cymbidium ensifolium* (L.) Sw. (sRNA de duas fases de floração), *Dendrobium officinale* Kimura & Migo (flores em diferentes estágios de desenvolvimento, PRJNA1191381), *Erycina pusilla* (L.) N.H. Williams & M.W. Chase (genoma e transcriptoma, GCA_000743925.1), *Orchis italica* Poir. (RNA-seq de inflorescência, SRX516901) e *Phalaenopsis aphrodite* Rchb. f. (sRNA-seq, SRA030409).

O Plant MicroRNA Database (PMRD – Zhang et al., 2010) é um banco de dados de anotação de miRNA específico para plantas, contendo sequências de miRNA publicadas, informações de alvos e dados de expressão. Complementar ao PMRD, o Plant Non-coding RNA Database (PNRD – Yi et al., 2015) também inclui, no banco de dados, outras categorias de RNAs, como os ncRNAs (RNAs não codificantes), além de apresentar ferramentas de análise e busca por função dos RNAs.

O miRNEST 2.0 (Szczesniak & Makalowska, 2014) contém informações provenientes de predições realizadas por outras ferramentas computacionais e recuperadas de outros repositórios, como o miRBase, além de integrar diversas ferramentas e permitir a interação do usuário com os dados.

Principais Métodos Computacionais para a Predição de miRNAs em Plantas

Os métodos computacionais para a predição de miRNAs em plantas podem ser organizados em três categorias principais: comparativos, não comparativos e baseados em dados de sequenciamento de nova geração (Tabela 2).

Os métodos comparativos utilizam a conservação de elementos funcionais entre genomas de espécies filogeneticamente próximas. Esses métodos baseiam-se no princípio de que miRNAs funcionais tendem a ser conservados evolutivamente. Ferramentas como o BLAST e abordagens baseadas em alinhamento de sequências enquadram-se nessa categoria. Já os métodos não comparativos não se utilizam de filogenia, baseando-se exclusivamente nas características intrínsecas das sequências, como estrutura secundária em grampo, composição de bases e termodinâmica do dobramento. O RNAfold é um exemplo de ferramenta utilizada nessa abordagem. Os métodos baseados em dados de sequenciamento de nova geração utilizam dados de sequenciamento profundo (RNA-seq, sRNA-seq) para identificar miRNAs expressos. Entre as ferramentas de bioinformática utilizadas nessa abordagem estão: miRDeep-P, miRDeep-P2, ShortStack, , miRNAFinder e PlantMirP2.

Tabela 2. Métodos e ferramentas computacionais para predição de miRNAs em plantas

Método	Ferramenta	Característica	URL/Referência
Baseado em NGS	miRDeep-P / miRDeep-P2 / miRDP2	Predição de novos miRNAs em plantas	Palani et al., 2025; Kuang et al., 2019; Yang & Li, 2011
Baseado em NGS	ShortStack	Análise de clusters de small RNAs	Palani et al., 2025; Axtell, 2013
Baseado em NGS	miRDeep2	Predição de miRNAs conhecidos e novos	https://www.mdc-berlin.de/
Predição de alvos	TargetFinder	Busca de sítios-alvo para miRNAs	Srivastava et al., 2014; Galaxy Tool Shed
Predição de alvos	psRNATarget	Específico para plantas	Dai et al., 2018; https://www.zhaolab.org/psRNATarget/
Análise estrutural	RNAfold	Predição de estrutura secundária	Lorenz et al., 2011 (ViennaRNA Package)
Montagem de transcriptoma	Trinity	Montagem de novo de RNA-seq	Grabherr et al., 2011
Comparativo	BLAST	Busca de similaridade	https://blast.ncbi.nlm.nih.gov/

Fonte: elaborado pelos autores (2026).

O miRDeep-P (Yang & Li, 2011) é um software específico para análise de miRNA de plantas, projetado para identificar novas sequências de pequenos RNAs em leituras curtas. Possui múltiplas versões disponíveis, como o miRDeep-P2 (Kuang et al., 2019), o miRDP2 e o miRDeep-2, com variações na parte de análise profunda e fina do dado inserido (Palani et al., 2025).

O ShortStack (Axtell, 2013) é uma ferramenta para análise de small RNAs que realiza o agrupamento e a classificação de miRNAs, identificando loci genômicos associados a pequenos RNAs. O ShortStack define parâmetros estruturais para miRNAs com base em miRNAs anotados selecionados no miRBase, dependendo do “miRType” especificado pelo usuário (“plant” ou “animal”). Essa ferramenta tem sido amplamente utilizada para a identificação de miRNAs a partir de dados de sRNA-seq (Palani et al., 2025).

O TargetFinder (Srivastava et al., 2014) é uma ferramenta que prediz computacionalmente sítios de ligação de small RNAs em transcritos-alvo a partir de um banco de dados de sequências, alinhando a sequência de small RNA de entrada contra todos os transcritos, com posterior pontuação dos sítios por meio de uma matriz de pontuação ponderada por posição. A saída inclui o nome da query, a descrição do alvo e o escore de predição, além de um diagrama de pareamento de bases.

O psRNATarget (Dai et al., 2018) é uma ferramenta de predição de alvos amplamente citada para plantas devido à sua sensibilidade às regras de pareamento específicas em vegetais. O psRNATarget incorpora descobertas recentes no reconhecimento de alvos de miRNA em plantas, distinguindo a inibição traducional e pós-transcricional, e reporta o número de pares miRNA/sítio-alvo que podem afetar a atividade de ligação do miRNA ao transcrito-alvo.

O Trinity (Grabherr et al., 2011) é um software para montagem de novo de transcriptomas a partir de dados de RNA-seq, essencial quando não há genoma de referência disponível. O Trinity opera em três estágios principais: Inchworm (montagem inicial de contigs), Chrysalis (agrupamento de contigs e construção de grafos de De Bruijn) e Butterfly (resolução de isoformas).

O RNAfold, parte do pacote ViennaRNA (Lorenz et al., 2011), é utilizado para a predição de estrutura secundária de precursores de miRNA, permitindo a identificação da estrutura em grampo.

Tutorial Descritivo para o Uso do TargetFinder na Plataforma Galaxy

Tutoriais são ferramentas de ensino e aprendizagem que auxiliam na realização de uma determinada tarefa e consistem em um recurso de grande utilidade na execução de ferramentas de informática, uma vez que a habilidade tecnológica que exigem não é de domínio geral. Assim, com o objetivo de guiar o usuário na utilização da ferramenta TargetFinder, disponível na plataforma Galaxy, para a predição de sítios-alvo de miRNAs em plantas, apresenta-se, a seguir, um tutorial para essa ferramenta.

Entre os pré-requisitos para a execução do tutorial estão: acesso à plataforma Galaxy (<https://usegalaxy.eu/> ou instância local), conhecimento básico de formato FASTA e arquivos de entrada — sequências de miRNA (formato FASTA) e banco de dados de sequências-alvo (formato FASTA). Os passos para a execução do tutorial estão descritos a seguir:

Passo 1: acessar a ferramenta TargetFinder

1. Acesse a plataforma Galaxy.
2. No painel de ferramentas à esquerda, utilize a barra de busca e digite “TargetFinder”.
3. Selecione a ferramenta TargetFinder (version 1.7.0+galaxy1).

Passo 2: preparar os arquivos de entrada

A ferramenta TargetFinder requer dois arquivos no formato FASTA: “-f”, arquivo com as sequências de small RNA (miRNA) de entrada; e “-d”, arquivo com o banco de dados de sequências-alvo (transcriptoma ou genoma). Nesse ponto, é importante certificar-se de que os arquivos estejam no formato FASTA correto, bem como utilizar sequências maduras de miRNA (18 a 25 nt).

Passo 3: configurar os parâmetros da ferramenta

1. No campo “Input small RNA sequences file”, selecione o arquivo FASTA contendo as sequências de miRNA (parâmetro “-f”).

2. No campo “Target sequence database file”, selecione o arquivo FASTA contendo as sequências-alvo (parâmetro “-d”).

Nesse passo, existem alguns parâmetros adicionais que são opcionais e incluem os comandos: “Cutoff score”, que define o escore de corte para a predição de alvos e tem como valor padrão 4.0 — escores mais baixos indicam predições mais rigorosas; “Use this option if the database is genomic DNA”, que deve ser marcado se o banco de dados for DNA genômico (em vez de transcriptoma); e “Output format”, para seleção do formato de saída desejado: “Classic” (formato padrão, com descrição e diagrama de pareamento), “GFF3” (formato geral de características, do inglês General Feature Format) e “Tab-delimited” (formato tabulado).

Passo 4: executar a análise

1. Clique no botão “execute” para iniciar a análise.
2. Aguarde a conclusão do processamento. O tempo de execução dependerá do tamanho dos arquivos de entrada.

Passo 5: interpretar os resultados

A saída do TargetFinder apresenta cada sítio-alvo predito separadamente. Para o formato “Classic”, cada saída contém:

1. Linha de descrição: com o nome da query (miRNA), a descrição do alvo e o escore de predição.

“query=miR399a, target=AT2G33770.1 Symbol: None ubiquitin-conjugating enzyme family protein, score=1.5”

2. Diagrama de pareamento de bases: apresenta a sequência do sítio-alvo na parte superior

(orientação 5'→3') e a sequência do miRNA na parte inferior (orientação 3'→5'). Os símbolos indicam: “.” (dois pontos), par Watson-Crick; “.” (ponto), par G:U; e “ ” (espaço), pareamento incorreto (mismatch).

target 5' UAGGGCAAUCUUCUUUGGCA 3'

.....

query 3' GUCCCGUUUAGAGGAAACCGU 5'

Passo 6: baixar os resultados

1. Após a conclusão, o arquivo de resultados estará disponível no histórico do Galaxy.
2. Clique no nome do arquivo e selecione “Download” para salvar os resultados localmente.

DISCUSSÃO

Principais Bancos de Dados Públicos de miRNAs em Plantas

Os resultados demonstraram que o miRBase (KOZOMARA et al., 2019; atualizado em GUO et al., 2020) continua sendo o repositório central e mais abrangente para sequências de miRNA, confirmando sua posição como a principal referência para esse tipo de pesquisa. Publicações recentes (2023 a 2025) ainda o citam como referência para a anotação de miRNAs.

O SRA (NCBI) permanece indispensável para o armazenamento de dados brutos de sequenciamento (FASTQ). Nas espécies de plantas selecionadas para as análises, a disponibilidade de dados no SRA variou consideravelmente: enquanto *P. aphrodite*, *E. pusilla* e *D. officinale* dispõem de dados extensos, *O. italica* e *C. ensifolium* apresentam conjuntos mais limitados. A integração de múltiplos repositórios tem se mostrado uma estratégia poderosa para estudos de miRNAs em plantas. Como demonstrado no estudo de Zhang et al. (2022) sobre melanoma cutâneo, a integração de dados de scRNA-seq do GEO com dados de RNA-seq em massa (bulk) do TCGA permitiu a identificação de novos biomarcadores. Essa abordagem integrativa pode ser adaptada para estudos em plantas,

combinando dados de expressão de diferentes fontes para uma caracterização mais completa.

Principais Métodos Computacionais para a Predição de miRNAs em Plantas

A diversidade de métodos computacionais disponíveis para a predição de miRNAs em plantas reflete a complexidade do problema e a evolução das tecnologias de sequenciamento. A categorização em métodos comparativos, não comparativos e baseados em sequenciamento de nova geração (NGS) fornece um arcabouço útil para a escolha da abordagem mais adequada.

O miRDeep-P (Yang & Li, 2011) revolucionou a predição de miRNAs em plantas ao integrar dados de sequenciamento profundo com análise de estrutura secundária. Sua evolução para o miRDeep-P2 (KUANG et al., 2019) trouxe melhorias significativas em velocidade e precisão, e a ferramenta foi validada em estudos recentes como superior às demais para análise de miRNAs em plantas. O miRDeep-P2 é considerado o mais sensível e rápido entre as ferramentas de predição de miRNAs avaliadas (Palani et al., 2025).

No entanto, a dependência de genomas de referência completos limita sua aplicação a espécies não modelo. Além disso, as ferramentas da família do miRDeep-P têm pouca acessibilidade e compatibilidade com o Galaxy (framework utilizado) e exigem domínio profundo e fino do estudo de microRNAs para a integração em um pipeline manual, sem travar o programa ou obter resultados indesejados.

O ShortStack emergiu como uma alternativa robusta para análise de small RNAs, especialmente em situações em que não há genoma de referência disponível. Sua capacidade de realizar o agrupamento e a classificação de miRNAs sem depender de anotação prévia o torna particularmente valioso para estudos em espécies pouco investigadas. No entanto, seu desempenho pode ser inferior ao do miRDeep-P2 em termos de número de miRNAs preditos, além de não ser nativo da plataforma utilizada neste estudo, sendo necessários plugins e extensões para o emprego concomitante à aplicação gerada.

O TargetFinder, disponível na plataforma Galaxy, representa uma ferramenta acessível para a predição de alvos de miRNA. Sua implementação na plataforma facilita o uso por pesquisadores sem experiência em linha de comando, embora as predições dependam da qualidade dos dados de entrada.

O psRNATarget continua sendo a ferramenta de predição de alvos mais citada para plantas devido à sua sensibilidade às regras de pareamento específicas de vegetais. No entanto, a interface web e a necessidade de carregar arquivos podem ser limitantes para análises em larga escala, além de apresentar o mesmo problema do ShortStack quanto à integração com outras etapas. O psRNATarget incorpora descobertas recentes no reconhecimento de alvos de miRNA em plantas, distinguindo a inibição traducional e pós-transcricional.

O Trinity, para montagem de novo de transcriptomas, é essencial quando não há genoma de referência disponível. Seu pipeline em três estágios (Inchworm, Chrysalis, Butterfly) permite a montagem de transcriptomas complexos a partir de dados de RNA-seq.

Desafios na Implementação de Pipelines de Bioinformática

Um dos principais desafios identificados neste estudo é a fragmentação das ferramentas de bioinformática e a complexidade de suas dependências. Grande parte das ferramentas analisadas é distribuída de forma esparsa e desorganizada, exigindo a instalação de múltiplas bibliotecas e dependências que frequentemente entram em conflito. Essa fragmentação dificulta significativamente a reprodução de análises e a adoção de pipelines por pesquisadores sem formação específica em bioinformática.

A documentação disponível para muitas ferramentas é limitada ou desatualizada, e a falta de tutoriais explícitos ou de uma comunidade de usuários ativa representa uma barreira adicional para novos usuários. Como observado nos relatórios de execução do pipeline, ferramentas como o Bowtie, o ShortStack e o miRDeep-P apresentaram falhas devido à ausência de dependências ou à

incompatibilidade de versões. O miRDeep-P2, por exemplo, requer um ambiente Linux com Perl, BioPerl, BLAST, Bowtie, SAMtools, R e DESeq2, o que pode dificultar a configuração.

A plataforma Galaxy destaca-se como uma solução promissora para mitigar esses desafios, oferecendo um ambiente integrado no qual as ferramentas podem ser executadas sem preocupação com dependências. Os tutoriais disponíveis na Galaxy Training Network (GTN) fornecem documentação estruturada e reproduzível para diversos fluxos de trabalho, incluindo análise de small RNAs, montagem de novo de transcriptomas e predição de alvos de miRNA. Contudo, mesmo na plataforma, a complexidade dos pipelines e a necessidade de compreender os parâmetros de cada ferramenta representam desafios. A integração do TargetFinder, do Trinity e dos programas de análise diferencial requer conhecimento sobre o fluxo de trabalho e a interpretação dos resultados.

A substituição do UEA sRNA Workbench (Stocks et al., 2012), descontinuado em 2019, por ferramentas mais modernas, como o ShortStack e o miRDeep-P2, representa um avanço, mas também exige adaptação por parte dos pesquisadores. A curva de aprendizado associada a essas novas ferramentas e a necessidade de validação cruzada dos resultados são aspectos que devem ser considerados no planejamento experimental.

Tutorial Descritivo para o Uso do TargetFinder

Quanto ao tutorial desenvolvido para o TargetFinder na plataforma Galaxy, este representa uma contribuição prática para a comunidade de pesquisadores em miRNAs de plantas. A escolha do TargetFinder justifica-se por sua ampla utilização em estudos de plantas, sua disponibilidade na plataforma e sua interface acessível.

O tutorial segue a estrutura recomendada pela Galaxy Training Network, com passos claros e objetivos, desde o acesso à ferramenta até a interpretação dos resultados. A inclusão de informações sobre os diferentes formatos de saída (Classic, GFF3, Tab-delimited) e a explicação dos símbolos no diagrama de pareamento de bases proporcionam ao usuário uma compreensão aprofundada dos

resultados.

A disponibilização deste tutorial em formato acessível contribui para a democratização do acesso a ferramentas bioinformáticas, reduzindo a barreira de entrada para pesquisadores sem formação em computação. Futuros trabalhos poderão expandir este tutorial para incluir outros métodos de predição de alvos (o psRNATarget e o miRanda) e fluxos de trabalho completos para a identificação e a caracterização de miRNAs em plantas.

CONSIDERAÇÕES FINAIS

Ao levantar os principais bancos de dados públicos e os métodos computacionais para a predição de miRNAs em plantas, constataram-se a diversidade e as peculiaridades inerentes às ferramentas de bioinformática disponíveis, o que se constitui em importantes desafios para a realização de pesquisas biológicas por pesquisadores que não dominam habilidades computacionais. Isso sugere que parcerias entre pesquisadores da área das Ciências Biológicas e afins e profissionais da tecnologia da informação são importantes no desenvolvimento desse tipo de pesquisa e podem indicar um caminho para encorajar esses pesquisadores no desenvolvimento de seus trabalhos.

Quanto às ferramentas propriamente ditas, observou-se que o TargetFinder, escolhido para o tutorial aqui disponibilizado, pode ser utilizado tanto para transcriptomas quanto para genomas completos. Para análises com múltiplos miRNAs, recomenda-se concatenar as sequências em um único arquivo FASTA, mas a qualidade da predição depende da qualidade das sequências de entrada e da abrangência do banco de dados-alvo. Para a validação dos resultados, recomenda-se a comparação com outras ferramentas de predição de alvos (o psRNATarget e o miRanda) e a validação experimental. Ainda é possível realizar múltiplas execuções, incluindo ferramentas diferentes (para tratamento de dados, verificação de qualidade, conversão de dados de múltiplos tipos) dentro de um pipeline, chamado na ferramenta do Galaxy de “Workflow”.

O desenvolvimento deste trabalho permitiu identificar e caracterizar os principais repositórios

de dados de miRNAs, entre os quais o miRBase, o PmiREN, o NCBI SRA e o miRNEST 2.0. O miRBase continua sendo o repositório central para sequências de miRNA, enquanto o PmiREN emerge como um recurso funcional abrangente para plantas. O SRA e o GEO fornecem dados brutos e processados essenciais para análises in silico.

Também foram mapeados os métodos computacionais para a predição de miRNAs em plantas, organizados em três categorias: comparativos (BLAST), não comparativos (RNAfold) e baseados em dados de NGS (miRDeep-P, miRDeep-P2, ShortStack e miRador). A evolução do miRDeep-P para o miRDeep-P2 representa um avanço significativo em velocidade e precisão. O surgimento de ferramentas baseadas em aprendizado de máquina, como o miRNAFinder e o PlantMirP2, e de conjuntos integrados de ferramentas, como o sRNAMiner, aponta para novas direções na área.

A análise da literatura revelou que a maioria dos estudos com miRNAs em plantas utiliza uma combinação de abordagens, com predominância do miRBase para a identificação de miRNAs conservados e de ferramentas como o miRDeep-P e o ShortStack para a predição de novo. A predição de alvos é realizada principalmente com o psRNATarget e o TargetFinder.

O tutorial prático para o uso do TargetFinder na plataforma Galaxy contribui para a disseminação desse conhecimento em bioinformática e facilita o acesso dos pesquisadores interessados na análise de miRNA vegetal, diminuindo o desafio na execução desse tipo de ferramenta.

Os resultados evidenciaram que, embora exista uma diversidade considerável de ferramentas bioinformáticas disponíveis, a fragmentação das soluções, a dependência de múltiplas bibliotecas, a defasagem dessas tecnologias e a escassez de documentação acessível constituem desafios significativos para pesquisadores, especialmente aqueles sem formação específica em bioinformática. A plataforma Galaxy destaca-se como uma solução promissora para mitigar esses desafios, oferecendo um ambiente integrado, uma comunidade ativa e tutoriais estruturados.

Pelo exposto, o presente trabalho forneceu um panorama abrangente dos recursos computacionais disponíveis para estudos de miRNAs em plantas, contribuindo para a disseminação do conhecimento e facilitando o acesso a ferramentas de bioinformática pela comunidade científica.

REFERÊNCIAS

- AXTELL, M. J. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA*, v. 19, n. 6, p. 740-751, 2013. DOI: <https://doi.org/10.1261/rna.035279.112>. Acesso em: 21 ago. 2026.
- BETEL, D. et al. The microRNA.org resource: targets and expression. *Nucleic Acids Research*, v. 36, supl. 1, p. D149-D153, 2008. DOI: <https://doi.org/10.1093/nar/gkm995>. Acesso em: 21 ago. 2026.
- CHENG, H. et al. Characterization of three SEPALLATA-like MADS-box genes associated with floral development in *Paphiopedilum henryanum* (Orchidaceae). *Frontiers in Plant Science*, v. 13, art. 916081, 2022. DOI: <https://doi.org/10.3389/fpls.2022.916081>. Acesso em: 21 ago. 2026.
- DAI, X.; ZHUANG, Z.; ZHAO, P. X. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research*, v. 46, n. W1, p. W49-W54, 2018. DOI: <https://doi.org/10.1093/nar/gky316>. Acesso em: 21 ago. 2026.
- GOMES, C. P. C. et al. A review of computational tools in microRNA discovery. *Frontiers in Genetics*, v. 4, art. 81, 2013. DOI: <https://doi.org/10.3389/fgene.2013.00081>. Acesso em: 21 ago. 2026.
- GRABHERR, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, v. 29, n. 7, p. 644-652, 2011. DOI: <https://doi.org/10.1038/nbt.1883>. Acesso em: 21 ago. 2026.
- GUO, Z. et al. PmiREN: a comprehensive encyclopedia of plant miRNAs. *Nucleic Acids Research*, v. 48, n. D1, p. D1114-D1121, 2020. DOI: <https://doi.org/10.1093/nar/gkz894>. Acesso em: 21 ago. 2026.
- KERTESZ, M. et al. The role of site accessibility in microRNA target recognition. *Nature Genetics*, v. 39, n. 10, p. 1278-1284, 2007. DOI: <https://doi.org/10.1038/ng2135>. Acesso em: 21 ago. 2026.
- KOZOMARA, A.; BIRGAOANU, M.; GRIFFITHS-JONES, S. miRBase: from microRNA sequences to function. *Nucleic Acids Research*, v. 47, n. D1, p. D155-D162, 2019. DOI: <https://doi.org/10.1093/nar/gky1141>. Acesso em: 21 ago. 2026.



KOZOMARA, A.; GRIFFITHS-JONES, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, v. 39, suppl. 1, p. D152-D157, 2011. DOI: <https://doi.org/10.1093/nar/gkq1027>. Acesso em: 21 ago. 2026.

KREK, A. et al. Combinatorial microRNA target predictions. *Nature Genetics*, v. 37, n. 5, p. 495-500, 2005. DOI: <https://doi.org/10.1038/ng1536>. Acesso em: 21 ago. 2026.

KUANG, Z. et al. miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. *Bioinformatics*, v. 35, n. 14, p. 2521-2522, 2019. DOI: <https://doi.org/10.1093/bioinformatics/bty972>. Acesso em: 21 ago. 2026.

LEINONEN, R.; SUGAWARA, H.; SHUMWAY, M. The Sequence Read Archive. *Nucleic Acids Research*, v. 39, suppl. 1, p. D19-D21, 2011. DOI: <https://doi.org/10.1093/nar/gkq1019>. Acesso em: 21 ago. 2026.

LORENZ, R. et al. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, v. 6, art. 26, 2011. DOI: <https://doi.org/10.1186/1748-7188-6-26>. Acesso em: 21 ago. 2026.

MIRANDA, K. C. et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, v. 126, n. 6, p. 1203-1217, 2006. DOI: <https://doi.org/10.1016/j.cell.2006.07.031>. Acesso em: 21 ago. 2026.

MOHER, D. et al. Reprint — preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Physical Therapy*, v. 89, n. 9, p. 873-880, 2009. DOI: <https://doi.org/10.1093/ptj/89.9.873>. Acesso em: 21 ago. 2026.

PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, v. 372, art. n71, 2021. DOI: <https://doi.org/10.1136/bmj.n71>. Acesso em: 21 ago. 2026.

PALANI, T. et al. Deciphering the impact of microRNAs in plant biology: a review of computational insights and experimental validation. *Molecular Biology Reports*, v. 52, n. 1, art. 209, 2025. DOI: <https://doi.org/10.1007/s11033-025-10273-0>. Acesso em: 21 ago. 2026.

PEREIRA, T. C. (org.). *Introdução ao mundo dos microRNAs*. Ribeirão Preto: Sociedade Brasileira de Genética, 2015. 342 p.

SHAKER, F. et al. Web-based tools for miRNA studies analysis. *Computers in Biology and Medicine*, v. 127, art. 104060, 2020. DOI: <https://doi.org/10.1016/j.compbimed.2020.104060>. Acesso em: 21 ago. 2026.

SILVA, E. M. et al. MicroRNAs e o desenvolvimento vegetal. In: PEREIRA, T. C. (org.). *Introdução ao mundo dos microRNAs*. Ribeirão Preto: Sociedade Brasileira de Genética, 2015. p. [inserir intervalo de páginas do capítulo].

SOUZA, M. T. S.; SILVA, M. D.; CARVALHO, R. Revisão integrativa: o que é e como fazer. *Einstein* (São Paulo), v. 8, n. 1, p. 102-106, 2010. DOI: <https://doi.org/10.1590/S1679-45082010RW1134>. Acesso em: 21 ago. 2026.

SRIVASTAVA, P. K. et al. A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, v. 15, art. 348, 2014. DOI: <https://doi.org/10.1186/1471-2164-15-348>. Acesso em: 21 ago. 2026.

STOCKS, M. B. et al. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, v. 28, n. 15, p. 2059-2061, 2012. DOI: <https://doi.org/10.1093/bioinformatics/bts311>. Acesso em: 21 ago. 2026.

SZCZEŚNIAK, M. W.; MAKALOWSKA, I. miRNEST 2.0: a database of plant and animal microRNAs. *Nucleic Acids Research*, v. 42, n. D1, p. D74-D77, 2014. DOI: <https://doi.org/10.1093/nar/gkt1156>. Acesso em: 21 ago. 2026.

TASTSOGLU, S. et al. DIANA-microT 2023: including predicted targets of virally encoded miRNAs. *Nucleic Acids Research*, v. 51, n. W1, p. W148-W153, 2023. DOI: <https://doi.org/10.1093/nar/gkad283>. Acesso em: 21 ago. 2026.

YANG, X.; LI, L. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, v. 27, n. 18, p. 2614-2615, 2011. DOI: <https://doi.org/10.1093/bioinformatics/btr430>. Acesso em: 21 ago. 2026.

YI, X. et al. PNRD: a plant non-coding RNA database. *Nucleic Acids Research*, v. 43, n. D1, p. D982-D989, 2015. DOI: <https://doi.org/10.1093/nar/gku1162>. Acesso em: 21 ago. 2026.

ZHANG, Z. et al. PMRD: plant microRNA database. *Nucleic Acids Research*, v. 38, supl. 1, p. D806-D813, 2010. DOI: <https://doi.org/10.1093/nar/gkp818>. Acesso em: 21 ago. 2026.

