v7.n11.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

# ESTUDO DESCRITIVO DAS FERRAMENTAS PARA ANÁLISE GENÔMICA: CARACTERÍSTICAS, POTENCIALIDADES E DESAFIOS

## *DESCRIPTIVE STUDY OF GENOMIC ANALYSIS TOOLS: CHARACTERISTICS, POTENTIAL, AND CHALLENGES*

## *ESTUDIO DESCRIPTIVO DE LAS HERRAMIENTAS DE ANÁLISIS GENÓMICO: CARACTERÍSTICAS, POTENCIAL Y DESAFÍOS*

Daniella Rezende dos Reis[1], Mariana Pires de Campos Telles[2], Mariana Oliveira Silva de Paula[3], Cimara Sales Vieira[4], Flávia Melo Rodrigues[5]

e7119450
https://doi.org/10.47820/recima21.v7i11.9450

PUBLICADO: 11/2026

**RESUMO**

O crescimento dos dados produzidos por tecnologias de sequenciamento de nova geração (NGS) ampliou a importância da bioinformática na análise genômica. Este estudo objetivou caracterizar ferramentas de bioinformática quanto às aplicações, funcionalidades, limitações, interfaces e condições de acesso. Realizou-se revisão sistematizada, qualitativa, exploratória e descritiva, na base Scopus. A busca identificou 14.051 registros; após aplicação do filtro territorial Brasil e dos critérios de elegibilidade, 212 estudos foram incluídos, dos quais 34 descreviam novas ferramentas. Um subconjunto intencional de 55 artigos, selecionado pela atualidade e pela completude das informações disponíveis, foi utilizado para extração detalhada, com identificação de 310 ferramentas. Foram examinados recursos clássicos e recentes, como BLAST, Galaxy, PGAP, Trimmomatic e SAMtools, além de diferentes interfaces e abordagens de aprendizado de máquina. No conjunto caracterizado, predominaram ferramentas de acesso gratuito, embora infraestrutura computacional, requisitos técnicos e conhecimentos especializados possam limitar sua utilização. A escolha das ferramentas deve considerar o objetivo analítico, as características dos dados, a disponibilidade de suporte e os recursos necessários. A revisão não realizou testes comparativos de desempenho nem análise econômica formal.

**PALAVRAS-CHAVE**: Análise Genômica. Softwares. Bioinformática.

***ABSTRACT***

*The expansion of data generated by next-generation sequencing (NGS) technologies has increased the importance of bioinformatics in genomic analysis. This study aimed to characterize bioinformatics tools according to their applications, functionalities, limitations, interfaces, and access conditions. A qualitative, exploratory, descriptive, systematized literature review was conducted in Scopus. The search identified 14,051 records; after applying the Brazil geographic*

[1] Graduanda em Ciências Biológicas – Licenciatura, Escola de Formação de Professores e Humanidades (EFPH); Pontifícia Universidade Católica de Goiás (PUC Goiás).
[2] Doutora em Ciências Ambientais pela Universidade Federal de Goiás (UFG), Programa de Pós-Graduação Stricto Sensu em Genética (PPGGEN), Programa de Pós-Graduação Stricto Sensu em Ciências Ambientais e Saúde (PPGCAS), Pontifícia Universidade Católica de Goiás (PUC Goiás).
[3] Doutoranda em Ciências Ambientais e Saúde (PPGCAS), Mestra pelo Programa de Pós-Graduação Stricto Sensu em Genética (PPGGEN), Pontifícia Universidade Católica de Goiás (PUC Goiás).
[4] Mestranda em Genética pelo Programa de Pós-Graduação Stricto Sensu em Genética (PPGGEN), Pontifícia Universidade Católica de Goiás (PUC Goiás).
[5] Doutora em Ciências Ambientais pela Universidade Federal de Goiás (UFG), Programa de Pós-Graduação Stricto Sensu em Genética (PPGGEN), Programa de Pós-Graduação Stricto Sensu em Ciências Ambientais e Saúde (PPGCAS), Pontifícia Universidade Católica de Goiás (PUC Goiás).



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

*filter and eligibility criteria, 212 studies were included, of which 34 described new tools. A purposively selected subset of 55 articles, chosen for recency and completeness of available information, underwent detailed extraction, identifying 310 tools. Established and recent resources, including BLAST, Galaxy, PGAP, Trimmomatic, and SAMtools, were considered alongside user interfaces and machine-learning approaches. Freely accessible tools predominated in the characterized subset, although computational infrastructure, technical requirements, and specialized knowledge may restrict their use. Tool selection should consider the analytical objective, data characteristics, available support, and required resources. No empirical performance benchmarking or formal economic evaluation was conducted.*

***KEYWORDS:*** *Genomic Analysis. Software. Bioinformatics.*

***RESUMEN***

*La expansión de los datos generados por las tecnologías de secuenciación de nueva generación (NGS) ha aumentado la importancia de la bioinformática para el análisis genómico. Este estudio tuvo como objetivo caracterizar herramientas bioinformáticas según sus aplicaciones, funcionalidades, limitaciones, interfaces y condiciones de acceso. Se realizó una revisión sistematizada de la literatura, cualitativa, exploratoria y descriptiva, en Scopus. La búsqueda identificó 14.051 registros; tras aplicar el filtro geográfico de Brasil y los criterios de elegibilidad, se incluyeron 212 estudios, de los cuales 34 describían herramientas nuevas. Se seleccionó intencionalmente un subconjunto de 55 artículos por su actualidad y la integridad de la información disponible, identificándose 310 herramientas en la extracción detallada. Se consideraron recursos clásicos y recientes, incluidos BLAST, Galaxy, PGAP, Trimmomatic y SAMtools, así como interfaces y enfoques de aprendizaje automático. Predominaron las herramientas de acceso gratuito en el subconjunto caracterizado, aunque la infraestructura informática, los requisitos técnicos y los conocimientos especializados pueden limitar su utilización. La selección debe considerar el objetivo analítico, las características de los datos, el soporte disponible y los recursos necesarios. No se realizaron pruebas empíricas comparativas de rendimiento ni una evaluación económica formal.*

***PALABRAS CLAVE:*** *Análisis genómico. Software. Bioinformática.*

## 1. INTRODUÇÃO

O sequenciamento do genoma humano representou um marco científico, cuja magnitude foi contextualizada por Lesk (2008). Desde então, as tecnologias de sequenciamento de nova geração ampliaram a produção de dados e a diversidade de organismos investigados. Repositórios internacionais como o National Center for Biotechnology Information (NCBI) e o Ensembl disponibilizam sequências, montagens e anotações genômicas, apoiando estudos comparativos e a reutilização de dados. Esses acervos são continuamente atualizados, e suas estatísticas dependem da unidade de contagem adotada (espécies, genomas ou montagens), razão pela qual números históricos não devem ser diretamente comparados sem padronização (NCBI, 2026; ENSEMBL, 2026).

Os avanços tecnológicos provenientes do surgimento e consolidação da revolução robótica no século XX permitiram, na década de 1980, o advento da bioinformática (Araújo *et al.,*



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2008). Essa área de conhecimento é a união entre a ciência da computação e a biologia, em que a computação contribui pela sua competência genuína em processamento e armazenamento de dados, além de seus métodos matemáticos sofisticados, permitindo a criação e utilização de programas de computador que possibilitam analisar grandes volumes de dados biológicos de forma rápida e eficiente (Lesk, 2008).

Diante desse cenário, a bioinformática tornou-se uma área fundamental para o armazenamento, processamento e interpretação de dados biológicos, permitindo descobertas importantes em diversas áreas, como medicina personalizada, farmacologia, agricultura, entre outras. A bioinformática evoluiu com os avanços em poder computacional e com projetos como o Genoma Humano, que impulsionaram a geração massiva de dados biológicos. O campo enfrenta o desafio constante de atualização tecnológica e metodológica, mas também oferece oportunidades crescentes, inclusive para jovens pesquisadores (Catanho; Miranda; Degrave, 2007; Verli, 2014).

Com o crescimento do acesso à internet, diversas ferramentas de análise genética passaram a estar disponíveis na web, facilitando o acesso de pesquisadores, professores e estudantes a metodologias antes restritas a laboratórios altamente especializados. Essas ferramentas possibilitam desde análises simples, como alinhamento de sequências, até investigações mais complexas, como inferências filogenéticas e identificação de variantes genéticas (Silva; Alves, 2024).

As tecnologias de sequenciamento de nova geração (NGS) vêm evoluindo rapidamente desde seu lançamento, em 2005. Novas plataformas vêm surgindo, com capacidade de gerar uma enorme quantidade de informações com grande economia de tempo e custo (Carvalho; Silva, 2010).

Nessa perspectiva, compreender as características e os requisitos de utilização dessas ferramentas é essencial para analisar dados genéticos de maneira criteriosa e eficiente (Silva; Alves, 2024). Este trabalho tem como objetivo caracterizar estudos nacionais que utilizam ferramentas de bioinformática destinadas à análise genômica quanto às suas aplicações, funcionalidades, limitações, interfaces de utilização e condições de acesso. Pretende-se oferecer subsídios para a escolha informada de recursos computacionais por estudantes e pesquisadores.

## 2. METODOLOGIA

Trata-se de uma pesquisa de abordagem qualitativa, de caráter exploratório e descritivo, desenvolvida por meio de uma revisão sistematizada da literatura, associada à análise



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

comparativa de ferramentas de bioinformática aplicadas às análises genômicas, transcriptômicas e metagenômicas. O levantamento bibliográfico foi realizado na base de dados Scopus, acessada por meio da Comunidade Acadêmica Federada (CAFe), disponibilizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Para a busca, foram utilizados os descritores *"genomic analysis platforms" OR "bioinformatics tools" OR "online genomic resources" OR "genome browser"*. O processo de identificação, triagem, elegibilidade e inclusão dos estudos foi sistematizado com base nas recomendações do PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (Page *et al.,* 2021).

A busca inicial resultou em 14.051 publicações. Posteriormente, aplicou-se o filtro territorial referente ao Brasil, com a finalidade de delimitar a produção científica com participação de pesquisadores vinculados a instituições brasileiras. Esse recorte foi adotado para identificar ferramentas de bioinformática desenvolvidas, descritas ou utilizadas em pesquisas com participação nacional, considerando a inserção da comunidade científica brasileira no desenvolvimento e na aplicação de recursos computacionais destinados às análises genômicas. Ressalta-se que a delimitação territorial se refere à vinculação institucional dos autores das publicações, não à origem ou à abrangência de utilização das ferramentas identificadas, que podem apresentar aplicações internacionais. Foi aplicado o filtro de país/território Brasil na Scopus, reduzindo o conjunto a 453 registros. Após leitura de títulos e resumos, excluíram-se 143 registros por inadequação temática e 27 por tipo de documento, totalizando 170 exclusões.

Foram considerados elegíveis estudos que descrevessem, apresentassem ou avaliassem ferramentas de bioinformática destinadas às análises genômicas, transcriptômicas ou metagenômicas. Após a aplicação dos critérios de elegibilidade, permaneceram 283 publicações, das quais 57 foram excluídas em razão da indisponibilidade de acesso ao conteúdo necessário para análise, resultando em 226 estudos. Posteriormente, foi realizada uma nova avaliação dos trabalhos previamente selecionados, com exclusão de 13 publicações por inviabilidade de acesso e uma por inadequação ao escopo temático da pesquisa. Ao final desse processo, 212 estudos foram incluídos na revisão. As etapas de seleção foram representadas em fluxograma elaborado com base nas recomendações do PRISMA 2020 (Figura 1).

Subsequentemente, os 212 estudos incluídos foram submetidos a uma análise inicial com o objetivo de identificar a presença de ferramentas de bioinformática relacionadas às análises genômicas, transcriptômicas e metagenômicas. Nessa etapa, foram identificados 34 artigos cuja finalidade principal consistia na apresentação ou descrição de ferramentas desenvolvidas pelos próprios autores.

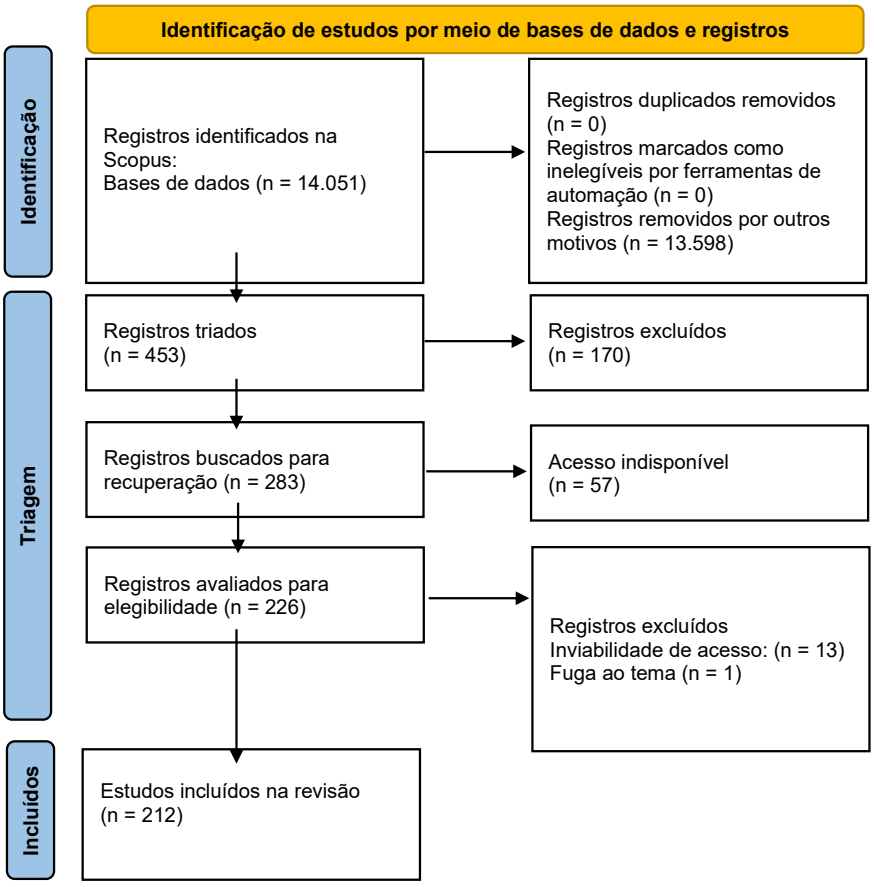
Os 212 estudos incluídos foram examinados para identificação de ferramentas de bioinformática. A extração detalhada das características foi realizada em um subconjunto



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

intencional de 55 artigos, selecionados pela atualidade das publicações e pela maior completude das informações técnicas disponíveis. Essa seleção não constituiu nova etapa de exclusão da revisão e não corresponde a amostragem probabilística. Foram registradas, quando disponíveis, finalidade, área de aplicação, funcionalidades, interface, limitações, acesso e licenciamento.

**Figura 1.** Fluxograma de identificação, triagem, elegibilidade e inclusão dos estudos, adaptado das recomendações PRISMA 2020




*Fonte: Elaborado pela autora, com base em PRISMA (2020).*

Nota: Dos 212 estudos incluídos na revisão, 55 foram selecionados por amostragem intencional para extração detalhada das informações sobre as ferramentas de bioinformática, considerando a atualidade das publicações e a completude das informações técnicas disponíveis.

A partir dos dados extraídos, realizou-se uma etapa de seleção das ferramentas destinadas à análise comparativa aprofundada. Foram considerados como critérios a atualidade



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

das publicações de apresentação ou descrição, a frequência de ocorrência das ferramentas entre os estudos analisados, sua relevância para as aplicações investigadas e a disponibilidade de informações necessárias à caracterização. Ferramentas de acesso pago também foram identificadas durante o levantamento; contudo, devido à baixa representatividade no conjunto analisado e à indisponibilidade de acesso integral aos respectivos artigos de descrição ou às próprias plataformas, não foram incluídas na análise aprofundada. Dessa forma, a etapa comparativa concentrou-se predominantemente em ferramentas gratuitas e de código aberto para as quais foi possível obter informações suficientes e verificáveis.

As informações extraídas foram organizadas em planilhas e quadros comparativos. A avaliação das condições de acesso e dos aspectos relacionados ao custo-benefício das ferramentas foi realizada de maneira qualitativa e descritiva, com base nas informações disponibilizadas nas publicações analisadas. Foram considerados aspectos como gratuidade, modalidade de licenciamento, disponibilidade de acesso, funcionalidades oferecidas, exigências de infraestrutura computacional e conhecimentos técnicos necessários à utilização. Nesse contexto, o custo-benefício foi interpretado em termos de acessibilidade e adequação dos recursos às necessidades das análises genômicas, sem a realização de estimativas monetárias ou de comparações experimentais de desempenho. Dessa forma, a avaliação buscou identificar potencialidades e limitações práticas das ferramentas, considerando as informações disponíveis na literatura consultada.

Ferramentas de inteligência artificial foram utilizadas exclusivamente como recurso auxiliar durante a elaboração do manuscrito, especialmente para revisão da escrita acadêmica, organização textual e síntese das informações apresentadas. Esses recursos não foram empregados na seleção dos estudos, na extração, na validação ou na interpretação dos dados, permanecendo essas etapas sob responsabilidade dos autores.

## 3. RESULTADOS

Dos 212 estudos selecionados para análise, 34 correspondiam a artigos destinados à descrição de novas ferramentas de bioinformática desenvolvidas pelos próprios autores. A análise inicial dos estudos permitiu identificar uma ampla diversidade de ferramentas de bioinformática. A extração detalhada e sistematizada das ferramentas foi realizada em 55 artigos. Nesses estudos, foram identificadas 310 ferramentas de bioinformática, evidenciando a variedade de recursos empregados nas análises consideradas neste trabalho.

Diante da quantidade e da diversidade de ferramentas identificadas, tornou-se necessário estabelecer um recorte para a análise aprofundada, de modo a contemplar aquelas



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

mais relacionadas aos objetivos desta pesquisa. Assim, as ferramentas foram selecionadas considerando sua atualidade e a disponibilidade de acesso aos respectivos artigos de descrição.

As ferramentas selecionadas foram analisadas quanto às suas aplicações, funcionalidades, limitações e disponibilidade de acesso. Os resultados dessa análise foram sintetizados e descritos nas Tabelas 1 e 2.

A Tabela 1 reúne ferramentas bioinformáticas direcionadas à análise e caracterização de dados genômicos, transcriptômicos e proteômicos, contemplando diferentes aplicações e abordagens computacionais. A Tabela 2 apresenta ferramentas consolidadas para processamento, análise e interpretação de sequências e dados genômicos, destacando suas principais aplicações, limitações e disponibilidade.

**Tabela 1.** Ferramentas de bioinformática mais recentes identificadas entre os 34 artigos destinados à descrição de novas ferramentas, com resumo descritivo, aplicações, limitações e condições de acesso

| FERRAMENTA | DESCRIÇÃO RESUMIDA | APLICAÇÃO PRINCIPAL | LIMITAÇÃO PRINCIPAL | ACESSO E LICENÇA | FONTE |
|---|---|---|---|---|---|
| **PANGENEWHALE** | Plataforma para análises pangenômicas. | Construção e análise de pangenomas. | Resultados variam conforme ferramenta e parâmetros. | Público/Gratuito; licença GPL-3.0 | Netto *et al*., 2026 |
| **BIOCIDE-RESISTANCE-SCANNER** | Ferramenta para identificar genes de resistência a biocidas. | Detecção de resistência em bactérias. | Limitada à identificação gênica. | Gratuito; licença MIT | Rosa; Kremer, 2024 |
| **BIOPIPELINE CREATOR** | Interface para criação de pipelines bioinformáticos. | Análises genômicas, metagenômicas e transcriptômicas. | Servidor requer Linux. | Público/Gratuito; licença GPL-3.0 | Maia Cordeiro *et al*., 2024 |
| **GENTLE** | Ferramenta com aprendizado de máquina para repertórios de TCR. | Análise de TCR e identificação de biomarcadores. | Não analisa séries temporais. | Gratuito; licença MIT | Andrade *et al*., 2023 |
| **CONJUNTO DE PIPELINES** | Pipelines para processamento, montagem e anotação genômica. | Análise de genomas bacterianos. | Exige recursos computacionais; específico para bactérias. | Público/Gratuito; licença GPL-3.0 | Almeida *et al*., 2023 |
| **INPACTOR2** | Ferramenta de aprendizado profundo para LTR-retrotransposons. | Anotação de genomas vegetais. | Depende da qualidade dos dados de treinamento. | Público/Gratuito; licença GPL-3.0 | Orozco-Arias *et al*., 2023 |



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

| | | | | | |
|---|---|---|---|---|---|
| **PIPELINE ÔMICO PARA COVID-19** | Fluxo integrado para análise transcriptômica e proteômica. | Identificação de genes, proteínas e biomarcadores. | Exige conhecimento em linha de comando e espaço de armazenamento. | Todos os comandos utilizados são disponibilizados gratuitamente pelos autores. | Macedo-Da-Silva *et al*., 2022 |
| **CRISPRLOCI** | Ferramenta para identificação e classificação de sistemas CRISPR-Cas. | Caracterização de CRISPR-Cas em procariotos. | Depende de modelos de aprendizado de máquina. | Gratuito; licença não especificada. | Alkhnbashi *et al*., 2021 |

Fonte: Elaborado pelas autoras.
Nota: As informações sobre funcionalidades, limitações, condições de acesso e licenciamento foram sistematizadas a partir das publicações analisadas. A classificação como ferramenta gratuita não implica necessariamente disponibilização do código-fonte ou ausência de custos associados à infraestrutura computacional. As condições de acesso e licenciamento podem sofrer alterações ao longo do tempo.

**Tabela 2.** Principais ferramentas clássicas de bioinformática mencionadas nos 55 estudos analisados sistematicamente, com resumo descritivo, aplicações, limitações, condições de acesso e número de citações entre os 212 artigos

| FERRAMENTA | DESCRIÇÃO RESUMIDA | APLICAÇÕES PRINCIPAIS | LIMITAÇÕES | ACESSO | FONTE | Nº REPORTADO |
|---|---|---|---|---|---|---|
| **BLAST** | Busca rápida de sequências semelhantes (DNA/proteínas). | Identificação de genes e regiões conservadas. | Difícil equilibrar velocidade e sensibilidade; muitos hits pouco informativos. | Gratuito | Altschul *et al.,* 1990; Altschul *et al.,* 1997 | 78 |
| **GALAXY** | Plataforma web para análise genômica sem programação. | Combinar, filtrar e visualizar dados genômicos. | Processamento lento em consultas grandes; limitações de formato. | Gratuito, código aberto | Giardine *et al*., 2005 | 12 |
| **PGAP** | Pipeline para análise de pan-genomas bacterianos. | Estudos evolutivos e funcionais de genes. | Processamento pesado; parametrização complexa. | Gratuito | Zhao *et al*., 2012 | 16 |
| **TRIMMOMATIC** | Pré-processamento de dados Illumina (corte, filtragem, adaptadores). | Limpeza de dados de sequenciamento paired-end. | Dificuldade em detectar adaptadores curtos/parciais. | Público, GPL-3.0 | Bolger *et al*., 2014 | 18 |
| **SAMTOOLS** | Manipulação de arquivos SAM/BAM (alinhamentos). | Conversão, ordenação, remoção de duplicatas, SNPs. | Conversão e ordenação de grandes volumes são lentas. | Gratuito, MIT | Li *et al*., 2009 | 6 |

Fonte: Elaborado pelas autoras.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Nota: As informações sobre funcionalidades, limitações, condições de acesso e licenciamento foram sistematizadas a partir das publicações analisadas. A classificação como ferramenta gratuita não implica necessariamente disponibilização do código-fonte ou ausência de custos associados à infraestrutura computacional. As condições de acesso e licenciamento podem sofrer alterações ao longo do tempo.

## 4. DISCUSSÃO

Os resultados obtidos neste trabalho evidenciam a ampla diversidade e disponibilidade de ferramentas computacionais destinadas às análises genômicas, abrangendo diferentes aplicações e funcionalidades. A quantidade e a diversidade de ferramentas identificadas sugerem um cenário de rápida expansão e especialização dos recursos computacionais destinados à análise de dados biológicos, indicando que o desenvolvimento das ferramentas bioinformáticas tem acompanhado a ampliação das questões biológicas que podem ser investigadas por meio de dados genômicos. Os detalhes técnicos das ferramentas analisadas, embora variados, convergem em desafios comuns de parametrização, tempo de processamento e curva de aprendizado.

Além da expansão do número de ferramentas, observa-se a integração de diferentes etapas analíticas, buscando tornar os fluxos de trabalho mais acessíveis, automatizados e reprodutíveis. A exemplo disso, a plataforma PanGeneWhale integra diversas ferramentas em um ambiente unificado, visando simplificar a realização de análises pangenômicas (*Netto et al.,* 2026). De maneira semelhante, o biocide-resistance-scanner reúne diferentes softwares e bancos de dados para a identificação de genes relacionados à resistência a biocidas em genomas bacterianos (Rosa; Kremer, 2024), enquanto o conjunto de pipelines desenvolvido por Almeida *et al.* (2023) integra etapas de pré-processamento, montagem e anotação genômica. Essa integração também é observada no CRISPRloci, que combina diferentes abordagens para a identificação e caracterização de sistemas CRISPR-Cas (Alkhnbashi *et al.,* 2021).

Entre os recursos discutidos, Inpactor2 utiliza aprendizado profundo para identificar e classificar retrotransposons LTR em genomas vegetais (Orozco-Arias *et al.,* 2023). Abordagens baseadas em aprendizado de máquina também aparecem em ferramentas destinadas a outras tarefas, como a análise de repertórios de receptores de células T (Andrade *et al.,* 2023). Métodos baseados em modelos treinados podem reconhecer padrões complexos, enquanto estratégias heurísticas e algoritmos tradicionais podem oferecer maior transparência de parâmetros ou menor exigência computacional em determinadas aplicações. Entretanto, a precisão depende dos dados de treinamento, do domínio de aplicação e das métricas utilizadas; os estudos reunidos não permitem estabelecer uma comparação universal de acurácia, tempo de execução



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

ou consumo de recursos. Assim, a escolha entre abordagens deve considerar validação independente, reprodutibilidade e infraestrutura disponível.

Também foram observadas diferentes modalidades de interação com as ferramentas identificadas, incluindo interfaces de linha de comando (CLI), interfaces gráficas (GUI) e plataformas web. Embora as interfaces gráficas e web possam facilitar a utilização por usuários com menor familiaridade com recursos computacionais, algumas ferramentas ainda são executadas por meio de linha de comando, como o BLAST, que se destacou como a ferramenta mais mencionada entre os 212 artigos. O BLAST e suas derivações, BLASTn, BLASTp (Altschul *et al.,* 1990; Altschul *et al.,* 1997), foram reportados em 65 artigos e, entre as novas ferramentas encontradas neste estudo, 13 delas incorporaram essa ferramenta.

O biocide-resistance-scanner (Rosa; Kremer, 2024) e os pipelines ngs-preprocess, MpGAP (Genome Assembly Pipeline) e bacannot (Genome Annotation Pipeline) desenvolvidos por Almeida *et al.* (2023) são outros exemplos de ferramentas executadas por meio de interface de linha de comando. Essa característica pode representar uma barreira para usuários com pouca experiência computacional, além de exigir conhecimentos específicos para sua configuração, execução e interpretação dos resultados (Altschul *et al.,* 1990; Almeida *et al.,* 2023; Rosa; Kremer, 2024; Maia Cordeiro *et al.,* 2024; Netto *et al.,* 2026). Nesse sentido, Maia Cordeiro *et al. (*2024) destacam que a maioria das ferramentas desenvolvidas para análises ômicas é destinada a aplicações específicas e frequentemente baseia-se em interfaces de linha de comando, o que pode dificultar sua utilização e integração em pipelines personalizados. Os autores também ressaltam que ferramentas que apresentam maior facilidade de uso ainda estão, em muitos casos, disponíveis principalmente em versões pagas.

Em contraste com essa perspectiva, os resultados obtidos neste estudo demonstraram predominância de ferramentas de acesso gratuito, sendo parte delas também disponibilizada como software de código aberto (PanGeneWhale, biocide-resistance-scanner, BioPipeline Creator, GENTLE, Galaxy, conjunto de pipelines, Inpactor2, SAMtools). Essa característica foi observada em diferentes modalidades de utilização, incluindo ferramentas de linha de comando, interfaces gráficas, plataformas web e ferramentas que empregam aprendizado de máquina. Dessa forma, os resultados indicam a existência de diferentes alternativas de acesso para a realização de análises genômicas, embora aspectos relacionados à interface, à infraestrutura computacional e ao conhecimento necessário para a utilização das ferramentas ainda possam influenciar sua acessibilidade e aplicabilidade pelos usuários.

A predominância de ferramentas gratuitas no subconjunto caracterizado pode favorecer atividades de ensino e pesquisa, mas não deve ser extrapolada para todo o ecossistema internacional de bioinformática, dado o recorte territorial e a seleção intencional de artigos.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Embora a gratuidade e a disponibilização de código aberto favoreçam o acesso às ferramentas de bioinformática, essas características não implicam necessariamente menores custos operacionais. Algumas ferramentas exigem infraestrutura computacional especializada, capacidade elevada de processamento e armazenamento, além de conhecimentos técnicos específicos para instalação, parametrização e interpretação dos resultados. Assim, a escolha de uma ferramenta deve considerar não apenas sua disponibilidade financeira, mas também os recursos computacionais e humanos necessários à sua implementação. Ressalta-se que a presente investigação não realizou uma análise econômica quantitativa nem testes comparativos de desempenho, de modo que as considerações sobre custo-benefício se restringem às características técnicas e às condições de acesso descritas nas publicações analisadas.

De modo geral, os resultados demonstram que o desenvolvimento das ferramentas bioinformáticas tem sido marcado não apenas pelo aumento da quantidade de recursos disponíveis, mas também pela diversificação de suas aplicações, pela integração de diferentes etapas analíticas e pela incorporação de abordagens computacionais mais avançadas, como o aprendizado de máquina. A predominância de ferramentas gratuitas e/ou de acesso público amplia as possibilidades de utilização desses recursos em diferentes contextos de pesquisa. Entretanto, diferenças quanto às interfaces de interação, aos requisitos de infraestrutura e ao conhecimento computacional necessário para sua utilização evidenciam que a disponibilidade de uma ferramenta não implica, necessariamente, facilidade de acesso e uso. Nesse sentido, a escolha da ferramenta mais adequada deve considerar não apenas sua funcionalidade, mas também sua acessibilidade, requisitos técnicos e adequação às necessidades específicas de cada análise genômica (Lefin *et al.,* 2024).

A incorporação de técnicas de aprendizado profundo (*deep learning*) às ferramentas de bioinformática representa um avanço no processamento e na interpretação de dados genômicos, especialmente em aplicações que envolvem o reconhecimento de padrões complexos e a classificação de sequências biológicas. Diferentemente de abordagens tradicionais baseadas predominantemente em regras e procedimentos heurísticos, os modelos de aprendizado profundo podem identificar relações complexas a partir de grandes conjuntos de dados de treinamento, apresentando potencial para melhorar a acurácia em determinadas tarefas. Entretanto, esses benefícios dependem da qualidade e da representatividade dos dados utilizados no treinamento, da arquitetura dos modelos e das condições de validação. Além disso, a implementação dessas abordagens pode exigir maior capacidade de processamento, memória e infraestrutura computacional, particularmente durante o treinamento dos modelos. Portanto, a adoção de ferramentas baseadas em aprendizado profundo deve considerar o equilíbrio entre desempenho preditivo, interpretabilidade, requisitos computacionais e adequação ao problema



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

biológico investigado (Lefin *et al.,* 2024). Ressalta-se que o presente estudo não realizou comparações experimentais de desempenho, de modo que não é possível estabelecer superioridade de acurácia ou eficiência computacional entre as ferramentas analisadas.

## 5. CONSIDERAÇÕES FINAIS

Este estudo permitiu caracterizar a diversidade de ferramentas de bioinformática destinadas às análises genômicas e outras abordagens ômicas, contemplando diferentes aplicações, funcionalidades, limitações e condições de acesso. Os resultados evidenciaram a disponibilidade de recursos computacionais voltados a etapas específicas de análise, bem como de plataformas que integram diferentes procedimentos, favorecendo a automatização e a organização dos fluxos de trabalho. Destaca-se, ainda, a incorporação de abordagens de aprendizado de máquina e inteligência artificial, que ampliam as possibilidades de processamento e interpretação de dados biológicos.

A predominância de ferramentas gratuitas e de código aberto entre aquelas analisadas representa uma oportunidade para ampliar o acesso às análises genômicas. Entretanto, a gratuidade não implica necessariamente facilidade de utilização ou ausência de custos operacionais, uma vez que determinadas ferramentas exigem infraestrutura computacional especializada e conhecimentos técnicos específicos.

Nesse contexto, recomenda-se que a seleção de ferramentas de bioinformática considere, prioritariamente, a finalidade da investigação, a natureza e o volume dos dados, as funcionalidades disponíveis, os requisitos computacionais, a facilidade de utilização, as condições de acesso e o licenciamento. Também é importante avaliar a disponibilidade de documentação técnica, atualizações e suporte, bem como a possibilidade de integração com outras ferramentas e a reprodutibilidade dos procedimentos analíticos. Esses critérios podem auxiliar pesquisadores na identificação de recursos mais adequados às necessidades científicas e às condições de infraestrutura disponíveis.

Por fim, ressalta-se que a presente investigação apresenta caráter descritivo e foi desenvolvida a partir de um recorte bibliográfico com participação institucional brasileira, utilizando uma amostragem intencional para a caracterização detalhada das ferramentas. A ausência de avaliações experimentais de desempenho e de análises econômicas quantitativas limita comparações diretas de acurácia, eficiência computacional e custo-benefício. Estudos futuros poderão incorporar testes comparativos padronizados, considerando diferentes conjuntos de dados e ambientes computacionais, de modo a subsidiar avaliações mais abrangentes sobre



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

o desempenho e a aplicabilidade dessas ferramentas.

**AGRADECIMENTOS**

Ao financiamento concedido pelo CEGGen – Centro de Excelência em Genética e Genômica (FAPEG proc. 202410267001291) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas de Mestrado e Doutorado concedidas a Cimara Sales Vieira e Mariana Oliveira Silva de Paula, respectivamente.

**REFERÊNCIAS**

ALKHNBASHI, O. S. et al. CRISPRloci: comprehensive and accurate annotation of CRISPR-Cas systems. ***Nucleic Acids Research***, v. 49, n. W1, p. W125-W130, 2021. DOI: 10.1093/nar/gkab456. Disponível em: https://academic.oup.com/nar/article/49/W1/W125/6300616. Acesso em: 27 set. 2026.

ALMEIDA, F. M. de; CAMPOS, T. A. de; PAPPAS JR., G. J. Scalable and versatile container-based pipelines for *de novo* genome assembly and bacterial annotation. ***F1000Research***, v. 12, art. 1205, 2023. DOI: 10.12688/f1000research.139488.1. Disponível em: https://f1000research.com/articles/12-1205. Acesso em: 27 set. 2026.

ALTSCHUL, S. F. et al. Basic local alignment search tool. ***Journal of Molecular Biology***, v. 215, n. 3, p. 403-410, 1990. DOI: 10.1016/S0022-2836(05)80360-2. Disponível em: https://www.sciencedirect.com/science/article/pii/S0022283605803602. Acesso em: 27 set. 2026.

ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. ***Nucleic Acids Research***, v. 25, n. 17, p. 3389-3402, 1997. DOI: 10.1093/nar/25.17.3389. Disponível em: https://academic.oup.com/nar/article/25/17/3389/1061651. Acesso em: 27 set. 2026.

ANDRADE, D. S. et al. GENTLE: a novel bioinformatics tool for generating features and building classifiers from T cell repertoire cancer data. ***BMC Bioinformatics***, v. 24, n. 1, art. 32, 2023. DOI: 10.1186/s12859-023-05155-w. Disponível em: https://link.springer.com/article/10.1186/s12859-023-05155-w. Acesso em: 27 set. 2026.

ARAÚJO, N. D. de et al. A era da bioinformática: seu potencial e suas implicações para as ciências da saúde. ***Estudos de Biologia***, v. 30, n. 70/72, p. 143-148, 2008. DOI: 10.7213/reb.v30i70/72.22819. Disponível em: https://periodicos.pucpr.br/estudosdebiologia/article/view/22819. Acesso em: 27 set. 2026.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. ***Bioinformatics***, v. 30, n. 15, p. 2114-2120, 2014. DOI: 10.1093/bioinformatics/btu170. Disponível em: https://academic.oup.com/bioinformatics/article/30/15/2114/2390096. Acesso em: 27 set. 2026.

CARVALHO, M. C. da C. G. de; SILVA, D. C. G. da. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. ***Ciência Rural***, v. 40, n. 3, p. 735-744, 2010. DOI: 10.1590/S0103-84782010000300040. Disponível em:



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

https://www.scielo.br/j/cr/a/5R3MnWFC6jNXQhBzCPFnsWH/?lang=pt. Acesso em: 27 set. 2026.

CATANHO, M. P.; MIRANDA, A. B. de; DEGRAVE, W. M. S. Comparando genomas: bancos de dados e ferramentas computacionais para a análise comparativa de genomas procarióticos. ***Revista Eletrônica de Comunicação, Informação & Inovação em Saúde***, v. 1, n. 2, p. 217-226, 2007. DOI: 10.3395/reciis.v1i2.933. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/933. Acesso em: 27 set. 2026.

ENSEMBL. **Genome selector.** EMBL-EBI, 2026. Disponível em: https://www.ensembl.org/genome-selector. Acesso em: 21 out. 2026.

GIARDINE, B. et al. Galaxy: a platform for interactive large-scale genome analysis. ***Genome Research***, v. 15, n. 10, p. 1451-1455, 2005. DOI: 10.1101/gr.4086505. Disponível em: https://genome.cshlp.org/content/15/10/1451. Acesso em: 27 set. 2026.

LEFIN, N. et al. Review and perspective on bioinformatics tools using machine learning and deep learning for predicting antiviral peptides. ***Molecular Diversity***, v. 28, n. 4, p. 2365-2374, 2024. DOI: 10.1007/s11030-023-10718-3. Disponível em: https://link.springer.com/article/10.1007/s11030-023-10718-3. Acesso em: 21 out. 2026.

LESK, A. M. ***Introdução à bioinformática***. 2. ed. Porto Alegre: Artmed, 2008.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. ***Bioinformatics***, v. 25, n. 16, p. 2078-2079, 2009. DOI: 10.1093/bioinformatics/btp352. Disponível em: https://academic.oup.com/bioinformatics/article/25/16/2078/204688. Acesso em: 27 set. 2026.

MACEDO-DA-SILVA, J. et al. Exploring COVID-19 pathogenesis on command-line: a bioinformatics pipeline for handling and integrating omics data. In: DONEV, R. (ed.). ***Protein Interaction Networks***. United States: Elsevier Academic Press, 2022. p. 311-339. (*Advances in Protein Chemistry and Structural Biology*, v. 131). DOI: 10.1016/bs.apcsb.2022.04.002. Disponível em: https://www.sciencedirect.com/science/chapter/bookseries/pii/S1876162322000311?via%3Dihub. Acesso em: 27 set. 2026.

MAIA CORDEIRO, C. et al. BioPipeline Creator: a user-friendly Java-based GUI for managing and customizing biological data pipelines. ***Scientific Reports***, v. 14, art. 16572, 2024. DOI: 10.1038/s41598-024-67409-8. Disponível em: https://www.nature.com/articles/s41598-024-67409-8. Acesso em: 27 set. 2026.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI). **Genomes**. Bethesda: NCBI, 2026. Disponível em: https://www.ncbi.nlm.nih.gov/home/genomes/. Acesso em: 21 out. 2026.

NETTO, W. de B. G. et al. PanGeneWhale: a dockerized Kotlin-based GUI platform for reproducible and user-friendly pangenomic analysis. ***Computational Biology and Chemistry***, v. 120, art. 108779, 2026. DOI: 10.1016/j.compbiolchem.2025.108779. Disponível em: https://www.sciencedirect.com/science/article/pii/S1476927125004414. Acesso em: 27 set. 2026.

OROZCO-ARIAS, S. et al. Inpactor2: a software based on deep learning to identify and classify LTR-retrotransposons in plant genomes. ***Briefings in Bioinformatics***, v. 24, n. 1, art. bbac511, 2023. DOI: 10.1093/bib/bbac511. Disponível em:



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

https://academic.oup.com/bib/article/24/1/bbac511/6887110. Acesso em: 27 set. 2026.

PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, v. 372, art. n71, 2021. DOI: 10.1136/bmj.n71. Disponível em: https://www.bmj.com/content/372/bmj.n71. Acesso em: 27 set. 2026.

ROSA, E. E. B. da; KREMER, F. S. The mobilome landscape of biocide-resistance in Brazilian ESKAPE isolates. ***Brazilian Journal of Microbiology***, v. 55, n. 4, p. 3603-3616, 2024. DOI: 10.1007/s42770-024-01450-7. Disponível em: https://link.springer.com/article/10.1007/s42770-024-01450-7. Acesso em: 27 set. 2026.

SILVA, R. C. C. da; ALVES, M. C. S. O uso de ferramentas de bioinformática para análise de dados genéticos: uma revisão. ***Scientific Electronic Archives***, v. 17, n. 1, p. 85-91, 2024. DOI: 10.36560/17120241872. Disponível em: https://sea.ufr.edu.br/SEA/article/view/1872. Acesso em: 27 set. 2026.

VERLI, H. (org.). ***Bioinformática: da biologia à flexibilidade molecular***. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.

ZHAO, Y. et al. PGAP: pan-genomes analysis pipeline. ***Bioinformatics***, v. 28, n. 3, p. 416-418, 2012. DOI: 10.1093/bioinformatics/btr655. Disponível em: https://academic.oup.com/bioinformatics/article/28/3/416/188161. Acesso em: 27 set. 2026.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.