

GenTreat – Pipeline computacional para montagem híbrida de genomas procarióticos

Victória Cardoso dos Santos

Graduanda de Engenharia de Computação

Instituição: Universidade Federal do Pará (UFPA) -Campus Tucuruí

Endereço: Rodovia BR 422, km 13, Canteiro de Obras UHE, Vila Permanente, Tucuruí -PA,
CEP: 68464-000

E-mail: victoriacardoso82@gmail.com

Adonney Allan de Oliveira Veras

Doutor em Genética e Biologia Molecular

Instituição: Faculdade de Computação da Universidade Federal do Pará (UFPA) -Campus
Castanhal

Endereço: Av. dos Universitários, s/n, Jaderlândia, Castanhal -PA, CEP: 68746-630

E-mail: allanverasce@ufpa.br

ABSTRACT

The development of omics sciences has been greatly influenced by sequencing technologies. However, the large volume of data generated by these technologies necessitates the development of new computational tools for processing and analysis, particularly in genome assembly. Two main approaches are commonly used for assembly: reference-based assembly, which maps sequencing reads against a reference genome, and *de novo* assembly, which performs assembly without a reference. *De novo* assembly techniques include Overlap Layout-consensus, *De Bruijn* graph, and greedy algorithms. Although numerous tools have been developed over the years, challenges persist, including fragmented assembly results and redundant contigs. As a result, new methods have emerged, such as hybrid assembly strategies that combine results from different assemblers. However, existing tools utilizing this strategy often involve extensive and complex command lines. GenTreat is a computational pipeline with an intuitive graphical interface, was developed for automated hybrid assembly of prokaryotic genomes, performs assembly using two assemblers, merges the results, and then orders and annotates the assembled genome. Validation using raw reads from 61 organisms demonstrated that is a viable alternative for automated hybrid assembly, eliminating the need for using extensive command lines. The tool is available at: <https://sourceforge.net/projects/gentreat/>.

KEYWORDS: Pipeline, Hybrid Assembly, Assembly, NGS, Prokaryote.

RESUMO

O desenvolvimento das ciências ômicas foi muito influenciado pelas tecnologias de sequenciamento. No entanto, o grande volume de dados gerados por essas tecnologias torna necessário o desenvolvimento de novas ferramentas computacionais para processamento e análise, principalmente na montagem do genoma. Duas abordagens principais são comumente usadas para montagem: montagem baseada em referência, que mapeia as leituras de sequenciamento em relação a um genoma de referência, e montagem *de novo*, que executa a

montagem sem uma referência. Técnicas de montagem *de novo* incluem *Overlap Layout-consensus*, grafo *De Bruijn* e algoritmos gulosos. Embora inúmeras ferramentas tenham sido desenvolvidas ao longo dos anos, os desafios persistem, incluindo resultados de montagem fragmentados e contigs redundantes. Como resultado, surgiram novos métodos, como estratégias de montagem híbrida que combinam resultados de diferentes montadores. No entanto, as ferramentas existentes que utilizam essa estratégia geralmente envolvem linhas de comando extensas e complexas. O GenTreat é um pipeline computacional com uma interface gráfica intuitiva, foi desenvolvido para montagem híbrida automatizada de genomas procarióticos, realiza a montagem usando dois montadores, mescla os resultados e, em seguida, ordena e anota o genoma montado. A validação usando leituras brutas de 61 organismos demonstrou que é uma alternativa viável para montagem híbrida automatizada, eliminando a necessidade do uso de extensas linhas de comando. A ferramenta está disponível em: <https://sourceforge.net/projects/gentreat/>.

PALAVRAS-CHAVE: Pipeline, Montagem Híbrida, Montagem, NGS, Procarioto.

1 INTRODUÇÃO

Os sequenciadores de DNA de Nova Geração (NGS) revolucionaram o sequenciamento de genomas ao oferecer melhorias na velocidade, custo-efetividade e geração de leituras altamente precisas. No entanto, o grande volume de dados produzidos por esses sequenciadores em um curto período, juntamente com as características específicas das leituras geradas, exigiu o desenvolvimento de ferramentas computacionais capazes de lidar e analisar esses dados de forma eficaz (1).

Como resultado, inúmeras ferramentas computacionais foram desenvolvidas para diversas análises, incluindo o processo de montagem de genomas. Alguns exemplos notáveis dessas ferramentas incluem SPAdes (2), Velvet (3), Megahit (4), Edena (5) e ALLPATHS (6).

A montagem de genomas pode ser abordada por meio de dois métodos principais. O primeiro método, conhecido como montagem por referência, envolve mapear as leituras de sequenciamento brutas contra um genoma de referência de um organismo que é filogeneticamente próximo ao organismo alvo sendo montado. O segundo método é chamado de montagem *de novo*, que não depende de um genoma de referência durante o processo de montagem. Diversas técnicas são empregadas na montagem de novo, incluindo algoritmos gulosos, *Overlap Layout-consensus* e grafos de *De Bruijn* (7).

Apesar da existência de várias ferramentas de alta acurácia para a montagem de genomas, a tarefa continua sendo desafiadora no campo da bioinformática, com problemas persistentes como contigs redundantes e resultados fragmentados (8,9). Consequentemente, há a necessidade de explorar novas estratégias para melhorar a qualidade dos resultados de montagem. Uma dessas abordagens é a adoção da montagem híbrida, que combina os resultados de diferentes montadores. Várias ferramentas, incluindo Unicycler (10), HASLR (11),

MaSuRCA (12), DBG2OLC (13) e Cerulean (14), foram desenvolvidas para implementar a montagem híbrida, empregando diferentes abordagens que combinam leituras longas e curtas. No entanto, essas ferramentas frequentemente exigem linhas de comando extensas e complexas com múltiplos parâmetros (9).

Neste artigo, apresenta-se o GenTreat, um pipeline computacional com uma interface gráfica amigável, projetado para montagem híbrida automatizada. O GenTreat combina os resultados de diferentes montadores, em seguida, realiza a ordenação e busca melhorar os resultados da montagem, aproximando-os da sequência original do genoma. Além disso, o GenTreat possui um processo de anotação para os resultados da montagem.

2 MATERIAIS E MÉTODOS

2.1 Linguagem de Programação e Banco de Dados

O pipeline foi desenvolvido usando a linguagem de programação Java (15) e sua interface gráfica foi construída utilizando a biblioteca Swing. Para o gerenciamento de banco de dados, foi utilizado o SQLite versão 3 (16). O gerenciamento de dependências foi realizado pelo Maven, o qual simplifica o manuseio das dependências do projeto ao fornecer um repositório centralizado conhecido como Maven Central Repository (17).

2.2 Montadores, Integrador de Contigs e Ordenação

A montagem *de novo* das leituras brutas foi realizada usando o software SPAdes versão 3.15.5 e Megahit versão 1.2.9. Posteriormente, os resultados dos diferentes montadores foram integrados usando o CISA versão 1.3 (18). Para ordenar os *contigs* montados, foi utilizada a versão 2.4.0 do Mauve (19). É importante destacar que todos os valores de parâmetros para esses programas podem ser configurados diretamente na interface gráfica do GenTreat.

2.3 Análise de completude gênica

A análise de completude gênica foi realizada usando o software BUSCO versão 5.4.7 (20), que é uma ferramenta amplamente utilizada na pesquisa genômica para avaliar a integridade e completude gênica. Os parâmetros utilizados nesta análise foram "-m genome" para o modo genoma e "--auto-lineage-prok" para seleção automática de linhagem, ignorando eucariotos.

2.4 Avaliação da qualidade da montagem

A versão 5.2.0 do QCAST (Quality Assessment Tool for Genome Assemblies) foi utilizada para avaliar a qualidade das montagens de genomas. Ela fornece métricas e estatísticas abrangentes para avaliar a precisão dos genomas montados (21).

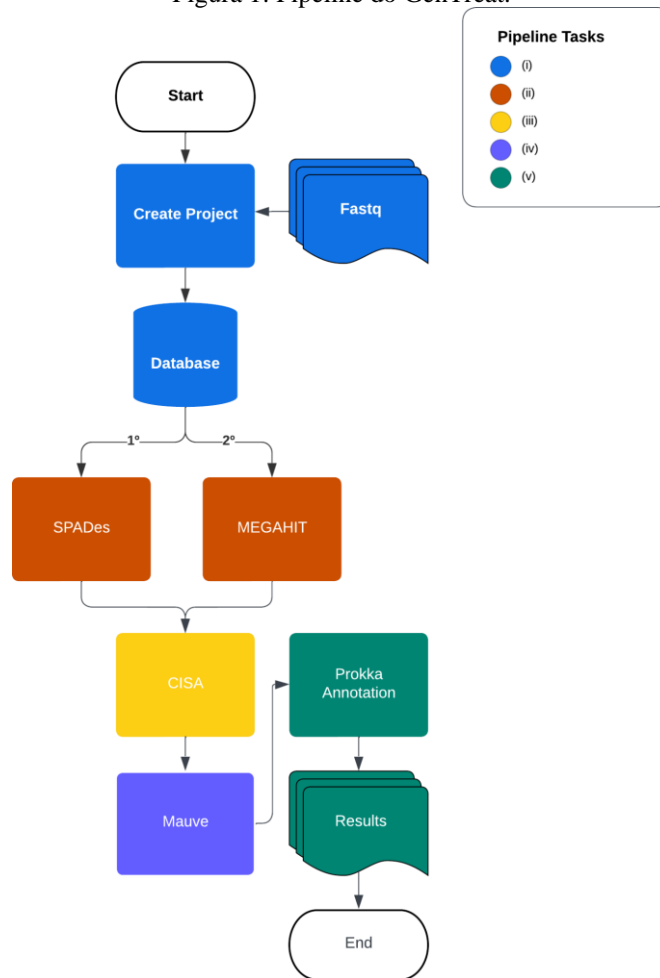
2.5 Anotação Automática

Após o processo de ordenação, as sequências resultantes foram submetidas à anotação automática. Esta etapa envolveu a previsão de elementos genéticos, como CDS, tRNAs e rRNAs, por meio da transferência de informações de um genoma de referência contido no banco de dados. Para esse propósito, foi utilizado o anotador Prokka versão 1.14.6 (22).

2.6 Pipeline do GenTreat

A Figura 1 ilustra o pipeline do GenTreat, que pode ser descrito nas seguintes etapas: (i) Criação de um projeto, entrada de parâmetros e arquivos de entrada no formato FASTQ (aceitando leituras single-end ou paired-end). Adicionalmente, um arquivo GenBank (GB) é usado para o processo de ordenação. É importante notar que todas as informações de entrada são salvas no banco de dados do GenTreat, permitindo que o processo seja retomado do ponto de interrupção em caso de falha externa. (ii) Montagem dos dados brutos usando os montadores SPAdes e Megahit, com os resultados da montagem sendo então enviados para integração. (iii) Integração dos resultados da montagem usando o CISA. (iv) Ordenação dos resultados integrados usando o Mauve. (v) Anotação automática dos resultados ordenados usando a ferramenta Prokka.

Figura 1: Pipeline do GenTreat.



Fonte: Autoria própria.

2.7 Validação da ferramenta

Para a validação da ferramenta, um total de 61 organismos foi utilizado. Esses organismos foram obtidos do banco de dados público do National Center for Biotechnology Information (NCBI), disponível em (23). Os organismos foram baixados no formato FASTQ usando a ferramenta fastq-dump, que é um componente do SRA-Toolkit. Os valores de parâmetros padrão foram usados para leituras single-end, e para leituras paired-end, o parâmetro "-split-files" foi aplicado.

Para obter os arquivos de referência no formato GB, foi acessado o banco de dados de nucleotídeos do NCBI em (24). Os organismos usados nas análises, juntamente com suas respectivas informações, estão listados na Tabela 1 do Apêndice. Durante o processo de validação do GenTreat, os valores de parâmetros padrão de cada programa incluído no pipeline foram utilizados.

O GenTreat passou por validação em duas etapas. A primeira etapa envolveu a comparação dos resultados de processamento obtidos com o GenTreat com aqueles obtidos com

as ferramentas de montagem híbrida HASLR e Cerulean. Essa comparação foi realizada utilizando a cepa *Escherichia coli K-12 substr. MG1655* (número SRA: SRR10971019). Vale ressaltar que esse organismo foi o mesmo utilizado para validar as ferramentas HASLR (11) e Cerulean (14) em seus respectivos artigos científicos.

Na segunda etapa de validação, as montagens dos organismos listados na Tabela 3, no Apêndice, foram realizadas utilizando as ferramentas SPAdes, Megahit, Velvet e GenTreat para fins de comparação. Todos os resultados das duas etapas foram posteriormente avaliados utilizando o software QUAST para avaliar a qualidade das montagens.

3 RESULTADOS E DISCUSSÕES

A Tabela 2 exibe os resultados de montagem obtidos utilizando as ferramentas HASLR, Cerulean e GenTreat. As métricas de avaliação consideradas nesta comparação são as seguintes: a porcentagem de bases alinhadas no genoma de referência (Fração do Genoma), o número de contigs, o comprimento da menor sequência de contig representando 50% do comprimento total da montagem (N50), o comprimento da menor sequência de contig representando 75% do comprimento total da montagem (N75), o comprimento dos blocos alinhados cobrindo pelo menos metade do genoma de referência (NGA50), o comprimento do contig mais longo, o comprimento total da montagem, a taxa de variação no comprimento devido a inserções ou deleções de bases (Indel), a taxa de divergência em todas as bases alinhadas (Incompatibilidade) e o número de pontos de quebra nos contigs (Misassemblies).

É importante notar que, ao conduzir a análise das ferramentas de montagem híbrida para comparação (Tabela 2), certas métricas não foram apresentadas nos artigos científicos das ferramentas HASLR e Cerulean. Conseqüentemente, alguns valores na Tabela 2 não puderam ser contabilizados devido à indisponibilidade dessas métricas nos respectivos artigos científicos.

Tabela 2: Análise dos resultados utilizando as ferramentas HASLR, Cerulean e GenTreat.

Número SRA	Montador	Fração do Genoma (%)	Quantidade de Contigs	N50	N75	NGA50
SRR10971019	HASLR	99.992	2	-	-	3970011
	Cerulean	-	21	694	507	-
	GenTreat	99.994	14	4642256	4642256	3818549
	Montador	Maior Contig	Comprimento Total	Taxa Indel	Taxa de divergências	Misassemblies
	HASLR	-	-	79.85	22.62	2
	Cerulean	1991897	4625935	-	-	4
	GenTreat	4642256	4856726	1.20	9.74	2

Fonte: Autoria própria.

Embora o GenTreat produza um número maior de contigs em comparação com a ferramenta HASLR, é evidente que o GenTreat alcança uma porcentagem maior de bases mapeadas contra a referência. Além disso, o GenTreat demonstra uma taxa menor de divergências (mismatches) e indels (inserções/deleções), indicando maior acurácia dos dados e redução de viés. Apesar da montagem ligeiramente mais fragmentada (maior número de contigs), as taxas mais baixas de divergências e indels alcançadas pelo GenTreat proporcionam maior confiabilidade em comparação com os resultados produzidos pelo HASLR.

Quando comparado com a ferramenta Cerulean, o GenTreat gera menos contigs, indicando uma montagem menos fragmentada. Além disso, o GenTreat alcança valores de N50 e N75 mais altos, indicando uma qualidade superior de montagem em termos de contiguidade de sequência. Ademais, o GenTreat gera contigs mais longos, o que significa que ele monta com sucesso sequências maiores e não fragmentadas (maior contig) em comparação com o Cerulean, indicando ainda mais a alta qualidade da montagem do genoma.

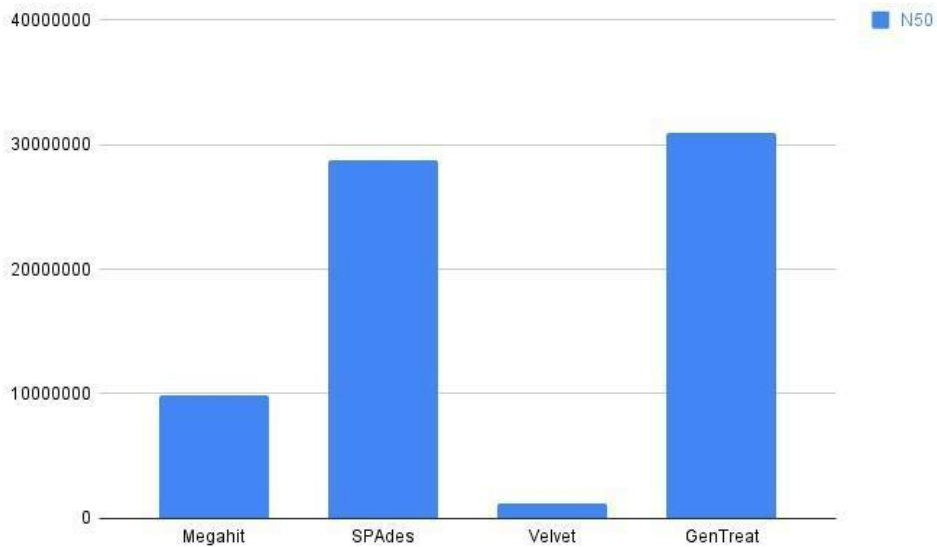
Os resultados da comparação entre as montagens realizadas pelas ferramentas SPAdes, Megahit, Velvet e GenTreat para cada organismo são apresentados na Tabela 3 do Apêndice. Com base nesses resultados, as Figuras 2 a 7 foram geradas.

As métricas consideradas na análise do software QUAST incluem a porcentagem de bases alinhadas no genoma de referência (Fração do Genoma), o número de contigs, o comprimento da menor sequência de contig representando 50% do comprimento total da montagem (N50), o comprimento dos blocos alinhados cobrindo pelo menos metade do genoma de referência (NGA50), o comprimento da menor sequência de contig, o comprimento do contig mais longo, o comprimento total da montagem e as desmontagens (misassemblies).

É importante observar que o QUAST avalia contigs com um comprimento de 500 ou mais pares de bases. Consequentemente, certos resultados gerados podem não ter sido considerados na análise por este software devido ao seu comprimento insuficiente, como indicado na Tabela 3 do Apêndice.

A Figura 2 ilustra a comparação dos valores de N50. A análise demonstra que entre os montadores Megahit, SPAdes, Velvet e o pipeline GenTreat, o último consistentemente produz os valores mais altos de N50. Isso indica que o GenTreat gera resultados de montagem com qualidade e contiguidade de sequência superiores em comparação com os outros montadores sendo comparados.

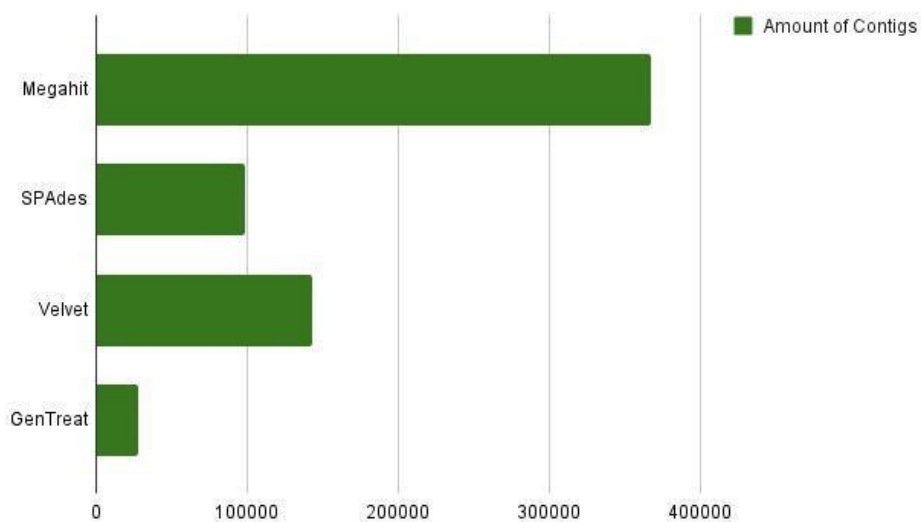
Figura 2: Análise Comparativa de N50.



Fonte: Autoria própria.

Ao analisar o número total de contigs gerados em cada resultado de montagem pelos respectivos montadores (Figura 3), torna-se evidente que o GenTreat produz o menor número de contigs entre eles. Essa descoberta sugere que as montagens geradas pelo GenTreat são menos fragmentadas e, portanto, exibem maior precisão. Essa observação está alinhada com os resultados apresentados na Figura 2, corroborando ainda mais a ideia de que o GenTreat se destaca na produção de montagens de alta qualidade e mais contíguas.

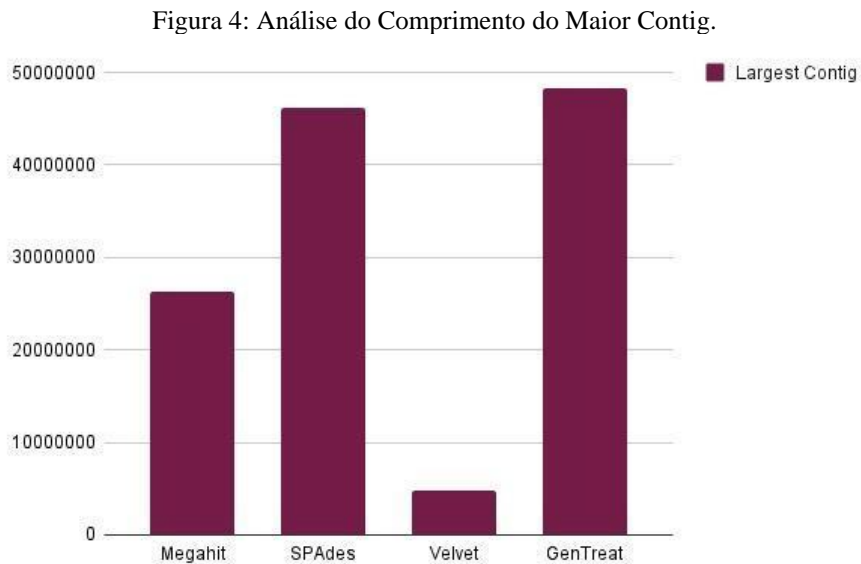
Figura 3: Análise do Número de Contigs por Ferramenta.



Fonte: Autoria própria.

A Figura 4 ilustra a comparação dos valores de comprimento para o maior contig gerado por cada ferramenta para cada organismo. A análise revela que o GenTreat exibe os valores

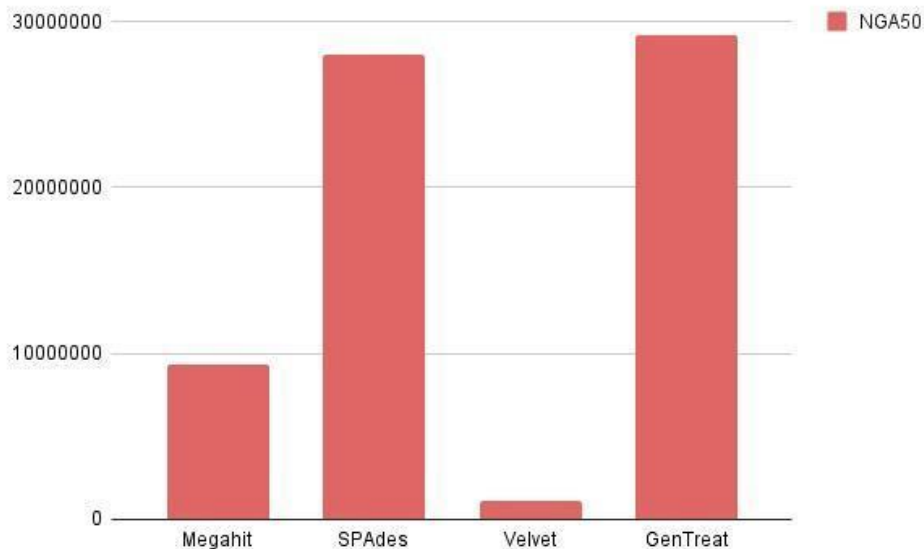
mais altos para o comprimento do maior contig, seguido pelas ferramentas SPAdes, Megahit e Velvet, respectivamente. O comprimento do maior contig significa a extensão da sequência de nucleotídeos gerada no processo de montagem. Conseqüentemente, é digno de nota que o GenTreat demonstra uma capacidade superior em produzir resultados de montagem que são mais precisos e confiáveis em comparação com outras ferramentas.



Fonte: Autoria própria.

A Figura 5 ilustra o comprimento dos blocos alinhados que cobrem pelo menos metade do genoma de referência, conhecido como NGA50. A análise revela que o pipeline GenTreat produz os valores mais altos de NGA50, seguido pelos resultados obtidos com a ferramenta SPAdes. Esta análise destaca a precisão das montagens do GenTreat em comparação com o genoma de referência, enfatizando sua qualidade, menor fragmentação e maior comprimento de sequência.

Figura 5: Análise do NGA50.



Fonte: Autoria própria.

A análise das figuras geradas, quando comparadas com outras ferramentas, revela que a maioria dos resultados de montagem obtidos com o pipeline GenTreat exibem menor fragmentação, valores mais altos de N50 e NGA50 e comprimentos de contig mais longos. Esses resultados indicam que as montagens do GenTreat possuem qualidade superior, maior precisão, contiguidade aprimorada e maior confiabilidade.

Os resultados da montagem foram submetidos ao software BUSCO para realizar uma análise de completude gênica. A análise de completude gênica, comparando os resultados de montagem obtidos usando o montador Velvet e o pipeline GenTreat, revelou que, dos 61 organismos analisados, 11 exibiram uma completude gênica superior em comparação com os resultados de montagem utilizando o Velvet. Entre eles, 2 organismos mostraram valores iguais de completude gênica (Tabela 4 do Apêndice).

Na análise de completude gênica comparando GenTreat com SPAdes, os resultados foram os seguintes: dos 61 organismos analisados, 3 não produziram nenhum resultado na análise BUSCO (Tabela 5 do Apêndice). Na comparação entre GenTreat e MEGAHIT, os resultados foram os seguintes: dos 61 organismos analisados, 9 apresentaram o mesmo nível de completude gênica, e 6 encontraram erros durante a execução do BUSCO, resultando na ausência de arquivos de resultados (Tabela 6 do Apêndice).

No entanto, observou-se que, na maioria dos casos, a diferença entre os resultados das comparações entre SPAdes, MEGAHIT e GenTreat não foi significativa. Por outro lado, ao analisar todas as métricas observadas, constatou-se que o pipeline GenTreat é uma alternativa viável para o processo de montagem. Além disso, ele realiza a ordenação automática dos

resultados e a subsequente anotação, adicionando ainda mais valor à análise. Em última análise, o desenvolvimento desta ferramenta visa fornecer à comunidade científica uma alternativa para a obtenção de resultados de montagem que devem ser analisados mais detalhadamente pelo pesquisador.

A Tabela 7 fornece uma visão comparativa das características de usabilidade e desempenho das ferramentas de montagem híbrida Cerulean, HASLR, Unicycler, MaSuRCA e GenTreat. É evidente que o GenTreat abrange várias funcionalidades relacionadas à montagem híbrida que estão ausentes nas demais, incluindo a utilização de leituras paired-end e single-end, bem como a presença de uma interface gráfica, entre outras.

Tabela 7: Resumo das tarefas realizadas pelas ferramentas Cerulean, HASLR, Unicycler, MaSuRCA e GenTreat.

Descrição das tarefas	Cerulean	HASLR	Unicycler	DBG2OLC	MaSuRCA	GenTreat
Realiza montagem híbrida.	X	X	X	X	X	X
Precisa de leituras curtas e longas juntas para realizar a montagem híbrida.	X	X	X	X	X	
Funciona com leituras single-end e paired-end.			X			X
Permite a escolha de valores de parâmetros.	X	X	X	X	X	X
Possui uma opção para ordenar o resultado da montagem.						X
Possui interface gráfica.						X
Realiza anotação automática.						X

Fonte: Autoria própria.

No entanto, é importante enfatizar que o GenTreat não é uma ferramenta de montagem híbrida que segue a mesma abordagem que as demais. Em vez disso, ele é um pipeline que permite a montagem utilizando dois montadores, unindo os resultados obtidos, seguido pela ordenação e anotação, sem a necessidade de diferentes tipos de leituras, como leituras longas e curtas. Portanto, uma comparação direta entre essas ferramentas não seria viável devido às diferenças nas abordagens. Assim, a comparação deve se basear nos resultados da montagem, utilizando métricas consolidadas como N50, número de contigs produzidos (refletindo a análise de fragmentação), comprimento do maior contig, entre outras.

4 CONCLUSÃO

GenTreat é uma ferramenta projetada com uma interface amigável, proporcionando uma experiência de análise mais simples e intuitiva. Na comparação realizada com os softwares que executam a mesma tarefa que o GenTreat, foi possível verificar que o uso do pipeline se mostrou eficiente e, na comparação das métricas, apresentou melhores resultados.

Portanto, ao final da avaliação de todas as montagens, foi possível observar que o GenTreat apresentou os melhores resultados de montagem de genomas, seguido pelas ferramentas SPAdes e Megahit, de acordo com as métricas avaliadas. Esse resultado é esperado, pois SPAdes e Megahit fazem parte do pipeline de montagem híbrida do GenTreat. Por essa razão, os resultados desses dois montadores foram muito próximos aos resultados do GenTreat em algumas métricas.

Assim, o pipeline GenTreat é uma alternativa viável para realizar montagem híbrida, apresentando montagens confiáveis e precisas. Além disso, ele executa esse processo automaticamente por meio de uma interface gráfica, sem a necessidade de usar linhas de comando longas e complexas, facilitando seu uso por usuários que não possuem formação em computação.

AGRADECIMENTOS

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Agradecemos à Universidade Federal do Pará. Este trabalho faz parte da pesquisa desenvolvida pelo grupo de pesquisa BIOD (Bioinformática, Ômicas e Desenvolvimento - www.biod.ufpa.br). AAOV agradece à Universidade Federal do Pará (UFPA), PHCGS e JTCA agradecem à Universidade Federal Rural da Amazônia (UFRA).

REFERÊNCIAS

1. J. Henson, G. Tischler, Z. Ning, Next-generation sequencing and Large Genome Assemblies, *Pharmacogenomics*. 13 (2012) 901–915. doi:10.2217/pgs.12.72.
2. A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, et al., Spades: A new genome assembly algorithm and its applications to single-cell sequencing, *Journal of Computational Biology*. 19 (2012) 455–477. doi:10.1089/cmb.2012.0021.
3. D.R. Zerbino, Using the velvet *de novo* assembler for short-read Sequencing Technologies, *Current Protocols in Bioinformatics*. 31 (2010). doi:10.1002/0471250953.bi1105s31.
4. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, Megahit: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de bruijn* graph, *Bioinformatics*. 31 (2015) 1674–1676. doi:10.1093/bioinformatics/btv033.
5. D. Hernandez, P. François, L. Farinelli, M. Østerås, J. Schrenzel, De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer, *Genome Research*. 18 (2008) 802–809. doi:10.1101/gr.072033.107.
6. J. Butler, I. MacCallum, M. Kleber, I.A. Shlyakhter, M.K. Belmonte, E.S. Lander, et al., Allpaths: De Novo Assembly of whole-genome shotgun microreads, *Genome Research*. 18 (2008) 810–820. doi:10.1101/gr.7337908.
7. K. Paszkiewicz, D.J. Studholme, De Novo Assembly of short sequence reads, *Briefings in Bioinformatics*. 11 (2010) 457–472. doi:10.1093/bib/bbq020.
8. A.S. Gaia, P.H. de Sá, M.S. de Oliveira, A.A. Veras, Ngsreadstreatment – a cuckoo filter-based tool for removing duplicate reads in NGS Data, *Scientific Reports*. 9 (2019). doi:10.1038/s41598-019-48242-w.
9. E.W. Myers Jr, A history of DNA sequence assembly, *It - Information Technology*. 58 (2016) 126–132. doi:10.1515/itit-2015-0047.
10. R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads, *PLOS Computational Biology*. 13 (2017). doi:10.1371/journal.pcbi.1005595.
11. E. Haghshenas, H. Asghari, J. Stoye, C. Chauve, F. Hach, HASLR: Fast hybrid assembly of long reads, *iScience*. 23 (2020) 101389. doi:10.1016/j.isci.2020.101389.
12. A.V. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, J.A. Yorke, The MASURCA genome assembler, *Bioinformatics*. 29 (2013) 2669–2677. doi:10.1093/bioinformatics/btt476.
13. C. Ye, C.M. Hill, S. Wu, J. Ruan, Z. Ma, DBG2OLC: Efficient Assembly of large genomes using long erroneous reads of the Third Generation Sequencing Technologies, *Scientific Reports*. 6 (2016). doi:10.1038/srep31900.

14. V. Deshpande, E.D. Fung, S. Pham, V. Bafna, Cerulean: A hybrid assembly using high throughput short and long reads, *Lecture Notes in Computer Science*. (2013) 349–363. doi:10.1007/978-3-642-40453-5_27.
15. Cloud applications and cloud platform, Oracle. (Internet) (accessed July 22, 2023). Available in: <https://www.oracle.com/>.
16. SQLite Home Page. (Internet) (accessed July 22, 2023). Available in: <https://www.sqlite.org/>.
17. B. Porter, J. van Zyl, O. Lamy, Welcome to Apache Maven, Maven. (Internet) (accessed July 22, 2023). <https://maven.apache.org/>.
18. S.-H. Lin, Y.-C. Liao, CISA: Contig Integrator for Sequence Assembly of bacterial genomes, *PLoS ONE*. 8 (2013). doi:10.1371/journal.pone.0060843.
19. A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: Multiple alignment of conserved genomic sequence with rearrangements, *Genome Research*. 14 (2004) 1394–1403. doi:10.1101/gr.2289704.
20. M. Manni, M.R. Berkeley, M. Seppey, F.A. Simão, E.M. Zdobnov, Busco update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, *Molecular Biology and Evolution*. 38 (2021) 4647–4654. doi:10.1093/molbev/msab199.
21. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, Quast: Quality Assessment Tool for Genome Assemblies, *Bioinformatics*. 29 (2013) 1072–1075. doi:10.1093/bioinformatics/btt086.
22. T. Seemann, Prokka: Rapid Prokaryotic Genome Annotation, *Bioinformatics*. 30 (2014) 2068–2069. doi:10.1093/bioinformatics/btu153.
23. National Center for Biotechnology Information. (Internet) (accessed July 22, 2023). Available in: <https://www.ncbi.nlm.nih.gov/>.
24. Nucleotide - NCBI, National Center for Biotechnology Information. (Internet) (accessed July 22, 2023). Available in: <https://www.ncbi.nlm.nih.gov/nucleotide>.

APÊNDICE

Tabela 1. Organismos usados para validar a ferramenta.

Número SRA	Nome do Organismo	Biblioteca	Tamanho do Genoma	Acesso do GenBank	Plataforma de Sequenciamento	Ano da Plataforma de sequenciamento	Informação de referência
ERR007646	<i>Escherichia coli</i> 042	Paired-end	4692707	CP042934.2	Illumina Genome Analyzer	2006	Illumina
ERR351258	<i>Escherichia Coli</i> O111:H - str. 11128	Paired-end	5575231	AP010960.1	Illumina HiSeq 2000	2010	Illumina
ERR351259	<i>Escherichia Coli</i> O26:H11 str. 11368	Paired-end	5697240	AP010953.1	Illumina HiSeq 2000	2010	Illumina
SRR3465539	<i>Escherichia coli</i> strain Eco889	Paired-end	5012253	CP015159.1	Illumina MiSeq	2018	Illumina
SRR3999078	<i>Escherichia coli</i> Ecol_545	Paired-end	5031843	CP018976.1	Illumina MiSeq	2018	Illumina
SRR3999096	<i>Escherichia coli</i> Ecol_AZ146	Paired-end	5434745	CP018991.1	Illumina MiSeq	2018	Illumina
SRR5194991	<i>Escherichia coli</i> O25b:H4-ST131	Paired-end	5109767	HG941718.1	Illumina HiSeq 2500	2015	Illumina
ERR351260	<i>Escherichia Coli</i> O103:H2 str. 120009	Paired-end	5449314	AP010958.1	Illumina HiSeq 2000	2010	Illumina
SRR933455	<i>Escherichia coli</i> strain ZH193	Paired-end	5046610	CP014497.1	Illumina Genome Analyzer IIx	2009	Illumina
SRR10971019	<i>Escherichia coli</i> str. K-12 substr. MG1655	Single-end	4641652	U00096.3	Sequel	2017	PacBio
SRR638977	<i>Corynebacterium glutamicum</i> strain ATCC 13032	Single-end	3316624	CP025533.1	Ion Torrent PGM	2017	ThermoFisher
SRR3947906	<i>Corynebacterium doosanense</i> CAU 212 = DSM 45436	Paired-end	2671798	CP006764.1	Illumina HiSeq 2000	2010	Illumina
DRR174983	<i>Corynebacterium ulcerans</i> 0211	Paired-end	2579078	AP019662.1	Illumina Genome Analyzer IIx	2009	Illumina
SRR12270040	<i>Corynebacterium diphtheriae</i> bv. <i>mitis</i> strain PC0646	Paired-end	2386265	CP040557.1	Illumina MiSeq	2018	Illumina
SRR013523	<i>Corynebacterium matruchotii</i> strain ATCC 14266	Single-end	2866540	CP050134.2	454 GS FLX	2008	Roche

Número SRA	Nome do Organismo	Biblioteca	Tamanho do Genoma	Acesso do GenBank	Plataforma de Sequenciamento	Ano da Plataforma de sequenciamento	Informação de referência
SRR3312980	<i>Corynebacterium pseudotuberculosis strain 262</i>	Single-end	2361125	CP080542.1	Ion Torrent PGM	2017	ThermoFisher
SRR3932423	<i>Corynebacterium striatum strain 215</i>	Paired-end	2993983	CP024931.1	Illumina MiSeq	2018	Illumina
DRR249160	<i>Mycobacterium tuberculosis H37Rv</i>	Paired-end	4411709	CP003248.2	Illumina Genome Analyzer IIx	2009	Illumina
SRR974839	<i>Mycobacterium tuberculosis F11</i>	Paired-end	4424435	CP000717.1	Illumina HiSeq 2000	2010	Illumina
DRR161272	<i>Mycobacterium kansasii strain JCM 13573</i>	Single-end	5958133	CP065047.1	MinION	2015	Nanopore Tech
ERR5104572	<i>Mycobacterium canettii CIPT 140010059</i>	Paired-end	4482059	HE572590.1	Illumina HiSeq 2500	2015	Illumina
DRR255737	<i>Mycobacterium paraintracellulare M001</i>	Paired-end	5739349	AP024240.1	Illumina MiSeq	2018	Illumina
DRR286167	<i>Mycobacterium avium subsp. hominissuis strain OCU464</i>	Paired-end	5178488	CP009360.4	Illumina MiSeq	2018	Illumina
DRR161197	<i>Mycobacterium simiae JCM 12377</i>	Paired-end	5788994	AP022568.1	Illumina HiSeq 2500	2015	Illumina
DRR259154	<i>Mycobacterium heckeshornense JMUB5695</i>	Single-end	4865109	AP024310.1	MinION	2015	Nanopore Tech
SRR12998338	<i>Klebsiella variicola strain KPN029</i>	Paired-end	5573106	CP065162.1	Illumina NovaSeq 6000	2017	Illumina
DRR223366	<i>Klebsiella pneumoniae SMKP03</i>	Paired-end	5155320	AP023148.1	Illumina MiSeq	2018	Illumina
SRR2965748	<i>Klebsiella aerogenes strain CAV1320</i>	Paired-end	5124987	CP011574.1	Illumina HiSeq 2000	2010	Illumina
DRR199345	<i>Klebsiella aerogenes WP5-W18-CRE-01</i>	Paired-end	5135838	AP022108.1	Illumina NextSeq 500	2014	Illumina
SRR12233579	<i>Klebsiella africana strain FF1003</i>	Paired-end	5291121	CP059391.1	Illumina NextSeq 500	2014	Illumina
SRR13355318	<i>Klebsiella quasipneumoniae strain KqPF26</i>	Paired-end	5242686	CP065838.1	Illumina NextSeq 500	2014	Illumina
SRR5386043	<i>Klebsiella quasivariicola strain KPN1705</i>	Paired-end	5540188	CP022823.1	Illumina NextSeq 500	2014	Illumina
DRR286070	<i>Klebsiella variicola JCM 12419</i>	Single-end	5564085	AP024592.1	MinION	2015	Nanopore Tech
SRR6442873	<i>Klebsiella variicola strain WCHKV030666</i>	Paired-end	5553346	CP027064.2	Illumina HiSeq X Ten	2016	Illumina

Número SRA	Nome do Organismo	Biblioteca	Tamanho do Genoma	Acesso do GenBank	Plataforma de Sequenciamento	Ano da Plataforma de sequenciamento	Informação de referência
DRR199814	<i>Klebsiella michiganensis</i> THO-011	Paired-end	5935402	AP022547.1	Illumina NextSeq 500	2014	Illumina
SRR7945280	<i>Rhodococcus rhodochrous</i> strain ATCC BAA870	Paired-end	5370536	CP032675.1	Illumina HiSeq 2500	2010	Illumina
SRR16168733	<i>Rhodococcus qingshengii</i> strain VER34	Paired-end	6286261	CP085042.1	Illumina NextSeq 500	2014	Illumina
SRR15616889	<i>Rhodococcus opacus</i> strain DSM 44186	Paired-end	8044513	CP082160.1	Illumina HiSeq 4000	2015	Illumina
SRR10609577	<i>Rhodococcus ruber</i> strain R1	Paired-end	5352622	CP038030.2	Illumina MiSeq	2018	Illumina
SRR5056410	<i>Salmonella enterica</i> strain SA20021456	Paired-end	1477517	CP030219.1	Illumina MiSeq	2018	Illumina
SRR5056619	<i>Salmonella enterica</i> strain SA20100201	Paired-end	5195044	CP030180.1	Illumina MiSeq	2018	Illumina
SRR17276215	<i>Salmonella enterica</i> strain SZL 38	Paired-end	5052615	CP085987.1	Illumina MiSeq	2018	Illumina
SRR12342991	<i>Salmonella enterica</i> strain SLR1_8094	Paired-end	5017156	CP060515.1	Illumina HiSeq 2500	2015	Illumina
SRR12170550	<i>Salmonella enterica</i> strain SRC27	Paired-end	4865373	CP058807.1	Illumina MiSeq	2018	Illumina
SRR5581941	<i>Salmonella enterica</i> strain 85-0120	Paired-end	4794154	CP054715.1	Illumina HiSeq 2500	2015	Illumina
SRR10394909	<i>Pseudomonas putida</i> strain MX-2	Paired-end	5952420	NZ_CP046872.1	Illumina HiSeq 4000	2015	Illumina
DRR296284	<i>Pseudomonas izuensis</i>	Single-end	6732290	AP017423.2	Sequel	2017	PacBio
SRR9841424	<i>Pseudomonas aeruginosa</i> strain HOU1	Paired-end	6123373	NZ_CP042269.1	Illumina MiSeq	2018	Illumina
SRR18890661	<i>Pseudomonas aeruginosa</i> strain PAO1-N	Single-end	6217942	NZ_CP099797.1	Illumina HiSeq 2000	2010	Illumina
SRR9435726	<i>Bacillus thuringiensis</i> serovar konkukian str. 97-27	Single-end	5237682	AE017355.1	AB 310 Genetic Analyzer	2004	ThermoFisher
SRR4187667	<i>Bacillus velezensis</i> strain CBMB205	Paired-end	3929792	CP011937.1	Illumina HiSeq 2500	2015	Illumina
SRR20036353	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	Paired-end	4215619	CP010052.1	Illumina NovaSeq 6000	2017	Illumina
SRR572717	<i>Bacillus altitudinis</i> strain BA06	Paired-end	3755709	CP039020.1	Illumina HiSeq 2000	2010	Illumina

Número SRA	Nome do Organismo	Biblioteca	Tamanho do Genoma	Acesso do GenBank	Plataforma de Sequenciamento	Ano da Plataforma de sequenciamento	Informação de referência
SRR9052983	<i>Streptococcus sp. 1643</i>	Paired-end	1898954	NZ_CP040231.1	Illumina HiSeq 2500	2015	Illumina
SRR9641535	<i>Streptococcus salivarius strain FDAARGOS_771</i>	Paired-end	1999501	NZ_CP053998.1	Illumina HiSeq 4000	2015	Illumina
SRR5868502	<i>Streptococcus gordonii strain FDAARGOS_371</i>	Paired-end	2121309	NZ_CP023511.1	Illumina HiSeq 4000	2015	Illumina
SRR16999781	<i>Nocardia huaxiensis strain BCHNH01</i>	Paired-end	8314373	CP088007.1	Illumina NovaSeq 6000	2017	Illumina
SRR9592370	<i>Bacteroides ovatus strain 3725 D1 iv</i>	Paired-end	7079631	NZ_CP041395.1	Illumina HiSeq 2500	2015	Illumina
SRR11176973	<i>Acinetobacter baumannii strain K09-14</i>	Paired-end	3972439	CP043953.1	Illumina HiSeq 4000	2015	Illumina
SRR10020330	<i>Corynebacterium diphtheriae CN2000</i>	Paired-end	2530683	CP039522.1	Illumina MiSeq	2018	Illumina
SRR231104	<i>Mycobacterium colombiense CECT 3035</i>	Single-end	5581643	CP020821.1	454 GS FLX Titanium	2009	Roche

Fonte: Autoria própria.

Tabela 3. Análise comparativa dos resultados da montagem do GenTreat com as ferramentas SPAdes, Megahit e Velvet.

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
ERR007646	Megahit	84.559	660	23406	13993	200	102137	5160692	85
	SPAdes	85.086	1303	76266	30375	34	226206	5282686	103
	Velvet	66.720	1432	4051	2098	100	37558	4138672	43
	GenTreat	81.371	99	83733	26471	434	226151	4839279	101
ERR351258	Megahit	95.410	454	96036	95539	202	344095	5523789	15
	SPAdes	93.730	1575	95600	95459	56	290627	5408843	1

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	Velvet	90.752	1409	20271	18842	100	90766	5239974	0
	GenTreat	94.653	393	99798	97167	105	374185	5485734	12
ERR351259	Megahit	93.994	534	92062	91649	200	244184	5535631	12
	SPAdes	92.327	1567	100938	86997	56	228430	5405404	0
	Velvet	87.405	2055	19745	16644	100	68863	5106364	0
	GenTreat	93.390	202	102625	101863	135	249984	5544310	11
ERR351260	Megahit	95.182	2027	80602	79080	208	250061	5587299	8
	SPAdes	93.681	5540	77758	80807	56	249905	5869005	2
	Velvet	90.674	1371	19803	17513	100	120100	5017712	0
	GenTreat	93.895	631	94245	94245	223	250061	5634237	7
SRR933455	Megahit	98.458	801	131613	123872	201	308337	5696931	15
	SPAdes	96.996	2303	106476	109208	56	314036	5527484	3
	Velvet	95.025	1685	18024	17262	100	82292	4920108	0
	GenTreat	94.578	113	144119	139366	2016	314036	5427025	8
SRR3465539	Megahit	99.225	18546	94969	127216	200	400244	6827490	70
	SPAdes	98.823	664	402167	402167	128	912119	5263312	1
	Velvet	93.786	3103	4067	3866	100	19096	4866537	0

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	GenTreat	95.676	122	359847	359847	378	912119	5092764	8
SRR3999078	Megahit	98.815	35678	67535	130043	200	455011	6914327	81
	SPAdes	98.683	715	196167	198618	128	490217	5295213	1
	Velvet	94.916	1673	10548	10374	100	47311	4987693	0
	GenTreat	98.022	59	208205	207993	2918	528881	5252057	5
SRR3999096	Megahit	98.978	52718	68312	121515	128	312442	7792682	91
	SPAdes	98.677	685	309340	268871	200	578310	5687429	3
	Velvet	94.903	1601	10721	10412	100	41015	5397035	0
	GenTreat	97.566	192	312489	268871	158	592843	5713273	15
SRR5194991	Megahit	98.532	193	186020	175149	213	334080	5180465	5
	SPAdes	98.173	1104	180433	180433	56	476971	5150731	0
	Velvet	96.285	810	27505	26939	100	107926	5049145	0
	GenTreat	98.568	170	191169	191169	106	477804	5301039	3
SRR10971019	Megahit	98.990	30838	100414	133182	200	285273	6308120	16
	SPAdes	99.994	2	4642144	3815627	128	4642144	4642144	2
	Velvet	35.850	13529	758	-	100	3686	1665961	0
	GenTreat	99.994	14	4642256	3818549	302	4642256	4856726	2

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
DRR174983	Megahit	99.303	6287	63952	219568	202	844964	4779121	3
	SPAdes	99.305	10999	378618	857309	56	885994	5066822	0
	Velvet	72.208	4935	1174	887	100	4579	1904573	0
	GenTreat	99.312	94	857407	857309	266	891864	2801574	0
SRR013523	Megahit	99.300	61	191372	191282	215	588728	2854528	1
	SPAdes	99.375	63	298034	298028	129	685539	2853326	1
	Velvet	94.450	949	7193	6771	100	45946	2711955	1
	GenTreat	99.389	31	298048	298048	315	685539	2858774	2
SRR638977	Megahit	97.705	807	16161	15916	274	56718	3328950	43
	SPAdes	97.820	1210	112536	111156	81	258165	3325568	7
	Velvet	95.479	1279	4976	4792	100	20216	3175741	4
	GenTreat	96.939	387	112653	111156	273	227269	3339034	20
SRR3932423	Megahit	96.366	1033	96761	95313	200	253681	2914129	16
	SPAdes	96.407	265	150657	150654	128	274446	2890771	1
	Velvet	93.795	356	36234	35260	100	251384	2809228	0
	GenTreat	95.634	340	150781	150570	125	275107	3021431	8
SRR3947906	Megahit	98.368	966	54969	54639	78	145580	2703932	27

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	SPAdes	98.263	1606	191497	191497	200	593299	2658112	0
	Velvet	25.535	9265	674	-	100	3280	683820	0
	GenTreat	90.478	301	169874	169874	209	421557	2671911	12
SRR12270040	Megahit	98.108	144	134496	134496	206	321789	2352152	2
	SPAdes	98.152	121	179515	179515	128	339972	2351444	0
	Velvet	96.950	213	46287	46287	100	112368	2319880	0
	GenTreat	93.028	126	185676	179521	268	339998	2229730	2
SRR3312980	Megahit	99.135	259	21156	21156	227	56349	2378304	13
	SPAdes	99.131	69	629395	628605	132	803408	2345559	0
	Velvet	51.016	5832	834	510	100	3321	1208345	0
	GenTreat	98.683	14	410680	365587	274	629395	2335107	1
DRR199345	Megahit	98.982	90	219485	222758	200	593062	5405644	0
	SPAdes	98.933	178	398129	398029	78	1070432	5401138	0
	Velvet	97.972	463	33353	31846	100	125834	5344289	0
	GenTreat	95.831	19	446223	480991	55440	1070557	5495565	1
DRR199814	Megahit	98.903	137	291906	279274	206	655454	5980990	5
	SPAdes	98.926	666	230872	230544	78	659922	5982252	1

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	Velvet	97.889	537	33205	33139	100	125501	5916826	0
	GenTreat	89.353	21	340397	311594	99504	709253	6184262	3
DRR223366	Megahit	99.070	4402	124011	123799	200	351193	5460223	24
	SPAdes	99.033	172	388413	316024	128	574907	5358885	2
	Velvet	97.810	637	26080	26080	100	95681	5234987	0
	GenTreat	97.495	128	378840	307675	226	574966	5319569	16
DRR286070	Megahit	98.807	20397	1943	3557	200	20326	18757074	86
	SPAdes	94.402	24	484342	420404	128	991762	5407998	6
	Velvet	52.670	11999	1182	550	100	9468	2978687	0
	GenTreat	89.324	35	537355	503051	2662	1037960	5293032	6
SRR2965748	Megahit	99.162	179	249899	249674	202	693804	5116767	1
	SPAdes	99.133	1664	503867	503867	78	1041695	5145833	0
	Velvet	98.021	414	39131	38240	100	150555	5039577	0
	GenTreat	93.918	10	693614	580801	193144	1243674	5333965	1
SRR5386043	Megahit	97.339	223	128347	131734	200	435954	5789881	5
	SPAdes	97.058	524	134029	149084	56	621294	5755339	0
	Velvet	96.150	902	29534	30244	100	128977	5648218	0

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	GenTreat	96.376	58	157106	157106	11762	730724	5691622	5
SRR6442873	Megahit	99.176	1162	245617	271524	233	988362	5967064	2
	SPAdes	99.096	15188	437393	448925	78	682208	6271383	0
	Velvet	98.176	587	36729	37353	100	141198	5676626	0
	GenTreat	99.139	32	489543	489543	295	1316678	5776583	1
SRR12233579	Megahit	98.971	128	268786	268786	203	517987	5369339	1
	SPAdes	98.830	188	346863	346863	78	1043015	5362055	0
	Velvet	98.034	441	31865	30725	100	148926	5313291	0
	GenTreat	99.071	68	434774	480976	155	1043143	5953891	2
SRR12998338	Megahit	98.751	182	170653	192190	200	970771	6152956	2
	SPAdes	98.622	984	306022	305813	78	1129470	6139564	0
	Velvet	97.700	633	34331	34214	100	128791	6047088	0
	GenTreat	95.626	27	313515	328793	30694	1129534	6080151	2
SRR13355318	Megahit	99.305	199	175027	208563	230	334391	5369430	6
	SPAdes	99.198	123	375850	375850	78	990790	5351248	0
	Velvet	98.231	708	16208	15714	100	61775	5303067	0
	GenTreat	94.601	18	375850	375850	76189	990790	5705455	1

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
DRR161197	Megahit	99.330	104	594381	594381	215	1176307	5755097	0
	SPAdes	99.313	259	798759	798759	56	1261049	5751674	0
	Velvet	98.512	470	43212	42504	100	180430	5706459	0
	GenTreat	99.352	40	1261190	1261190	81	1712624	5755110	1
DRR161272	Megahit	1.192	13917	970	-	301	15928	2594774	0
	SPAdes	10.128	12	70990	-	128	92072	646612	4
	Velvet	29.887	18331	704	-	100	2123	1789618	0
	GenTreat	2.619	11295	951	-	97	44233	2536615	0
DRR249160	Megahit	95.309	313	36928	31572	207	137199	4232318	25
	SPAdes	95.283	349	41929	37185	56	212612	4228445	33
	Velvet	93.840	782	15010	14092	100	60636	4164335	11
	GenTreat	95.350	273	41917	36789	102	212612	4234495	37
DRR255737	Megahit	98.895	4751	101564	102201	201	302101	5753807	7
	SPAdes	98.821	296	114827	105904	128	302159	5718869	2
	Velvet	96.643	983	19045	18313	100	56345	5553460	0
	GenTreat	92.523	1055	115569	102409	98	302154	5588856	8
DRR259154	Megahit	99.269	39337	1957	3612	200	20983	39182617	40

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	SPAdes	29.379	26	79313	-	128	134510	1434028	5
	Velvet	85.669	4838	1918	1646	100	9516	4187114	0
	GenTreat	48.253	250	15214	-	803	158487	3644370	6
DRR286167	Megahit	99.350	77467	45637	70198	200	335776	7433653	400
	SPAdes	98.698	2924	175130	189726	128	393340	5360637	2
	Velvet	95.863	1871	6141	5902	100	33806	5064506	0
	GenTreat	98.411	160	175143	175143	112	397827	5366050	44
ERR5104572	Megahit	97.824	224	84650	73126	201	244161	4394031	3
	SPAdes	97.540	334	117992	117992	56	256869	4376046	2
	Velvet	94.771	941	20899	20019	100	78467	4253499	1
	GenTreat	97.942	100	121230	118250	101	257210	4438926	5
SRR974839	Megahit	96.210	866	17917	17459	200	89286	4311818	117
	SPAdes	95.744	602	18357	17068	56	105928	4272021	23
	Velvet	89.224	2171	4554	3938	100	21572	3953290	0
	GenTreat	93.109	528	22166	19639	106	106336	4172050	67
SRR7945280	Megahit	99.187	7160	261708	263423	200	881138	5951984	17
	SPAdes	99.173	195	857710	857397	128	1215560	5914941	1

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	Velvet	97.512	971	25286	26074	100	107164	5765527	0
	GenTreat	98.788	14	857842	857528	36881	1215573	5811144	1
SRR16168733	Megahit	99.928	119	203978	224751	56	796291	6575612	2
	SPAdes	99.984	481	645125	658871	224	1340023	6591167	2
	Velvet	98.213	2150	5873	5886	100	30853	6447968	1
	GenTreat	99.026	16	658896	658896	43535	1340109	6295568	2
SRR15616889	Megahit	96.528	2618	9448	5662	200	416255	11902301	13
	SPAdes	96.846	4247	11995	6891	56	414662	11918791	19
	Velvet	4.843	1523	75468	-	100	174783	3959080	1
	GenTreat	58.798	478	22048	7091	403	416255	8702866	14
SRR10609577	Megahit	99.421	799	48573	50603	200	208670	5553029	3
	SPAdes	99.294	414	62996	66354	128	208911	5547747	0
	Velvet	97.147	1194	11574	11353	100	73133	5386890	0
	GenTreat	97.489	547	71228	76528	149	208804	5584921	3
SRR5056410	Megahit	99.190	9499	153510	115485	200	459196	5769280	5
	SPAdes	99.386	160	321097	295186	128	732785	5645413	0
	Velvet	94.045	1285	30784	27352	100	144769	5421328	0

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	GenTreat	1.801	3	546521	-	394327	732799	1673647	0
SRR5056619	Megahit	98.096	2133	80619	80586	200	284221	5266856	29
	SPAdes	97.851	382	97332	97174	128	318992	5182404	2
	Velvet	92.138	801	17841	15356	100	63804	4795900	0
	GenTreat	92.103	590	97327	90332	115	334793	4945619	11
SRR17276215	Megahit	98.903	5169	208254	208254	200	389574	5191829	7
	SPAdes	98.936	1867	282782	282782	128	582171	5165805	1
	Velvet	97.296	1210	32650	31939	100	121813	5004568	0
	GenTreat	99.022	435	282809	282809	105	582198	5514098	4
SRR12342991	Megahit	98.759	279	143377	149900	202	775578	775578	6
	SPAdes	98.897	171	224389	267025	131	623959	623959	0
	Velvet	97.330	643	30207	29547	100	116739	116739	0
	GenTreat	93.467	28	267134	267038	57188	775578	775578	3
SRR12170550	Megahit	98.915	803	149628	182701	205	600218	5303574	4
	SPAdes	98.982	1305	222888	225813	128	521451	5461308	0
	Velvet	97.539	674	25385	24009	100	96010	5126826	0
	GenTreat	96.317	26	257746	257746	35884	600218	5317920	3

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
SRR5581941	Megahit	97.109	970	88831	91696	200	233740	5018141	4
	SPAdes	97.196	545	117993	123554	128	335555	5052857	2
	Velvet	95.350	636	32360	32360	100	114284	4849993	0
	GenTreat	97.245	441	123569	134676	273	335677	5269253	5
DRR296284	Megahit	99.521	11888	277543	341301	200	544609	10838953	9
	SPAdes	100.000	40	6732418	6732418	203	6732418	6744667	0
	Velvet	88.502	6575	1792	1610	100	13426	6655690	0
	GenTreat	100.000	11	6732432	6732432	146	6732432	6879464	1
SRR9841424	Megahit	99.099	120	210806	210806	200	475096	6255264	1
	SPAdes	99.103	244	301731	301731	78	665805	6280738	1
	Velvet	97.651	831	26867	26769	100	123125	6216122	0
	GenTreat	94.063	22	412270	379725	74016	672856	6273954	1
SRR10394909	Megahit	98.185	420	114474	108644	200	286188	6012756	14
	SPAdes	98.123	1583	201514	201514	78	587231	6293654	0
	Velvet	96.042	1182	11586	11030	100	63210	5853773	0
	GenTreat	95.579	433	203320	235454	73	595724	6675875	9
SRR18890661	Megahit	98.067	1350	7942	7792	237	33165	6192605	11

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	SPAdes	98.562	570	80945	80942	56	286694	6201096	1
	Velvet	97.656	1107	15322	14896	100	86236	6164041	2
	GenTreat	97.716	205	83187	79317	144	287919	6093799	4
SRR572717	Megahit	99.008	37	554749	554749	207	997579	3723233	1
	SPAdes	98.758	214	278368	278368	56	833724	3730624	0
	Velvet	98.138	212	42656	42338	100	133794	3697403	0
	GenTreat	99.007	33	565990	565990	113	998267	3722915	1
SRR4187667	Megahit	98.737	92	581066	581066	228	1011980	3916481	0
	SPAdes	98.733	3593	2059719	2059719	78	2059719	4855461	0
	Velvet	98.154	221	81391	79813	100	273170	3876594	0
	GenTreat	99.006	1387	2060398	2060398	258	2060398	4310845	1
SRR9435726	Megahit	93.572	756	22129	21881	209	108452	5136062	4
	SPAdes	94.325	257	46424	43391	129	164561	5034956	3
	Velvet	31.978	1219	2828	-	100	18708	1783597	1
	GenTreat	93.731	308	46616	46450	120	164572	5107541	3
SRR20036353	Megahit	11.236	4896	590	-	217	3357	2075779	189
	SPAdes	1.181	9135	861	-	77	1471	1402039	6

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	Velvet	1.923	13540	576	-	100	1593	2356749	0
	GenTreat	5.673	4266	597	-	72	3294	1426671	92
SRR5868502	Megahit	98.903	298	507830	291330	200	1068595	2244910	8
	SPAdes	99.020	2772	638211	552761	78	1074161	2875382	3
	Velvet	86.669	2240	1740	1512	100	7207	2102340	1
	GenTreat	99.057	413	638275	555151	72	1074225	2293920	5
SRR9641535	Megahit	98.211	516	145789	145789	202	429585	2235767	10
	SPAdes	98.061	3004	193901	193805	78	516293	2845530	0
	Velvet	85.535	2034	1676	1464	100	6633	1952360	0
	GenTreat	98.227	214	174029	194591	78	516293	2314977	7
SRR9052983	Megahit	98.818	49	1044659	250158	207	1044659	1888967	3
	SPAdes	98.848	75	1055027	1054635	78	1055027	1891738	0
	Velvet	97.493	179	35466	30617	100	153784	1867324	0
	GenTreat	67.949	2	1055091	1054699	1044659	1055091	2099750	3
SRR16999781	Megahit	99.801	125	344481	344395	254	1349969	8335039	1
	SPAdes	99.782	360	778975	778862	78	1575507	8387996	0
	Velvet	98.489	1372	18371	18072	100	83892	8305901	0

SRA	Montador	Fração do Genoma (%)	Contigs	N50	NGA50	Contig Mínimo	Maior Contig	Comprimento Total	Misassemblies
	GenTreat	99.806	342	881287	881148	121	1575484	8397185	1
SRR11176973	Megahit	98.706	1422	252011	279887	201	252011	279887	2
	SPAdes	98.699	5281	298086	371636	78	298086	371636	0
	Velvet	97.176	482	29828	29528	100	29828	29528	0
	GenTreat	98.732	163	412461	412461	81	412461	412461	1
SRR9592370	Megahit	97.660	179	129917	129917	200	469781	7097003	5
	SPAdes	97.384	6866	68002	67415	78	200200	7689079	2
	Velvet	95.631	1167	25706	24974	100	118679	7034490	0
	GenTreat	97.418	350	149791	140799	80	583937	7178737	5
SRR10020330	Megahit	96.592	64	123733	94348	214	242044	2463926	3
	SPAdes	96.435	543	105391	99837	78	244267	2584812	2
	Velvet	95.196	263	36833	34276	100	99498	2443685	1
	GenTreat	96.622	512	123733	99955	228	244331	2618433	6
SRR231104	Megahit	99.884	8615	148334	176621	200	475799	9081554	42
	SPAdes	99.878	42	658658	658459	128	977945	5585348	0
	Velvet	97.963	1544	7821	7651	100	34397	5582559	0
	GenTreat	99.846	353	658694	658495	95	978873	6101387	23

Fonte: Autoria propria.

Tabela 4. Análise de completude gênica dos resultados de montagem das ferramentas Velvet e GenTreat usando o software BUSCO. Onde C – BUSCOs Completos; S – BUSCOs Completos e de Cópia Única; D – BUSCOs Completos e Duplicados; F – BUSCOs Fragmentados; M – BUSCOs Ausentes; n – Total de grupos BUSCO pesquisados.

SRA Number	Velvet	GenTreat
DRR161197	C:97.1% [S:96.8%,D:0.3%],F:1.9%,M:1.0%,n:743	C:99.2% [S:98.9%,D:0.3%],F:0.4%,M:0.4%,n:743
DRR161272	C:3.2% [S:3.2%,D:0.0%],F:44.4%,M:52.4%,n:124	C:0.5% [S:0.5%,D:0.0%],F:0.0%,M:99.5%,n:194
DRR174983	C:36.2% [S:36.2%,D:0.0%],F:36.6%,M:27.2%,n:743	C:99.1% [S:99.1%,D:0.0%],F:0.4%,M:0.5%,n:743
DRR199345	C:98.7% [S:98.2%,D:0.5%],F:0.2%,M:1.1%,n:440	C:88.8% [S:82.7%,D:6.1%],F:0.0%,M:11.2%,n:440
DRR199814	C:96.0% [S:95.5%,D:0.5%],F:3.0%,M:1.0%,n:440	C:84.8% [S:73.2%,D:11.6%],F:0.0%,M:15.2%,n:440
DRR223366	C:96.9% [S:96.4%,D:0.5%],F:1.8%,M:1.3%,n:440	C:98.0% [S:97.3%,D:0.7%],F:0.2%,M:1.8%,n:440
DRR249160	C:97.4% [S:97.0%,D:0.4%],F:2.2%,M:0.4%,n:743	C:98.2% [S:97.7%,D:0.5%],F:1.1%,M:0.7%,n:743
DRR255737	C:97.2% [S:96.8%,D:0.4%],F:2.0%,M:0.8%,n:743	C:95.4% [S:94.2%,D:1.2%],F:0.3%,M:4.3%,n:743
DRR259154	C:16.9% [S:16.9%,D:0.0%],F:54.0%,M:29.1%,n:124	C:1.6% [S:1.6%,D:0.0%],F:9.7%,M:88.7%,n:124
DRR286070	C:22.0% [S:21.8%,D:0.2%],F:30.7%,M:47.3%,n:440	C:37.5% [S:37.3%,D:0.2%],F:29.1%,M:33.4%,n:440
DRR286167	C:74.2% [S:74.2%,D:0.0%],F:16.3%,M:9.5%,n:743	C:100.0% [S:99.2%,D:0.8%],F:0.0%,M:0.0%,n:124
ERR007646	C:67.0% [S:66.8%,D:0.2%],F:15.7%,M:17.3%,n:440	C:95.0% [S:94.5%,D:0.5%],F:0.0%,M:5.0%,n:440
ERR351258	C:99.4% [S:98.9%,D:0.5%],F:0.7%,M:0.1%,n:440	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440
ERR351259	C:99.6% [S:99.1%,D:0.5%],F:0.5%,M:0.1%,n:440	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
ERR351260	C:99.4% [S:98.9%,D:0.5%],F:0.7%,M:0.1%,n:440	C:100.0% [S:97.3%,D:2.7%],F:0.0%,M:0.0%,n:440
ERR5104572	C:99.3% [S:98.9%,D:0.4%],F:0.3%,M:0.4%,n:743	C:99.6% [S:98.4%,D:1.2%],F:0.0%,M:0.4%,n:743
SRR013523	C:86.9% [S:86.8%,D:0.1%],F:7.5%,M:5.6%,n:743	C:96.9% [S:96.8%,D:0.1%],F:1.2%,M:1.9%,n:743
SRR231104	C:65.7% [S:65.4%,D:0.3%],F:20.1%,M:14.2%,n:743	C:99.3% [S:92.2%,D:7.1%],F:0.3%,M:0.4%,n:743
SRR638977	C:60.0% [S:59.9%,D:0.1%],F:26.2%,M:13.8%,n:743	C:86.0% [S:84.5%,D:1.5%],F:8.7%,M:5.3%,n:743
SRR933455	C:97.1% [S:96.6%,D:0.5%],F:1.8%,M:1.1%,n:440	C:97.7% [S:92.0%,D:5.7%],F:0.5%,M:1.8%,n:440
SRR974839	C:73.5% [S:73.4%,D:0.1%],F:18.2%,M:8.3%,n:743	C:92.4% [S:91.9%,D:0.5%],F:4.3%,M:3.3%,n:743
SRR2965748	C:98.4% [S:98.2%,D:0.2%],F:0.5%,M:1.1%,n:440	C:91.1% [S:83.6%,D:7.5%],F:0.2%,M:8.7%,n:440
SRR3312980	C:22.3% [S:22.3%,D:0.0%],F:34.2%,M:43.5%,n:743	C:88.6% [S:88.6%,D:0.0%],F:7.3%,M:4.1%,n:743
SRR3465539	C:72.2% [S:72.0%,D:0.2%],F:23.6%,M:4.2%,n:440	C:96.0% [S:95.5%,D:0.5%],F:0.0%,M:4.0%,n:440
SRR3932423	C:96.8% [S:96.1%,D:0.7%],F:2.2%,M:1.0%,n:743	C:96.8% [S:91.0%,D:5.8%],F:0.5%,M:2.7%,n:743
SRR3947906	C:8.3% [S:8.3%,D:0.0%],F:32.3%,M:59.4%,n:743	C:90.6% [S:80.8%,D:9.8%],F:0.5%,M:8.9%,n:743
SRR3999078	C:92.3% [S:91.8%,D:0.5%],F:6.8%,M:0.9%,n:440	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440
SRR3999096	C:88.9% [S:88.2%,D:0.7%],F:9.8%,M:1.3%,n:440	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
SRR5056410	C:98.5% [S:98.0%,D:0.5%],F:0.9%,M:0.6%,n:440	C:17.7% [S:17.7%,D:0.0%],F:0.8%,M:81.5%,n:124

SRA Number	Velvet	GenTreat
SRR5056619	C:93.9% [S:93.4%,D:0.5%],F:3.6%,M:2.5%,n:440	C:87.9% [S:87.7%,D:0.2%],F:0.2%,M:11.9%,n:440
SRR5194991	C:98.7% [S:98.2%,D:0.5%],F:0.7%,M:0.6%,n:440	C:100.0% [S:96.6%,D:3.4%],F:0.0%,M:0.0%,n:440
SRR5386043	C:98.8% [S:98.6%,D:0.2%],F:0.2%,M:1.0%,n:440	C:98.7% [S:98.2%,D:0.5%],F:0.2%,M:1.1%,n:440
SRR5581941	C:99.8% [S:99.3%,D:0.5%],F:0.0%,M:0.2%,n:440	C:99.8% [S:93.4%,D:6.4%],F:0.0%,M:0.2%,n:440
SRR6442873	C:98.5% [S:98.0%,D:0.5%],F:0.5%,M:1.0%,n:440	C:98.9% [S:98.2%,D:0.7%],F:0.2%,M:0.9%,n:440
SRR7945280	C:91.8% [S:91.7%,D:0.1%],F:4.4%,M:3.8%,n:743	C:99.8% [S:99.3%,D:0.5%],F:0.1%,M:0.1%,n:743
SRR10020330	C:97.8% [S:97.7%,D:0.1%],F:0.7%,M:1.5%,n:743	C:98.0% [S:97.7%,D:0.3%],F:0.5%,M:1.5%,n:743
SRR10609577	C:92.6% [S:91.8%,D:0.8%],F:5.4%,M:2.0%,n:743	C:97.3% [S:91.1%,D:6.2%],F:0.8%,M:1.9%,n:743
SRR10971019	C:20.2% [S:19.4%,D:0.8%],F:66.9%,M:12.9%,n:124	C:100.0% [S:94.8%,D:5.2%],F:0.0%,M:0.0%,n:440
SRR12170550	C:97.7% [S:97.5%,D:0.2%],F:1.6%,M:0.7%,n:440	C:90.2% [S:83.6%,D:6.6%],F:0.0%,M:9.8%,n:440
SRR12233579	C:97.2% [S:97.0%,D:0.2%],F:0.9%,M:1.9%,n:440	C:99.1% [S:89.3%,D:9.8%],F:0.2%,M:0.7%,n:440
SRR12270040	C:97.2% [S:97.2%,D:0.0%],F:0.7%,M:2.1%,n:743	C:89.1% [S:88.8%,D:0.3%],F:0.3%,M:10.6%,n:743
SRR12342991	C:98.4% [S:98.2%,D:0.2%],F:0.9%,M:0.7%,n:440	C:91.9% [S:88.0%,D:3.9%],F:0.0%,M:8.1%,n:440
SRR12998338	C:98.6% [S:98.4%,D:0.2%],F:0.2%,M:1.2%,n:440	C:92.2% [S:87.0%,D:5.2%],F:0.2%,M:7.6%,n:440
SRR13355318	C:85.0% [S:84.8%,D:0.2%],F:9.5%,M:5.5%,n:440	C:88.9% [S:78.9%,D:10.0%],F:0.0%,M:11.1%,n:440
SRR15616889	C:98.5% [S:98.5%,D:0.0%],F:1.0%,M:0.5%,n:402	C:96.8% [S:59.7%,D:37.1%],F:2.4%,M:0.8%,n:124
SRR16168733	C:88.7% [S:87.8%,D:0.9%],F:8.2%,M:3.1%,n:743	C:99.4% [S:98.3%,D:1.1%],F:0.0%,M:0.6%,n:743
SRR17276215	C:98.9% [S:98.4%,D:0.5%],F:0.5%,M:0.6%,n:440	C:99.3% [S:94.8%,D:4.5%],F:0.0%,M:0.7%,n:440
DRR296284	C:65.7% [S:65.6%,D:0.1%],F:23.3%,M:11.0%,n:782	C:99.8% [S:93.9%,D:5.9%],F:0.1%,M:0.1%,n:782
SRR572717	C:100.0% [S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450	C:100.0% [S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450
SRR4187667	C:99.1% [S:99.1%,D:0.0%],F:0.9%,M:0.0%,n:450	C:99.8% [S:99.6%,D:0.2%],F:0.2%,M:0.0%,n:450
SRR9435726	C:28.2% [S:27.4%,D:0.8%],F:12.1%,M:59.7%,n:124	C:84.7% [S:84.0%,D:0.7%],F:3.6%,M:11.7%,n:450
SRR9841424	C:98.1% [S:97.8%,D:0.3%],F:0.8%,M:1.1%,n:782	C:93.4% [S:85.2%,D:8.2%],F:0.4%,M:6.2%,n:782
SRR10394909	C:98.1% [S:97.8%,D:0.3%],F:1.2%,M:0.7%,n:782	C:97.3% [S:76.5%,D:20.8%],F:0.1%,M:2.6%,n:782
SRR18890661	C:94.9% [S:94.8%,D:0.1%],F:3.6%,M:1.5%,n:782	C:98.1% [S:97.7%,D:0.4%],F:0.8%,M:1.1%,n:782
SRR20036353	C:3.2% [S:3.2%,D:0.0%],F:44.4%,M:52.4%,n:124	C:2.4% [S:2.4%,D:0.0%],F:21.8%,M:75.8%,n:124
SRR9052983	C:95.2% [S:95.2%,D:0.0%],F:4.8%,M:0.0%,n:124	C:49.2% [S:15.3%,D:33.9%],F:1.6%,M:49.2%,n:124
SRR9641535	C:33.9% [S:33.9%,D:0.0%],F:52.4%,M:13.7%,n:124	C:99.2% [S:66.9%,D:32.3%],F:0.8%,M:0.0%,n:124
SRR5868502	C:49.2% [S:49.2%,D:0.0%],F:41.9%,M:8.9%,n:124	C:99.2% [S:95.2%,D:4.0%],F:0.8%,M:0.0%,n:124
SRR16999781	C:97.6% [S:94.4%,D:3.2%],F:2.4%,M:0.0%,n:124	C:99.2% [S:96.0%,D:3.2%],F:0.8%,M:0.0%,n:124
SRR9592370	C:97.6% [S:94.4%,D:3.2%],F:2.4%,M:0.0%,n:124	C:98.4% [S:91.9%,D:6.5%],F:1.6%,M:0.0%,n:124
SRR11176973	C:92.7% [S:92.7%,D:0.0%],F:6.5%,M:0.8%,n:124	C:98.4% [S:75.0%,D:23.4%],F:1.6%,M:0.0%,n:124

Fonte: Autoria própria.

Tabela 5. Análise de completude gênica dos resultados de montagem das ferramentas SPAdes e GenTreat usando o software BUSCO. Onde C – BUSCOs Completos; S – BUSCOs Completos e de Cópia Única; D – BUSCOs Completos e Duplicados; F – BUSCOs Fragmentados; M – BUSCOs Ausentes; n – Total de grupos BUSCO pesquisados.

SRA Number	SPAdes	GenTreat
DRR161197	C:99.2% [S:98.9%,D:0.3%],F:0.4%,M:0.4%,n:743	C:99.2% [S:98.9%,D:0.3%],F:0.4%,M:0.4%,n:743
DRR161272	C:0.0% [S:0.0%,D:0.0%],F:0.0%,M:100.0%,n:151	C:0.5% [S:0.5%,D:0.0%],F:0.0%,M:99.5%,n:194
DRR174983	-	C:99.1% [S:99.1%,D:0.0%],F:0.4%,M:0.5%,n:743
DRR199345	C:99.1% [S:98.4%,D:0.7%],F:0.0%,M:0.9%,n:440	C:88.8% [S:82.7%,D:6.1%],F:0.0%,M:11.2%,n:440
DRR199814	C:99.1% [S:98.6%,D:0.5%],F:0.0%,M:0.9%,n:440	C:84.8% [S:73.2%,D:11.6%],F:0.0%,M:15.2%,n:440
DRR223366	C:98.7% [S:98.0%,D:0.7%],F:0.2%,M:1.1%,n:440	C:98.0% [S:97.3%,D:0.7%],F:0.2%,M:1.8%,n:440
DRR249160	C:98.3% [S:97.8%,D:0.5%],F:1.1%,M:0.6%,n:743	C:98.2% [S:97.7%,D:0.5%],F:1.1%,M:0.7%,n:743
DRR255737	C:99.3% [S:98.9%,D:0.4%],F:0.3%,M:0.4%,n:743	C:95.4% [S:94.2%,D:1.2%],F:0.3%,M:4.3%,n:743
DRR259154	C:3.2% [S:3.2%,D:0.0%],F:14.5%,M:82.3%,n:124	C:1.6% [S:1.6%,D:0.0%],F:9.7%,M:88.7%,n:124
DRR286070	C:40.4% [S:40.2%,D:0.2%],F:30.0%,M:29.6%,n:440	C:37.5% [S:37.3%,D:0.2%],F:29.1%,M:33.4%,n:440
DRR286167	-	C:100.0% [S:99.2%,D:0.8%],F:0.0%,M:0.0%,n:124
ERR007646	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:95.0% [S:94.5%,D:0.5%],F:0.0%,M:5.0%,n:440
ERR351258	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440
ERR351259	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
ERR351260	C:100.0% [S:98.6%,D:1.4%],F:0.0%,M:0.0%,n:440	C:100.0% [S:97.3%,D:2.7%],F:0.0%,M:0.0%,n:440
ERR5104572	C:99.6% [S:99.1%,D:0.5%],F:0.0%,M:0.4%,n:743	C:99.6% [S:98.4%,D:1.2%],F:0.0%,M:0.4%,n:743
SRR013523	C:96.9% [S:96.8%,D:0.1%],F:1.2%,M:1.9%,n:743	C:96.9% [S:96.8%,D:0.1%],F:1.2%,M:1.9%,n:743
SRR231104	C:99.4% [S:99.1%,D:0.3%],F:0.3%,M:0.3%,n:743	C:99.3% [S:92.2%,D:7.1%],F:0.3%,M:0.4%,n:743
SRR638977	C:88.4% [S:88.3%,D:0.1%],F:8.5%,M:3.1%,n:743	C:86.0% [S:84.5%,D:1.5%],F:8.7%,M:5.3%,n:743
SRR933455	C:98.2% [S:97.7%,D:0.5%],F:0.9%,M:0.9%,n:440	C:97.7% [S:92.0%,D:5.7%],F:0.5%,M:1.8%,n:440
SRR974839	C:93.7% [S:93.0%,D:0.7%],F:4.8%,M:1.5%,n:743	C:92.4% [S:91.9%,D:0.5%],F:4.3%,M:3.3%,n:743
SRR2965748	C:99.1% [S:98.6%,D:0.5%],F:0.0%,M:0.9%,n:440	C:91.1% [S:83.6%,D:7.5%],F:0.2%,M:8.7%,n:440
SRR3312980	C:89.5% [S:89.5%,D:0.0%],F:7.1%,M:3.4%,n:743	C:88.6% [S:88.6%,D:0.0%],F:7.3%,M:4.1%,n:743
SRR3465539	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:96.0% [S:95.5%,D:0.5%],F:0.0%,M:4.0%,n:440
SRR3932423	C:98.5% [S:97.8%,D:0.7%],F:0.5%,M:1.0%,n:743	C:96.8% [S:91.0%,D:5.8%],F:0.5%,M:2.7%,n:743
SRR3947906	C:99.0% [S:98.9%,D:0.1%],F:0.4%,M:0.6%,n:743	C:90.6% [S:80.8%,D:9.8%],F:0.5%,M:8.9%,n:743
SRR3999078	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440
SRR3999096	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
SRR5056410	C:99.6% [S:99.1%,D:0.5%],F:0.0%,M:0.4%,n:440	C:17.7% [S:17.7%,D:0.0%],F:0.8%,M:81.5%,n:124
SRR5056619	C:99.5% [S:99.3%,D:0.2%],F:0.0%,M:0.5%,n:440	C:87.9% [S:87.7%,D:0.2%],F:0.2%,M:11.9%,n:440

SRA Number	SPAdes	GenTreat
SRR5194991	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:96.6%,D:3.4%],F:0.0%,M:0.0%,n:440
SRR5386043	C:99.1% [S:98.6%,D:0.5%],F:0.2%,M:0.7%,n:440	C:98.7% [S:98.2%,D:0.5%],F:0.2%,M:1.1%,n:440
SRR5581941	C:99.8% [S:99.3%,D:0.5%],F:0.0%,M:0.2%,n:440	C:99.8% [S:93.4%,D:6.4%],F:0.0%,M:0.2%,n:440
SRR6442873	C:76.9% [S:70.9%,D:6.0%],F:2.0%,M:21.1%,n:151	C:98.9% [S:98.2%,D:0.7%],F:0.2%,M:0.9%,n:440
SRR7945280	C:99.8% [S:99.3%,D:0.5%],F:0.1%,M:0.1%,n:743	C:99.8% [S:99.3%,D:0.5%],F:0.1%,M:0.1%,n:743
SRR10020330	C:98.1% [S:98.0%,D:0.1%],F:0.5%,M:1.4%,n:743	C:98.0% [S:97.7%,D:0.3%],F:0.5%,M:1.5%,n:743
SRR10609577	C:99.2% [S:98.3%,D:0.9%],F:0.8%,M:0.0%,n:743	C:97.3% [S:91.1%,D:6.2%],F:0.8%,M:1.9%,n:743
SRR10971019	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:94.8%,D:5.2%],F:0.0%,M:0.0%,n:440
SRR12170550	C:98.6% [S:97.7%,D:0.9%],F:0.9%,M:0.5%,n:440	C:90.2% [S:83.6%,D:6.6%],F:0.0%,M:9.8%,n:440
SRR12233579	C:98.7% [S:98.2%,D:0.5%],F:0.2%,M:1.1%,n:440	C:99.1% [S:89.3%,D:9.8%],F:0.2%,M:0.7%,n:440
SRR12270040	C:98.3% [S:97.8%,D:0.5%],F:0.3%,M:1.4%,n:743	C:89.1% [S:88.8%,D:0.3%],F:0.3%,M:10.6%,n:743
SRR12342991	C:99.5% [S:99.3%,D:0.2%],F:0.0%,M:0.5%,n:440	C:91.9% [S:88.0%,D:3.9%],F:0.0%,M:8.1%,n:440
SRR12998338	C:98.9% [S:98.4%,D:0.5%],F:0.2%,M:0.9%,n:440	C:92.2% [S:87.0%,D:5.2%],F:0.2%,M:7.6%,n:440
SRR13355318	C:98.7% [S:98.2%,D:0.5%],F:0.0%,M:1.3%,n:440	C:88.9% [S:78.9%,D:10.0%],F:0.0%,M:11.1%,n:440
SRR15616889	C:87.2% [S:85.5%,D:1.7%],F:10.8%,M:2.0%,n:743	C:96.8% [S:59.7%,D:37.1%],F:2.4%,M:0.8%,n:124
SRR16168733	C:99.8% [S:98.7%,D:1.1%],F:0.0%,M:0.2%,n:743	C:99.4% [S:98.3%,D:1.1%],F:0.0%,M:0.6%,n:743
SRR17276215	C:99.4% [S:98.9%,D:0.5%],F:0.0%,M:0.6%,n:440	C:99.3% [S:94.8%,D:4.5%],F:0.0%,M:0.7%,n:440
DRR296284	C:99.7% [S:99.2%,D:0.5%],F:0.1%,M:0.2%,n:782	C:99.8% [S:93.9%,D:5.9%],F:0.1%,M:0.1%,n:782
SRR572717	C:100.0% [S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450	C:100.0% [S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450
SRR4187667	C:99.8% [S:99.6%,D:0.2%],F:0.2%,M:0.0%,n:450	C:99.8% [S:99.6%,D:0.2%],F:0.2%,M:0.0%,n:450
SRR9435726	C:83.8% [S:83.1%,D:0.7%],F:2.7%,M:13.5%,n:450	C:84.7% [S:84.0%,D:0.7%],F:3.6%,M:11.7%,n:450
SRR9841424	C:99.4% [S:99.0%,D:0.4%],F:0.3%,M:0.3%,n:782	C:93.4% [S:85.2%,D:8.2%],F:0.4%,M:6.2%,n:782
SRR10394909	C:99.7% [S:99.2%,D:0.5%],F:0.1%,M:0.2%,n:782	C:97.3% [S:76.5%,D:20.8%],F:0.1%,M:2.6%,n:782
SRR18890661	C:99.2% [S:98.8%,D:0.4%],F:0.6%,M:0.2%,n:782	C:98.1% [S:97.7%,D:0.4%],F:0.8%,M:1.1%,n:782
SRR20036353	C:0.0% [S:0.0%,D:0.0%],F:0.0%,M:100.0%,n:151	C:2.4% [S:2.4%,D:0.0%],F:21.8%,M:75.8%,n:124
SRR9052983	C:99.8% [S:99.8%,D:0.0%],F:0.0%,M:0.2%,n:402	C:49.2% [S:15.3%,D:33.9%],F:1.6%,M:49.2%,n:124
SRR9641535	-	C:99.2% [S:66.9%,D:32.3%],F:0.8%,M:0.0%,n:124
SRR5868502	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:402	C:99.2% [S:95.2%,D:4.0%],F:0.8%,M:0.0%,n:124
SRR16999781	C:99.8% [S:98.3%,D:1.5%],F:0.1%,M:0.1%,n:743	C:99.2% [S:96.0%,D:3.2%],F:0.8%,M:0.0%,n:124
SRR9592370	C:98.7% [S:98.5%,D:0.2%],F:1.1%,M:0.2%,n:541	C:98.4% [S:91.9%,D:6.5%],F:1.6%,M:0.0%,n:124
SRR11176973	C:99.7% [S:96.0%,D:3.7%],F:0.1%,M:0.2%,n:782	C:98.4% [S:75.0%,D:23.4%],F:1.6%,M:0.0%,n:124

Fonte: Autoria própria.

Tabela 6. Análise de completude gênica dos resultados de montagem das ferramentas Megahit e GenTreat usando o software BUSCO. Onde C – BUSCOs Completos; S – BUSCOs Completos e de Cópia Única; D – BUSCOs Completos e Duplicados; F – BUSCOs Fragmentados; M – BUSCOs Ausentes; n – Total de grupos BUSCO pesquisados.

SRA Number	MEGAHIT	GenTreat
DRR161197	C:99.2% [S:98.9%,D:0.3%],F:0.4%,M:0.4%,n:743	C:99.2% [S:98.9%,D:0.3%],F:0.4%,M:0.4%,n:743
DRR161272	C:0.5% [S:0.5%,D:0.0%],F:0.0%,M:99.5%,n:194	C:0.5% [S:0.5%,D:0.0%],F:0.0%,M:99.5%,n:194
DRR174983	-	C:99.1% [S:99.1%,D:0.0%],F:0.4%,M:0.5%,n:743
DRR199345	C:99.1% [S:98.4%,D:0.7%],F:0.0%,M:0.9%,n:440	C:88.8% [S:82.7%,D:6.1%],F:0.0%,M:11.2%,n:440
DRR199814	C:99.1% [S:98.6%,D:0.5%],F:0.0%,M:0.9%,n:440	C:84.8% [S:73.2%,D:11.6%],F:0.0%,M:15.2%,n:440
DRR223366	C:98.9% [S:98.2%,D:0.7%],F:0.2%,M:0.9%,n:440	C:98.0% [S:97.3%,D:0.7%],F:0.2%,M:1.8%,n:440
DRR249160	C:98.2% [S:97.8%,D:0.4%],F:1.2%,M:0.6%,n:743	C:98.2% [S:97.7%,D:0.5%],F:1.1%,M:0.7%,n:743
DRR255737	C:99.3% [S:98.9%,D:0.4%],F:0.3%,M:0.4%,n:743	C:95.4% [S:94.2%,D:1.2%],F:0.3%,M:4.3%,n:743
DRR259154	C:1.7% [S:1.7%,D:0.0%],F:3.4%,M:94.9%,n:174	C:1.6% [S:1.6%,D:0.0%],F:9.7%,M:88.7%,n:124
DRR286070	C:25.8% [S:25.8%,D:0.0%],F:12.6%,M:61.6%,n:151	C:37.5% [S:37.3%,D:0.2%],F:29.1%,M:33.4%,n:440
DRR286167	C:44.9% [S:44.3%,D:0.6%],F:9.2%,M:45.9%,n:174	C:100.0% [S:99.2%,D:0.8%],F:0.0%,M:0.0%,n:124
ERR007646	C:99.1% [S:98.9%,D:0.2%],F:0.5%,M:0.4%,n:440	C:95.0% [S:94.5%,D:0.5%],F:0.0%,M:5.0%,n:440
ERR351258	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440
ERR351259	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0% [S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
ERR351260	C:100.0% [S:99.1%,D:0.9%],F:0.0%,M:0.0%,n:440	C:100.0% [S:97.3%,D:2.7%],F:0.0%,M:0.0%,n:440
ERR5104572	C:99.6% [S:99.2%,D:0.4%],F:0.0%,M:0.4%,n:743	C:99.6% [S:98.4%,D:1.2%],F:0.0%,M:0.4%,n:743
SRR013523	C:97.1% [S:97.0%,D:0.1%],F:0.9%,M:2.0%,n:743	C:96.9% [S:96.8%,D:0.1%],F:1.2%,M:1.9%,n:743
SRR231104	-	C:99.3% [S:92.2%,D:7.1%],F:0.3%,M:0.4%,n:743
SRR638977	C:81.4% [S:81.3%,D:0.1%],F:13.3%,M:5.3%,n:743	C:86.0% [S:84.5%,D:1.5%],F:8.7%,M:5.3%,n:743
SRR933455	C:100.0% [S:98.9%,D:1.1%],F:0.0%,M:0.0%,n:440	C:97.7% [S:92.0%,D:5.7%],F:0.5%,M:1.8%,n:440
SRR974839	C:94.5% [S:94.2%,D:0.3%],F:4.2%,M:1.3%,n:743	C:92.4% [S:91.9%,D:0.5%],F:4.3%,M:3.3%,n:743
SRR2965748	C:99.1% [S:98.6%,D:0.5%],F:0.0%,M:0.9%,n:440	C:91.1% [S:83.6%,D:7.5%],F:0.2%,M:8.7%,n:440
SRR3312980	C:82.5% [S:82.5%,D:0.0%],F:13.7%,M:3.8%,n:743	C:88.6% [S:88.6%,D:0.0%],F:7.3%,M:4.1%,n:743
SRR3465539	C:71.8% [S:69.5%,D:2.3%],F:6.3%,M:21.9%,n:174	C:96.0% [S:95.5%,D:0.5%],F:0.0%,M:4.0%,n:440
SRR3932423	C:98.4% [S:97.7%,D:0.7%],F:0.7%,M:0.9%,n:743	C:96.8% [S:91.0%,D:5.8%],F:0.5%,M:2.7%,n:743
SRR3947906	C:98.4% [S:98.3%,D:0.1%],F:1.1%,M:0.5%,n:743	C:90.6% [S:80.8%,D:9.8%],F:0.5%,M:8.9%,n:743
SRR3999078	C:70.7% [S:68.4%,D:2.3%],F:6.3%,M:23.0%,n:174	C:100.0% [S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440

SRA Number	MEGAHIT	GenTreat
SRR3999096	C:71.3%[S:68.4%,D:2.9%],F:6.3%,M:22.4%,n:174	C:100.0%[S:99.3%,D:0.7%],F:0.0%,M:0.0%,n:440
SRR5056410	-	C:17.7%[S:17.7%,D:0.0%],F:0.8%,M:81.5%,n:124
SRR5056619	C:99.4%[S:98.9%,D:0.5%],F:0.0%,M:0.6%,n:440	C:87.9%[S:87.7%,D:0.2%],F:0.2%,M:11.9%,n:440
SRR5194991	C:100.0%[S:99.5%,D:0.5%],F:0.0%,M:0.0%,n:440	C:100.0%[S:96.6%,D:3.4%],F:0.0%,M:0.0%,n:440
SRR5386043	C:99.1%[S:98.6%,D:0.5%],F:0.2%,M:0.7%,n:440	C:98.7%[S:98.2%,D:0.5%],F:0.2%,M:1.1%,n:440
SRR5581941	C:99.6%[S:99.1%,D:0.5%],F:0.0%,M:0.4%,n:440	C:99.8%[S:93.4%,D:6.4%],F:0.0%,M:0.2%,n:440
SRR6442873	C:98.9%[S:98.2%,D:0.7%],F:0.2%,M:0.9%,n:440	C:98.9%[S:98.2%,D:0.7%],F:0.2%,M:0.9%,n:440
SRR7945280	C:99.9%[S:99.2%,D:0.7%],F:0.1%,M:0.0%,n:743	C:99.8%[S:99.3%,D:0.5%],F:0.1%,M:0.1%,n:743
SRR10020330	C:97.9%[S:97.8%,D:0.1%],F:0.5%,M:1.6%,n:743	C:98.0%[S:97.7%,D:0.3%],F:0.5%,M:1.5%,n:743
SRR10609577	C:99.3%[S:98.4%,D:0.9%],F:0.5%,M:0.2%,n:743	C:97.3%[S:91.1%,D:6.2%],F:0.8%,M:1.9%,n:743
SRR10971019	C:82.1%[S:78.1%,D:4.0%],F:2.0%,M:15.9%,n:151	C:100.0%[S:94.8%,D:5.2%],F:0.0%,M:0.0%,n:440
SRR12170550	C:99.4%[S:98.9%,D:0.5%],F:0.0%,M:0.6%,n:440	C:90.2%[S:83.6%,D:6.6%],F:0.0%,M:9.8%,n:440
SRR12233579	C:98.9%[S:98.4%,D:0.5%],F:0.2%,M:0.9%,n:440	C:99.1%[S:89.3%,D:9.8%],F:0.2%,M:0.7%,n:440
SRR12270040	C:98.2%[S:97.7%,D:0.5%],F:0.3%,M:1.5%,n:743	C:89.1%[S:88.8%,D:0.3%],F:0.3%,M:10.6%,n:743
SRR12342991	C:99.6%[S:95.7%,D:3.9%],F:0.0%,M:0.4%,n:440	C:91.9%[S:88.0%,D:3.9%],F:0.0%,M:8.1%,n:440
SRR12998338	C:98.9%[S:98.4%,D:0.5%],F:0.2%,M:0.9%,n:440	C:92.2%[S:87.0%,D:5.2%],F:0.2%,M:7.6%,n:440
SRR13355318	C:98.5%[S:98.0%,D:0.5%],F:0.2%,M:1.3%,n:440	C:88.9%[S:78.9%,D:10.0%],F:0.0%,M:11.1%,n:440
SRR15616889	C:85.4%[S:83.7%,D:1.7%],F:12.4%,M:2.2%,n:743	C:96.8%[S:59.7%,D:37.1%],F:2.4%,M:0.8%,n:124
SRR16168733	C:99.6%[S:98.5%,D:1.1%],F:0.1%,M:0.3%,n:743	C:99.4%[S:98.3%,D:1.1%],F:0.0%,M:0.6%,n:743
SRR17276215	-	C:99.3%[S:94.8%,D:4.5%],F:0.0%,M:0.7%,n:440
DRR296284	-	C:99.8%[S:93.9%,D:5.9%],F:0.1%,M:0.1%,n:782
SRR572717	C:100.0%[S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450	C:100.0%[S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:450
SRR4187667	C:99.8%[S:99.8%,D:0.0%],F:0.2%,M:0.0%,n:450	C:99.8%[S:99.6%,D:0.2%],F:0.2%,M:0.0%,n:450
SRR9435726	C:82.0%[S:81.3%,D:0.7%],F:5.6%,M:12.4%,n:450	C:84.7%[S:84.0%,D:0.7%],F:3.6%,M:11.7%,n:450
SRR9841424	C:99.5%[S:99.1%,D:0.4%],F:0.3%,M:0.2%,n:782	C:93.4%[S:85.2%,D:8.2%],F:0.4%,M:6.2%,n:782
SRR10394909	C:99.6%[S:99.2%,D:0.4%],F:0.3%,M:0.1%,n:782	C:97.3%[S:76.5%,D:20.8%],F:0.1%,M:2.6%,n:782
SRR18890661	C:90.9%[S:90.5%,D:0.4%],F:6.5%,M:2.6%,n:782	C:98.1%[S:97.7%,D:0.4%],F:0.8%,M:1.1%,n:782
SRR20036353	-	C:2.4%[S:2.4%,D:0.0%],F:21.8%,M:75.8%,n:124
SRR9052983	C:99.8%[S:99.8%,D:0.0%],F:0.0%,M:0.2%,n:402	C:49.2%[S:15.3%,D:33.9%],F:1.6%,M:49.2%,n:124
SRR9641535	C:99.5%[S:99.5%,D:0.0%],F:0.0%,M:0.5%,n:402	C:99.2%[S:66.9%,D:32.3%],F:0.8%,M:0.0%,n:124
SRR5868502	C:100.0%[S:100.0%,D:0.0%],F:0.0%,M:0.0%,n:402	C:99.2%[S:95.2%,D:4.0%],F:0.8%,M:0.0%,n:124
SRR16999781	C:99.8%[S:98.3%,D:1.5%],F:0.1%,M:0.1%,n:743	C:99.2%[S:96.0%,D:3.2%],F:0.8%,M:0.0%,n:124
SRR9592370	C:100.0%[S:99.8%,D:0.2%],F:0.0%,M:0.0%,n:541	C:98.4%[S:91.9%,D:6.5%],F:1.6%,M:0.0%,n:124
SRR11176973	C:99.8%[S:98.5%,D:1.3%],F:0.1%,M:0.1%,n:782	C:98.4%[S:75.0%,D:23.4%],F:1.6%,M:0.0%,n:124

Fonte: Autoria própria.