On Variants of the Genome Rearrangement Distance Problem

Alexsandro Oliveira Alexandrino[1], **Ulisses Dias**[2], **Zanoni Dias**[1]

[1]Institute of Computing – University of Campinas (Unicamp)
Campinas – SP – Brazil

[2]School of Technology – University of Campinas (Unicamp)
Limeira – SP – Brazil

{alexsandro,zanoni}@ic.unicamp.br, ulisses@ft.unicamp.br

Abstract. *In comparative genomics, evolutionary distance can be estimated by determining a minimum sequence of genome rearrangements required to transform one genome into another. The length of this sequence is called the rearrangement distance. Traditionally, most studies assumed that compared genomes shared the same set of genes, with only their relative gene order being used for comparison. Recent studies indicate that incorporating intergenic region sizes leads to more accurate distance estimates when analyzing real genomes. Furthermore, they started to consider genomes having a distinct set of genes. In this work, we investigate genome rearrangement problems by incorporating additional information and complexity into the models, in order to achieve more realistic practical results that can benefit several scientific fields.*

Resumo. *Na genômica comparativa, a distância evolucionária pode ser estimada determinando uma sequência mínima de rearranjos de genomas necessários para transformar um genoma em outro. O tamanho dessa sequência é chamado de distância de rearranjos. Tradicionalmente, a maioria dos estudos assumiu que os genomas comparados compartilhavam o mesmo conjunto de genes, utilizando apenas a ordem relativa dos genes para a comparação. Estudos recentes indicam que incorporar o tamanho das regiões intergênicas leva a estimativas de distância mais precisas ao analisar genomas reais. Além disso, esses estudos começaram a considerar que os genomas possuem um conjunto distinto de genes. Neste trabalho, investigamos problemas de rearranjos de genomas incorporando informações adicionais e complexidade aos modelos, a fim de alcançar resultados práticos mais realistas que possam beneficiar diversos campos científicos.*

1. Introduction

Comparative genomics is a branch of biological research focused on analyzing evolutionary relationships by comparing genomic data across different organisms. In this context, different gene structures and metrics can be used. A crucial aspect is the study of genome rearrangement events, which are genetic mutations that affect stretches of the genome, changing the quantity, order, and orientation of genetic material. The investigation of these events led to the development of genome rearrangement problem studies. Genome rearrangement problems aim to determine the shortest sequence of rearrangement events required to transform one genome into another. The length of this sequence is known as the *rearrangement distance*.

Genome rearrangement analyses contribute to biology by offering a robust metric for estimating evolutionary distances and enhancing the accuracy of phylogenetic tree construction [Hannenhalli and Pevzner 1995, Bochkareva et al. 2018]. In medicine, these results can be used to detect, prevent, and treat genetic diseases and disorders [Lupski 1998, Chen et al. 2010, Schuy et al. 2022], with these applications expanding each year driven by significant advancements in the commercialization of genetic testing and sequencing for the general population [Shendure et al. 2017]. Although genome rearrangement problems are biologically motivated, they have also been shown to be computationally challenging, as many of them are NP-hard [Caprara 1999, Bulteau et al. 2012, Oliveira et al. 2019]. Moreover, they are directly related to other problems known in the literature, such as the String Partition and Cycle Decomposition problems [Caprara 1999, Oliveira et al. 2021a, Oliveira et al. 2024, Siqueira et al. 2024, Siqueira et al. 2023a, Siqueira et al. 2023b].

In rearrangement problems, a genome is typically represented as a sequence of genes, and depending on the available genomic information, different mathematical models may be used. Assuming that there are no repeated genes and that the genomes contain the same set of genes, a genome can be modeled as a permutation of integers, where each element represents a gene. When the orientation of the genes is known, this information is represented by the sign (positive or negative) of the elements in the permutation. In this case, the permutation is referred to as *signed*. When the orientation of the genes is unknown, unsigned permutations are used to represent the genomes. By using permutations, the problem of transforming one genome into another becomes equivalent to the problem of sorting a permutation by rearrangements [Fertin et al. 2009].

A *rearrangement model* specifies the allowable operations used to compute the rearrangement distance. Two of the most extensively studied rearrangement events are reversals, which flip a genome segment, and transpositions, which swap the positions of two adjacent segments.

Early studies on genome rearrangement problems focused on individual types of rearrangements, leading to the development of Sorting Permutations by Reversals [Caprara 1999] and Sorting Permutations by Transpositions [Bafna and Pevzner 1995] problems. Later, these rearrangements were incorporated into a single model. The problem of Sorting Signed Permutations by Reversals has a polynomial exact algorithm [Hannenhalli and Pevzner 1999]. The problems of Sorting Unsigned Permutations by Reversals or Transpositions are NP-hard [Bulteau et al. 2012, Caprara 1999]. The problem that allows both reversals and transpositions is NP-hard for both signed and unsigned permutations [Oliveira et al. 2019].

Although the representation of gene order using permutations has limitations, making it difficult to apply it directly to many genomes, it tends to provide strong bounds on the rearrangement problems. These permutation-based bounds are critical to obtain good results in more realistic models that allow genes with multiple copies [Chen et al. 2005, Siqueira et al. 2021]. Thus, achieving better results in problems that represent genomes through permutations enhances the effectiveness of more complex models.

When genomes contain repeated genes or have distinct gene sets, they are mathe-

matically represented as strings. We say that genomes that have distinct sets of genes are *unbalanced genomes*; otherwise, we say that they are *balanced genomes*. Similarly to the representation using permutations, gene orientation is indicated by the sign (positive or negative) of each element in the string.

The aforementioned rearrangements are all conservative operations, as they do not alter the amount of genetic material in the genome. There are also non-conservative operations, such as insertions, which add a segment to a position in the genome, and deletions, which remove a segment from the genome [Willing et al. 2013]. Insertions and deletions are referred to as *indels*. In models that have indels, we deal with unbalanced genomes, and, consequently, strings are used to represent the genomes.

In addition to the relative order in which genes appear in the genome, recent studies have incorporated information about the size of intergenic regions (DNA sequences between each pair of genes) in genome representations, based on evidence that intergenic regions help infer more accurate evolutionary scenarios [Biller et al. 2016a, Biller et al. 2016b]. Existing intergenic models in the literature include both conservative operations and indels, however, these indels can only add or remove intergenic regions, thus restricting problems to compare balanced genomes.

In 2016, Bulteau and coauthors [Bulteau et al. 2016] presented one of the first works considering an intergenic representation. The rearrangement model was composed of the Double Cut and Join operation (DCJ, which cuts two points in the genome and joins the resulting segments following a specific criteria) and indels. In this variant of the problem, gene orientation is known, the insertions and deletions act only in the intergenic regions, and there exists an exact polynomial-time algorithm for solving the problem. However, when the rearrangement model is composed only of the DCJ event, Fertin and coauthors [Fertin et al. 2017] showed that the problem belongs to the NP-hard class and presented a $\frac{4}{3}$-factor approximation algorithm. Subsequent studies [Oliveira et al. 2021a, Brito et al. 2020, Oliveira et al. 2020, Brito et al. 2021] have shown that the rearrangement problems considering intergenic regions with reversals and/or transpositions are NP-hard for signed and unsigned permutations.

In the thesis, we studied the Rearrangement Distance problem which considers only conservative rearrangements, based on the gene sequence and their orientations (when known), and the Intergenic Rearrangement Distance problem that includes indels and incorporates the sizes of intergenic regions in the genome representation, in addition to using the gene sequence and their orientations (when known). We introduced new structures and concepts for problems involving reversals, transpositions, and the combination of reversals and transpositions, which are used in complexity proofs and approximation algorithms. Furthermore, we conducted experiments on both synthetic and real genomes, demonstrating the practical applicability of our algorithms.

The following sections present the main contributions of the thesis. Section 2 presents the results related to our new 1.375-approximation algorithm for Sorting Permutations by Transpositions, and also the hardness proofs for Sorting Permutations by Transpositions and Other Rearrangements problems. Section 3 considers rearrangement distance problems in unbalanced genomes, and Section 4 considers intergenic rearrangement distance problems in unbalanced genomes. Section 5 summarizes all the contribu-

tions of the thesis, and Section 6 presents our final considerations.

2. Sorting Permutations by Transpositions and Other Rearrangements

Two of the most well-known algorithms in genome rearrangement are the exact polynomial-time algorithm for Sorting Signed Permutations by Reversals [Hannenhalli and Pevzner 1999] and the 1.375-approximation algorithm for Sorting Permutations by Transpositions [Elias and Hartman 2006]. Recently, Silva and coauthors [Silva et al. 2022] showed that the approximation factor of the Elias and Hartman algorithm [Elias and Hartman 2006] exceeds 1.375 in some cases. They also showed how to solve this problem and presented an algorithm with a time complexity of $O(n^6)$.

Our **first major contribution** was the development of a new version of the algorithm proposed by Elias and Hartman [Elias and Hartman 2006] that guarantees the approximation factor of 1.375 in all cases and has a time complexity of $O(n^5)$. This improvement was achieved by establishing new theoretical results on cycle graph structures, enabling a more efficient 1.375-approximation algorithm. In addition to the improvement in time complexity, our tests on small permutations showed that our algorithm also enhances the quality of the solutions found, as our algorithm found optimal solutions in more cases than the other two algorithms, and it also achieved a lower average approximation factor compared to them.

Additional genome rearrangement operations explored in the literature include inverse transpositions and revrevs. Like standard transpositions, these operations affect two adjacent genome segments. An inverse transposition swaps their positions while also reversing one of the segments. In contrast, a revrev flips both segments without swapping their relative positions. The complexity of Sorting Permutations by Rearrangements with models that include inverse transpositions and revrevs was unknown, despite the existence of approximation algorithms for these problems [Fertin et al. 2009].

Our **second major contribution** was demonstrating that the problems of Sorting (Signed or Unsigned) Permutations by Weighted Rearrangements are NP-hard for 12 rearrangement models that include transpositions along with combinations of reversals, inverse transpositions, and revrevs, considering that the cost of a reversal is w_ρ, the costs of a transposition, an inverse transposition, or a revrev are the same and equal to w_τ, and under the constraint that $w_\tau/w_\rho \leq 1.5$. Note that when $w_\tau/w_\rho = 1$, the weighted version is equivalent to the traditional approach.

3. Rearrangement Distance with Indels

Starting in 1999, works considering genomes with distinct gene sets were introduced [Sankoff 1999]. In 2000, El-Mabrouk [El-Mabrouk 2000] studied the problem of Reversal and Indels Distance in Signed Strings, presenting heuristics based on the exact algorithm for the version of the problem with permutations [Hannenhalli and Pevzner 1999]. Later, Willing and coauthors [Willing et al. 2013] created an exact polynomial algorithm for Reversal and Indels Distance in Signed Strings, considering specific classes of breakpoint graphs. Only recently, the authors [Willing et al. 2021] extended the previous algorithm and developed an exact polynomial algorithm that works for any instance of the Reversal and Indels Distance problem in Signed Strings.

In the thesis, we studied Rearrangement Distance problems with indels, considering models that combine the operations of reversal, transposition, and another operation called block interchange, which swaps the relative position of two arbitrary genome segments. For these problems, we can list the following major contributions.

- Our **third major contribution** was proving the NP-hardness of multiple rearrangement problems involving indels, including: Reversal and Indels Distance in Unsigned Strings; Transposition and Indels Distance in Unsigned Strings; and Reversal, Transposition, and Indels Distance on Signed or Unsigned Strings.
- Our **fourth major contribution** was adapting the breakpoint concept for unbalanced genomes and introducing the Labeled Cycle Graph, an extension of the classic breakpoint graph tailored for these scenarios.
- Our **fifth major contribution** was the development of the algorithms summarized in Table 1, accompanied by extensive experiments using synthetic genomes to evaluate their performance.

Table 1. Summary of the algorithms presented for the Rearrangement Distance problems in unbalanced genomes. These algorithms use either breakpoints or the labeled cycle graph.

Model	Breakpoints	Cycle Graph
Reversals and Indels (unsigned)	2-approximation	-
Transpositions and Indels (unsigned)	3-approximation	2-approximation
Transpositions, Reversals, and Indels (unsigned)	3-approximation	-
Transpositions, Reversals, and Indels (signed)	-	2-approximation
Block Interchanges and Indels (unsigned)	-	2-approximation
Block Interchanges, Reversals, and Indels (signed)	-	2-approximation

4. Rearrangement Distance with Intergenic Regions

The study of genome rearrangements incorporating intergenic regions is relatively recent. These studies assume that genomes do not contain repeated genes and that insertions and deletions only affect intergenic regions. Thus, the genomes have the same set of genes and can be modeled using permutations and a list of numerical values representing the sizes of the intergenic regions.

Oliveira and coauthors [Oliveira et al. 2021a] presented a 2-approximation for intergenic reversals in signed permutations, along with a proof of NP-hardness for this problem. They also presented a 2-approximation for a version of the problem with intergenic reversals and indels of intergenic regions in signed permutations, the complexity of which remains open. For the model with intergenic reversals and transpositions in signed permutations, Oliveira and coauthors [Oliveira et al. 2021b] presented a 3-approximation and a proof of NP-hardness.

Considering unsigned permutations, Brito and coauthors [Brito et al. 2020] demonstrated that the Intergenic Rearrangement Distance problem is NP-hard for the following models: intergenic reversals; intergenic reversals and indels of intergenic regions; intergenic reversals and transpositions; and intergenic reversals, transpositions, and indels

of intergenic regions. The authors presented a 4-approximation for the models with intergenic reversals including or not indels of intergenic regions, and a 4.5-approximation for the other two models. For intergenic transpositions in unsigned permutations, Oliveira and coauthors [Oliveira et al. 2020] developed a 3.5-approximation and proved that the problem is NP-hard.

In the thesis, we studied Intergenic Rearrangement Distance problems on unbalanced genomes, considering models that combine the operations of reversal, transposition, and indels. For these problems, we can list the following major contributions.

- Our **sixth major contribution** was proving the NP-hardness of multiple intergenic rearrangement distance problems, including: Intergenic Reversal and Indels Distance in Unsigned Strings; Intergenic Transposition and Indels Distance in Unsigned Strings; and Intergenic Reversal, Transposition, and Indels Distance on Signed or Unsigned Strings.
- Our **seventh major contribution** was extending the breakpoint concept to accommodate intergenic regions and unbalanced genomes. Furthermore, we introduced the Weighted Labeled Cycle Graph structure, which is a graph that can represent an instance of intergenic problems considering unbalanced genomes.
- Our **eighth major contribution** was the development of the algorithms summarized in Table 2, along with extensive experiments using synthetic genomes to evaluate their performance.
- Our **ninth major contribution** involved experiments with real genomic data, utilizing cyanobacterial genomes from the Cyanorak 2.1 database [Garczarek et al. 2020]. In this experiment, we constructed phylogenetic trees using our algorithm for Intergenic Reversal and Indels Distance, as well as the exact polynomial algorithm for Sorting Signed Permutations by Reversals, developed by Hannenhalli and Pevzner [Hannenhalli and Pevzner 1999]. When compared to the phylogenetic tree created by the authors of the Cyanorak 2.1 database, the phylogenetic tree generated by our algorithm exhibited a higher level of topological congruence than the tree produced using the algorithm from Hannenhalli and Pevzner.

Table 2. Summary of the algorithms presented for the Intergenic Rearrangement Distance problems in unbalanced genomes. These algorithms use either intergenic breakpoints or the weighted labeled cycle graph.

Model	Breakpoints	Cycle Graph
Reversals and Indels (signed)	-	2.5-approximation
Reversals and Indels (unsigned)	4-approximation	-
Transpositions and Indels (unsigned)	4.5-approximation	4-approximation
Transpositions, Reversals and Indels (signed)	-	4-approximation
Transpositions, Reversals and Indels (unsigned)	6-approximation	-

5. Publications

The findings of this thesis led to the publication of seven articles in international journals and five full papers presented at conferences. Beyond the scope of the thesis, our

contributions to comparative genomics research during this period resulted in an additional nine journal articles and twelve conference papers. The research developed in the thesis allowed us to collaborate with leading computational biology researchers from the University of Nantes, including Géraldine Jean and Guillaume Fertin.

The algorithms developed and the databases created are available in the public repository of the Computational Biology research group of the Institute of Computing of the University of Campinas (Unicamp)[1].

Table 3 provides an overview of the publications resulting from this doctoral research.

Table 3. Summary of publications showing the number of papers published in each journal and conference. The "Thesis" column refers to publications directly related to this thesis, while the "Extra" column refers to additional contributions to the field of genome rearrangements made during the doctoral period.

Journal	**Qualis**	**Thesis**	**Extra**
ACM Computing Surveys	A1	1	-
IEEE-ACM Transactions on Computational Biology and Bioinformatics	A1	1	1
Journal of Computational Biology	A2	3	-
Algorithms for Molecular Biology	A2	-	2
Journal of Combinatorial Optimization	A2	-	3
Journal of Universal Computer Science	A3	1	-
Journal of Bioinformatics and Computational Biology	B1	1	3

Conference	**Qualis**	**Thesis**	**Extra**
Annual Satellite Conference of RECOMB on Comparative Genomics (RECOMB-CG)	A3	1	3
Simpósio Brasileiro de Pesquisa Operacional (SBPO)	A4	-	1
International Symposium on Bioinformatics Research and Applications (ISBRA)	B1	1	1
Brazilian Symposium on Bioinformatics (BSB)	B1	2	3
International Conference on Algorithms for Computational Biology (AlCoB)	B2	1	1
Latin and American Algorithms, Graphs and Optimization Symposium (LAGOS)	B4	-	2
International Conference on Bioinformatics and Computational Biology (BICOB)	-	-	1

[1]Public repository available at: https://github.com/compbiogroup

6. Final Considerations

The work developed in the thesis added more information and realism to the genome models created focusing on genome rearrangement problems. The thesis can be divided into three parts: (i) Advancing the understanding of computational complexity for Sorting Permutations by Transpositions when combined with other known rearrangements, and developing a more efficient algorithm for this problem; (ii) Investigating rearrangement distance problems in genomes with distinct gene sets, requiring the integration of non-conservative operations into rearrangement models; (iii) Exploring rearrangement problems that incorporate intergenic region information while integrating non-conservative operations. This research provided theoretical and practical contributions through complexity analyses, algorithm development, and experimental validation using real and synthetic genomic data. The results presented in the thesis have the potential to benefit more generic models, enabling their application to a wider range of organisms.

Future research can extend these models by incorporating additional features, such as the ability to map genomes with repeated genes.

References

Bafna, V. and Pevzner, P. A. (1995). Sorting Permutations by Transpositions. In *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA' 1995)*, pages 614–623, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.

Biller, P., Guéguen, L., Knibbe, C., and Tannier, E. (2016a). Breaking Good: Accounting for Fragility of Genomic Regions in Rearrangement Distance Estimation. *Genome Biology and Evolution*, 8(5):1427–1439.

Biller, P., Knibbe, C., Beslon, G., and Tannier, E. (2016b). Comparative Genomics on Artificial Life. In *Pursuit of the Universal*, pages 35–44. Springer International Publishing.

Bochkareva, O. O., Moroz, E. V., Davydov, I. I., and Gelfand, M. S. (2018). Genome Rearrangements and Selection in Multi-chromosome Bacteria Burkholderia spp. *BMC genomics*, 19(1):1–17.

Brito, K. L., Jean, G., Fertin, G., Oliveira, A. R., Dias, U., and Dias, Z. (2020). Sorting by Genome Rearrangements on both Gene Order and Intergenic Sizes. *Journal of Computational Biology*, 27(2):156–174.

Brito, K. L., Oliveira, A. R., Alexandrino, A. O., Dias, U., and Dias, Z. (2021). An Improved Approximation Algorithm for the Reversal and Transposition Distance Considering Gene Order and Intergenic Sizes. *Algorithms for Molecular Biology*, 16(1):1–21.

Bulteau, L., Fertin, G., and Rusu, I. (2012). Sorting by Transpositions is Difficult. *SIAM Journal on Discrete Mathematics*, 26(3):1148–1180.

Bulteau, L., Fertin, G., and Tannier, E. (2016). Genome Rearrangements with Indels in Intergenes Restrict the Scenario Space. *BMC Bioinformatics*, 17(14):426.

Caprara, A. (1999). Sorting Permutations by Reversals and Eulerian Cycle Decompositions. *SIAM Journal on Discrete Mathematics*, 12(1):91–110.

Chen, J.-M., Cooper, D. N., Férec, C., Kehrer-Sawatzki, H., and Patrinos, G. P. (2010). Genomic Rearrangements in Inherited Disease and Cancer. *Seminars in Cancer Biology*, 20(4):222–233.

Chen, X., Zheng, J., Fu, Z., Nan, P., Zhong, Y., Lonardi, S., and Jiang, T. (2005). Assignment of Orthologous Genes via Genome Rearrangement. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(4):302–315.

El-Mabrouk, N. (2000). Genome Rearrangement by Reversals and Insertions/Deletions of Contiguous Segments. In *Annual Symposium on Combinatorial Pattern Matching (CPM'2000)*, pages 222–234. Springer.

Elias, I. and Hartman, T. (2006). A 1.375-Approximation Algorithm for Sorting by Transpositions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4):369–379.

Fertin, G., Jean, G., and Tannier, E. (2017). Algorithms for Computing the Double Cut and Join Distance on both Gene Order and Intergenic Sizes. *Algorithms for Molecular Biology*, 12(1):16.

Fertin, G., Labarre, A., Rusu, I., Tannier, É., and Vialette, S. (2009). *Combinatorics of Genome Rearrangements*. Computational Molecular Biology. The MIT Press, London, England.

Garczarek, L., Guyet, U., Doré, H., Farrant, G. K., Hoebeke, M., Brillet-Guéguen, L., Bisch, A., Ferrieux, M., Siltanen, J., Corre, E., Corguillé, G. L., Ratin, M., Pitt, F. D., Ostrowski, M., Conan, M., Siegel, A., Labadie, K., Aury, J.-M., Wincker, P., Scanlan, D. J., and Partensky, F. (2020). Cyanorak v2.1: A Scalable Information System Dedicated to the Visualization and Expert Curation of Marine and Brackish Picocyanobacteria Genomes. *Nucleic Acids Research*, 49(D1):D667–D676.

Hannenhalli, S. and Pevzner, P. A. (1995). Transforming Men into Mice (Polynomial Algorithm for Genomic Distance Problem). In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS'1995)*, pages 581–592, Washington, DC, USA. IEEE Computer Society Press.

Hannenhalli, S. and Pevzner, P. A. (1999). Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals. *Journal of the ACM*, 46(1):1–27.

Lupski, J. R. (1998). Genomic Disorders: Structural Features of the Genome can Lead to DNA Rearrangements and Human Disease Traits. *Trends in Genetics*, 14(10):417–422.

Oliveira, A. R., Brito, K. L., Alexandrino, A. O., Siqueira, G., Dias, U., and Dias, Z. (2024). Rearrangement distance problems: An updated survey. *ACM Computing Surveys*, 56(8).

Oliveira, A. R., Brito, K. L., Dias, U., and Dias, Z. (2019). On the Complexity of Sorting by Reversals and Transpositions Problems. *Journal of Computational Biology*, 26:1223–1229.

Oliveira, A. R., Jean, G., Fertin, G., Brito, K. L., Bulteau, L., Dias, U., and Dias, Z. (2021a). Sorting Signed Permutations by Intergenic Reversals. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6):2870–2876.

Oliveira, A. R., Jean, G., Fertin, G., Brito, K. L., Dias, U., and Dias, Z. (2020). A 3.5-Approximation Algorithm for Sorting by Intergenic Transpositions. In *Proceedings of the 7th International Conference on Algorithms for Computational Biology (AlCoB'2020)*, pages 16–28. Springer International Publishing.

Oliveira, A. R., Jean, G., Fertin, G., Brito, K. L., Dias, U., and Dias, Z. (2021b). Sorting Permutations by Intergenic Operations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6):2080–2093.

Sankoff, D. (1999). Genome Rearrangement with Gene Families. *Bioinformatics*, 15(11):909–917.

Schuy, J., Grochowski, C. M., Carvalho, C. M., and Lindstrand, A. (2022). Complex Genomic Rearrangements: an Underestimated Cause of Rare Diseases. *Trends in Genetics*, 38(11):1134–1146.

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., and Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, 550(7676):345–353.

Silva, L. A. G., Kowada, L. A. B., Rocco, N. R., and Walter, M. E. M. T. (2022). A new 1.375-approximation algorithm for sorting by transpositions. *Algorithms for Molecular Biology*, 17(1):1–17.

Siqueira, G., Alexandrino, A. O., and Dias, Z. (2023a). Signed Rearrangement Distances Considering Repeated Genes, Intergenic Regions, and Indels. *Journal of Combinatorial Optimization*, 46(2):16.

Siqueira, G., Alexandrino, A. O., Oliveira, A. R., and Dias, Z. (2021). Approximation Algorithm for Rearrangement Distances Considering Repeated Genes and Intergenic Regions. *Algorithms for Molecular Biology*, 16(1):1–23.

Siqueira, G., Oliveira, A. R., Alexandrino, A. O., Jean, G., Fertin, G., and Dias, Z. (2024). Assignment of orthologous genes in unbalanced genomes using cycle packing of adjacency graphs. *Journal of Heuristics*.

Siqueira, G., Oliveira Alexandrino, A., Rodrigues Oliveira, A., Jean, G., Fertin, G., and Dias, Z. (2023b). Approximating Rearrangement Distances with Replicas and Flexible Intergenic Regions. In *Proceedings of the 19th International Symposium on Bioinformatics Research and Applications (ISBRA'2023)*, volume 14248, pages 241–254. Springer.

Willing, E., Stoye, J., and Braga, M. (2021). Computing the Inversion-Indel Distance. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6):2314–2326.

Willing, E., Zaccaria, S., Braga, M. D., and Stoye, J. (2013). On the Inversion-Indel Distance. *BMC Bioinformatics*, 14:S3.