

# Variações do Problema de Distância de Rearranjos

Alexsandro Oliveira Alexandrino

---

12 de Abril de 2021

**Orientador:** Prof. Dr. Zanoni Dias

**Coorientador:** Prof. Dr. Ulisses Dias

Instituto de Computação - Universidade Estadual de Campinas

1. Introdução
2. Conceitos
3. Trabalhos Relacionados
4. Objetivos e Metodologia
5. Plano de Trabalho
6. Resultados Preliminares

# Introdução

---

- Evolução:
  - Resultado de mutações ocorridas no material genético
- Mutações pontuais:
  - Alterações de bases na sequência do DNA
  - Ocorrem com maior frequência
- Rearranjos de genoma:
  - Mutações de larga escala que afetam a posição e a orientação de grandes trechos do DNA
- A distância de rearranjos é bastante utilizada como estimativa para a distância evolucionária entre espécies

# Distância de Rearranjos

- A ocorrência de um rearranjo  $\beta$  está vinculada a um custo  $w(\beta)$
- O custo de uma sequência de rearranjos é o somatório dos custos dos rearranjos da sequência
- Para dois genomas  $\mathcal{G}_1$  e  $\mathcal{G}_2$ , a distância de rearranjos é o menor custo possível para uma sequência de rearranjos transformar  $\mathcal{G}_1$  em  $\mathcal{G}_2$

# Variações da Distância de Rearranjos

- Os problemas de distância de rearranjos variam de acordo com:
  - Representação do genoma
    - Genes repetidos ou não repetidos
    - O conjunto de genes de cada genoma é igual ou distinto
    - Uso de regiões intergênicas ou apenas ordem dos genes
  - Conjunto de rearranjos permitidos (modelo de rearranjo)
  - Função de custo
    - Constante
    - Por tipo
    - Proporcional ao tamanho da região afetada

# Conceitos

---

## Representação: Ordem Relativa dos Genes

- A maioria dos estudos usa apenas a ordem relativa dos genes de cada genoma, supondo que:
  - Não existem genes duplicados
  - Os dois genomas comparados compartilham os mesmos genes
- Assim, um genoma  $\mathcal{G}$  com sequência de genes  $(g_1 g_2 \dots g_n)$  é representado por uma permutação  $\pi = (\pi_1 \pi_2 \dots \pi_n)$ , onde um número inteiro  $\pi_i$  corresponde a um gene  $g_i$  no genoma  $\mathcal{G}$
- Se a orientação dos genes é conhecida, cada elemento possui um sinal  $+$  ou  $-$  para indicar a orientação do gene correspondente

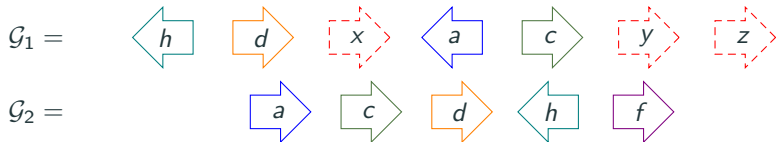


- Ao utilizar permutações, um dos genomas é representado como a permutação identidade  $\iota = (1\ 2\ \dots\ n)$
- O problema se torna equivalente ao problema da Ordenação de Permutações por Rearranjos

## Representação: Ordem Relativa dos Genes

- Quando uma das seguintes condições é atendida, a representação da ordem relativa dos genes é feita com strings
  - Genes repetidos
  - O conjunto de genes de cada genoma é distinto
- Assim, um genoma  $\mathcal{G} = (g_1 \ g_2 \ \dots \ g_n)$  é representado por uma string  $\sigma = (\sigma_1 \ \sigma_2 \ \dots \ \sigma_n)$
- Se a orientação dos genes é conhecida, cada elemento possui um sinal + ou - para indicar a orientação do gene correspondente

## Representação: Ordem Relativa dos Genes

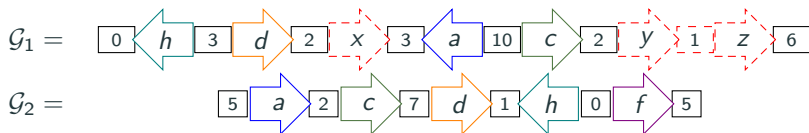


- Segmentos a serem removidos são mapeados com o valor  $\alpha$ , já que são removidos independente do seu conteúdo
- $\mathcal{G}_2$  é mapeado em  $\iota$ :
  - $\iota = (+1 +2 +3 +4 +5)$
- $\mathcal{G}_1$  é mapeado em  $\sigma$ :
  - $\sigma = (+4 +3 \alpha -1 +2 \alpha)$

# Representação: Regiões Intergênicas

- Regiões intergênicas:
  - Nucleotídeos entre genes e nas extremidades do genoma
  - Mais suscetíveis a mudanças do que genes
  - Normalmente não possuem correspondência entre genomas e, portanto, apenas a quantidade de nucleotídeos (tamanho) de cada região é usada
- Um genoma  $\mathcal{G} = (\check{g}_1 g_1 \check{g}_2 g_2 \dots \check{g}_n g_n \check{g}_{n+1})$  é representado por uma string  $\sigma$  e uma lista de regiões intergênicas  $\check{\sigma}$ .
  - Cada gene  $g_i$  é representado pelo caractere  $\sigma_i$
  - O tamanho da região intergênica  $\check{g}_i$  é representado pelo número  $\check{\sigma}_i$

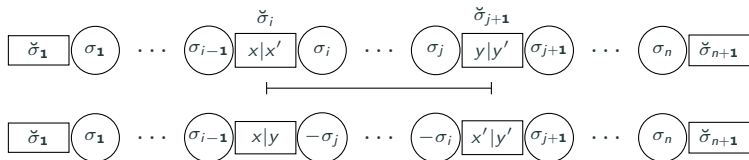
## Representação: Regiões Intergênicas



- $\mathcal{G}_2$  é mapeado em  $(\iota, \check{\iota})$ :
  - $\iota = (+1 +2 +3 +4 +5)$
  - $\check{\iota} = (5, 2, 7, 1, 0, 5)$
- $\mathcal{G}_1$  é mapeado em  $(\sigma, \check{\sigma})$ :
  - $\sigma = (+4 +3 \alpha -1 +2 \alpha)$
  - $\check{\sigma} = (0, 3, 2, 3, 10, 2, 6)$

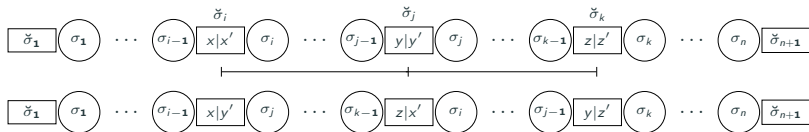
- Os rearranjos considerados nesta proposta de doutorado são:
  - Reversão  $\rho$
  - Transposição  $\tau$
  - Inserção  $\phi$  e Deleção  $\psi$  (Indels)
  - Transposição Inversa  $\rho\tau$
  - Revreves  $\rho\rho$
  - Block Interchange  $bi$
- Exceto pelos indels, todos os rearranjos mencionados são conservativos (*i.e.*, não alteram a quantidade de material genético)
- A ocorrência de um rearranjo  $\beta$  é denotada por  $\sigma \cdot \beta$  e  $\check{\sigma} \cdot \beta$

- Uma reversão  $\rho_{(x,y)}^{(i,j)}$  divide as regiões intergênicas  $\check{\sigma}_i$  e  $\check{\sigma}_{j+1}$  em  $x|x'$  e  $y|y'$ , respectivamente, e inverte o segmento  $(x' \sigma_i \check{\sigma}_{i+1} \dots \sigma_j y)$
- Em strings com sinais, os sinais dos elementos no segmento invertido são alterados



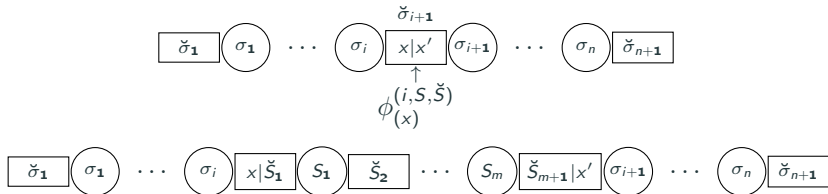
# Transposição

- Uma transposição  $\tau_{(x,y,z)}^{(i,j,k)}$  divide as regiões intergênicas  $\check{\sigma}_i$ ,  $\check{\sigma}_j$  e  $\check{\sigma}_k$  em  $x|x'$ ,  $y|y'$  e  $z|z'$ , respectivamente, e troca a posição dos segmentos  $(x' \sigma_i \check{\sigma}_{i+1} \dots \sigma_{j-1} y)$  e  $(y' \sigma_j \check{\sigma}_{j+1} \dots \sigma_{k-1} z)$

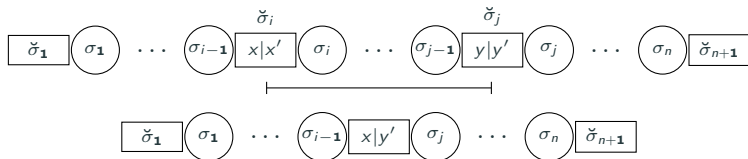




- Uma inserção  $\phi_{(x)}^{(i,S,\check{S})}$  é uma operação que adiciona o segmento  $(\check{S}_1 S_1 \check{S}_2 \dots S_m \check{S}_{m+1})$ , onde  $|S| = m$ , após o  $x$ -ésimo nucleotídeo da região intergênica  $\check{\sigma}_{i+1}$



- Uma deleção  $\psi_{(x,y)}^{(i,j)}$  divide as regiões intergênicas  $\check{\sigma}_i$  e  $\check{\sigma}_j$  em  $x|x'$  e  $y|y'$ , respectivamente, e remove o segmento  $(x' \sigma_i \check{\sigma}_{i+1} \dots \sigma_{j-1} y)$



- Transposição inversa:

$$\sigma = (\sigma_1 \dots \sigma_{i-1} \underline{\sigma_i \dots \sigma_{j-1}} \underline{\sigma_j \dots \sigma_{k-1}} \sigma_k \dots \sigma_n)$$

$$\sigma \cdot \bar{\rho}\tau_1(i, j, k) = (\sigma_1 \dots \sigma_{i-1} \underline{\sigma_j \dots \sigma_{k-1}} \underline{-\sigma_{j-1} \dots -\sigma_i} \sigma_k \dots \sigma_n)$$

$$\sigma \cdot \bar{\rho}\tau_2(i, j, k) = (\sigma_1 \dots \sigma_{i-1} \underline{-\sigma_{k-1} \dots -\sigma_j} \underline{\sigma_i \dots \sigma_{j-1}} \sigma_k \dots \sigma_n)$$

- Revrev:

$$\sigma = (\sigma_1 \dots \sigma_{i-1} \underline{\sigma_i \dots \sigma_{j-1}} \underline{\sigma_j \dots \sigma_{k-1}} \sigma_k \dots \sigma_n)$$

$$\sigma \cdot \bar{\rho}\bar{\rho}(i, j, k) = (\sigma_1 \dots \sigma_{i-1} \underline{-\sigma_{j-1} \dots -\sigma_i} \underline{-\sigma_{k-1} \dots -\sigma_j} \sigma_k \dots \sigma_n)$$

- Block interchange:

$$\sigma = (\sigma_1 \dots \sigma_{i-1} \underline{\sigma_i \dots \sigma_j} \dots \sigma_{x-1} \underline{\sigma_x \dots \sigma_y} \dots \sigma_n)$$

$$\sigma \cdot bi(i, j, x, y) = (\sigma_1 \dots \sigma_{i-1} \underline{\sigma_x \dots \sigma_y} \dots \sigma_{x-1} \underline{\sigma_i \dots \sigma_j} \dots \sigma_n)$$

- Modelo de rearranjo  $\mathcal{M}$ : define o conjunto de rearranjos permitidos no problema
- Função de custo  $w : \mathcal{M} \rightarrow \mathbb{R}$ : associa um custo à ocorrência de um rearranjo pertencente a  $\mathcal{M}$
- O tamanho de um rearranjo  $\beta$  é igual à quantidade de elementos afetados da string ( $|\beta|$ )
- Uma  $\lambda$ -operação é um rearranjo com tamanho menor ou igual a  $\lambda$

# Distância de Rearranjos em Permutações

- Distância de rearranjos (ordenação)  $d_{\mathcal{M}}^w(\pi) = \sum_{i=1}^{\ell} w(\beta_i)$  é tal que
  - $\beta_i \in \mathcal{M}$ , para  $1 \leq i \leq \ell$
  - $\pi \cdot \beta_1 \cdot \dots \cdot \beta_{\ell} = \iota$
  - $\sum_{i=1}^{\ell} w(\beta_i)$  é mínimo
- Abordagem tradicional:  $w(\beta) = 1$
- Ponderação pelo tamanho:  $w(\beta) = |\beta|^{\alpha}$ , um parâmetro  $\alpha > 0$
- Ponderação pelo tipo:
  - Reversões possuem peso  $w_{\rho}$
  - Transposições, transposições inversas e revrevs possuem peso  $w_{\tau}$
  - Normalmente,  $1 \leq w_{\tau}/w_{\rho} \leq 2$
- Ponderação pelo número de fragmentações (quebras de adjacência) causadas pela operação [2, 3]

# Variações da Distância de Rearranjos em Permutações

- Ordenação de Permutações por Rearranjos (não ponderados)
- Ordenação de Permutações por Rearranjos Ponderados pelo Tipo
- Ordenação de Permutações por Rearranjos Ponderados pelo Número de Fragmentações
- Ordenação de Permutações por  $\lambda$ -Operações (não ponderados)
- Ordenação de Permutações por  $\lambda$ -Operações Ponderadas pelo Tamanho

# Distância de Rearranjos: Strings e Regiões Intergênicas

- Para duas strings  $\sigma$  e  $\sigma'$ , a distância de rearranjos  $d_{\mathcal{M}}^w(\sigma, \sigma') = \sum_{i=1}^{\ell} w(\beta_i)$  é tal que
  - $\beta_i \in \mathcal{M}$ , para  $1 \leq i \leq \ell$
  - $\sigma \cdot \beta_1 \cdot \dots \cdot \beta_{\ell} = \sigma'$
  - $\sum_{i=1}^{\ell} w(\beta_i)$  é mínimo
- Variações para strings: conjuntos distintos de genes, genes duplicados, genes multiplicados
- Para dois genomas  $\mathcal{G}_1 = (\sigma, \check{\sigma})$  e  $\mathcal{G}_2 = (\sigma', \check{\sigma}')$ , a distância de rearranjos  $d_{\mathcal{M}}^w(\sigma, \sigma') = \sum_{i=1}^{\ell} w(\beta_i)$  é tal que
  - $\beta_i \in \mathcal{M}$ , para  $1 \leq i \leq \ell$
  - $\sigma \cdot \beta_1 \cdot \dots \cdot \beta_{\ell} = \sigma'$
  - $\check{\sigma} \cdot \beta_1 \cdot \dots \cdot \beta_{\ell} = \check{\sigma}'$
  - $\sum_{i=1}^{\ell} w(\beta_i)$  é mínimo
- As mesmas funções de custo previamente mencionadas podem ser usadas para essas representações

# Trabalhos Relacionados

---



- Hannenhalli e Pevzner (1999) [25] apresentaram um algoritmo polinomial para a Ordenação de Permutações com Sinais por Reversões
- Para permutações sem sinais, Caprara (1999) [20] provou que a Ordenação de Permutações por Reversões é NP-difícil
- Bultheau e coautores (2012) [18] provaram que a Ordenação de Permutações por Transposições é NP-difícil
- Oliveira e coautores (2019) [34] provaram que a Ordenação de Permutações por Reversões e Transposições é NP-difícil quando  $w_\tau/w_\rho \leq 1.5$ , para permutações com e sem sinais

- Para modelos com transposições inversas e revrevs, existem algoritmos de aproximação na literatura, apesar da complexidade ser desconhecida no momento da publicação desses trabalhos

	$\tau$	$\rho$	$\rho\tau$	$\rho\rho$	Sinais	Função de custo	Aproximação
Gu e coautores (1999) [24]	Sim	Sim	Sim	-	Sim	Constante	2
Lin e Xue (2001) [29]	Sim	Sim	Sim	Sim	Sim	Constante	1.75
Hartman e Sharan (2005) [26]	Sim	-	Sim	Sim	Sim	Constante	1.5
Lou e Zhu (2008) [31]	Sim	Sim	Sim	-	Não	Constante	2.25
Bader e coautores (2008) [9]	Sim	Sim	Sim	-	Sim	$1 \leq w_\tau/w_\rho \leq 2$	1.5
Eriksen (2002) [21]	Sim	Sim	Sim	-	Sim	$w_\tau/w_\rho = 2$	7/6

- Variações motivadas pela observação de que, em algumas espécies, os rearranjos que ocorreram durante o processo evolutivo não agem em regiões grandes [11, 28]
- O custo de um rearranjo  $\beta$  é igual a  $|\beta|^\alpha$ , onde  $\alpha \geq 0$  é um parâmetro de valor real
- Estudos existentes na literatura para ambas abordagens tratam genomas sem genes repetidos

- Para qualquer  $\lambda > 1$  e combinação de reversões e transposições, existem  $O(\lambda^2)$ -aproximações [32]
- Para  $\lambda = 3$  (operações curtas), existem aproximações constantes para combinações de reversões e transposições [23, 38]
- Para  $\lambda = 2$  (operações super curtas), todos os problemas envolvendo reversões e transposições possuem algoritmo exato polinomial [23, 27]

# Ordenação de Permutações Ponderadas pelo Tamanho

- Para  $\alpha = 1$ :
  - $O(\log n)$ -aproximação para reversões [10]
  - $O(\log^2 n)$ -aproximação para transposições e combinação de reversões e transposições [30]
- Para  $\alpha \geq 2$ :
  - Aproximações com fator constante para combinações de reversões e transposições [10, 30]
- Para  $\alpha \geq 3$  em permutações sem sinais:
  - Algoritmos exatos polinomiais para combinações de reversões e transposições [30]
- Nguyen e coautores (2005) [33] combinaram ambas as abordagens considerando apenas reversões em permutações com sinais
  - $O(\log n)$ -aproximação, para  $\lambda = \Omega(n)$  e  $\alpha = 1$
  - $(2 \log^2 n + \log n)$ -aproximação, para  $\lambda = o(n)$  e  $\alpha = 1$

- Braga e coautores (2011) [12] apresentaram um algoritmo exato polinomial para DCJs em strings com sinais
  - DCJ (*Double-Cut-and-Join*) é uma operação genérica que simula reversões e outros rearranjos de genomas
- Willing e coautores (2020) [39] apresentaram um algoritmo exato polinomial para reversões e indels em strings com sinais
- Não existem algoritmos conhecidos para os outros modelos

# Distância de Rearranjos de Genoma Considerando Regiões Intergênicas

- Nos trabalhos apresentados a seguir, inserções e deleções não afetam genes

	DCJ	$\rho$	$\tau$	Indels	Sinais	Classe	Resultado
Bulteau e coautores (2016) [19]	Sim	-	-	Sim	Sim	P	Algoritmo exato
Fertin e coautores (2017) [22]	Sim	-	-	-	Sim	NP-difícil	4/3-aproximação
Oliveira e coautores (2020) [35]	-	Sim	-	-	Sim	NP-difícil	2-aproximação
Oliveira e coautores (2020) [35]	-	Sim	-	Sim	Sim	Desconhecida	2-aproximação
Brito e coautores (2020) [16]	-	Sim	-	-	Não	NP-difícil	4-aproximação
Brito e coautores (2020) [16]	-	Sim	-	Sim	Não	NP-difícil	4-aproximação
Brito e coautores (2020) [16]	-	Sim	Sim	-	Não	NP-difícil	4.5-aproximação
Brito e coautores (2020) [16]	-	Sim	Sim	Sim	Não	NP-difícil	4.5-aproximação
Oliveira e coautores (2020) [36]	-	-	Sim	-	Não	NP-difícil	3.5-aproximação

# Objetivos e Metodologia

---



- O objetivo é estudar modelos que possuem transposições junto com combinações de reversões, transposições inversas e revrevs
- Consideramos o problema não ponderado e o ponderado por tipo
- Além disso, estudar a complexidade de problemas ponderados pelo número de fragmentações (quebras de elementos adjacentes)

# Ordenação de Permutações por $\lambda$ -Operações Ponderadas pelo Tamanho

- Para essa variação, consideramos os seguintes modelos:
  - $\lambda$ -Reversões para permutações com ou sem sinais
  - $\lambda$ -Transposições para permutações sem sinais
  - $\lambda$ -Reversões e  $\lambda$ -Transposições para permutações com ou sem sinais
- Investigação do problema para diferentes valores de  $\lambda$  e  $\alpha$
- Proposta de limitantes inferiores e algoritmos de aproximação

- Para essa variação, consideramos os seguintes modelos:
  - Reversões, Inserções e Deleções para strings com ou sem sinais
  - Transposições, Inserções e Deleções para strings sem sinais
  - Reversões, Transposições, Inserções e Deleções para strings com ou sem sinais
- Proposta de limitantes inferiores e algoritmos de aproximação

# Distância de Rearranjos de Genoma considerando Regiões Intergênicas

- Para essa variação, consideramos os seguintes modelos:
  - Reversões, Inserções e Deleções para strings com ou sem sinais
  - Transposições, Inserções e Deleções para strings sem sinais
  - Reversões, Transposições, Inserções e Deleções para strings com ou sem sinais
- As operações de inserção e deleção consideradas também afetam genes
- Proposta de limitantes inferiores e algoritmos de aproximação

- Para os dois últimos objetivos, dividiremos o estudo em:
  - Genomas sem genes repetidos
  - Genomas com genes repetidos (duplicados e multiplicados)

- O primeiro objetivo tem foco teórico
- Os outros objetivos possuem várias etapas, agrupadas em:
  - Definição formal das variações da distância de rearranjos
  - Definição de estruturas de dados
  - Definição de limitantes para a distância
  - Desenvolvimento de algoritmos
  - Criação de bases de testes
  - Análise dos resultados

# Plano de Trabalho

---

# Cronograma de Atividades

Atividades	Semestres							
	1	2	3	4	5	6	7	8
1	x	x						
2		x		x	x			
3	x	x	x	x				
4					x			
5	x	x	x					
6		x	x	x	x			
7				x	x	x		
8				x	x	x	x	
9							x	x
10								x
11								x

1. Obtenção dos créditos obrigatórios em disciplinas do programa de doutorado;
2. Participação no Programa de Estágio Docente (PED);
3. Revisão da literatura;
4. Exame de Qualificação Específico (EQE);



# Cronograma de Atividades

Atividades	Semestres							
	1	2	3	4	5	6	7	8
1	x	x						
2		x		x	x			
3	x	x	x	x				
4					x			
5	x	x	x					
6		x	x	x	x			
7				x	x	x		
8				x	x	x	x	
9							x	x
10								x
11								x

- Investigação da complexidade dos problemas de Ordenação de Permutações por Transposições e Outros Rearranjos;
- Investigação dos problemas de Ordenação de Permutações por Operações Ponderadas e Limitadas pelo Tamanho;
- Investigação dos problemas de Distância de Rearranjos de Genoma com modelos incluindo inserções e deleções;

# Cronograma de Atividades

Atividades	Semestres							
	1	2	3	4	5	6	7	8
1	x	x						
2		x		x	x			
3	x	x	x	x				
4					x			
5	x	x	x					
6		x	x	x	x			
7				x	x	x		
8				x	x	x	x	
9							x	x
10								x
11								x

- Investigação dos problemas de Distância de Rearranjos de Genoma com modelos incluindo inserções e deleções considerando regiões intergênicas;
- Escrita da tese;
- Revisão da tese;
- Defesa da tese.

# Resultados Preliminares

---

- Prova de NP-dificuldade para modelos que contém transposições e a combinação de reversões, transposições inversas e revrevs para  $w_\tau/w_\rho \leq 1.5$
- Prova de NP-dificuldade considerando a ponderação pelo número de fragmentações para modelos com transposições
- “*On the Complexity of Some Variations of Sorting by Transpositions*” (Alexsandro O. Alexandrino, Andre R. Oliveira, Ulisses Dias, Zanoni Dias), *Journal of Universal Computer Science* (2020) [5]

# Ordenação de Permutações por $\lambda$ -Operações Ponderadas pelo Tamanho

- Sumário dos resultados obtidos para essa variação do problema:

Parâmetros	Fator de Aproximação Obtido				
	Sem Sinais			Com Sinais	
	$\mathcal{M}_\rho^\lambda$	$\mathcal{M}_\tau^\lambda$	$\mathcal{M}_{\rho,\tau}^\lambda$	$\mathcal{M}_\rho^\lambda$	$\mathcal{M}_{\rho,\tau}^\lambda$
$\lambda > 1$ e $\alpha = 1$	$\lambda - 1$	$\lambda/2$	$\lambda - 1$	$\lambda$	$\lambda$
$\lambda = \Theta(n)$ e $\alpha = 1$	$O(\log n)$	$O(\log^2 n)$	$O(\log^2 n)$	$O(\log n)$	$O(\log^2 n)$
$\lambda = 3$ e $\alpha = 1$	2	4/3	2	3	7/3
$\lambda > 1$ e $\alpha \geq 2$	2	2	2	3	3
$\lambda > 1$ e $\alpha \geq 3$	Exato	Exato	Exato	Exato	Exato

- “Length-Weighted  $\lambda$ -Rearrangement Distance” (Alexsandro O. Alexandrino, Guilherme H. S. Miranda, Carla N. Lintzmayer, Zanoni Dias), *Journal of Combinatorial Optimization* (2021) [4]

- 2-aproximação para reversões em strings sem sinais
- 3-aproximação para transposições e a combinação de reversões e transposições em strings com e sem sinais
- “*Genome Rearrangement Distance with Reversals, Transpositions, and Indels*” (Alexsandro O. Alexandrino, Andre R. Oliveira, Ulisses Dias, Zanoni Dias), *Journal of Computational Biology* (2021) [6]

- 2-aproximação para block interchanges
- 2-aproximação para transposições
- “*Labeled Cycle Graph for Transposition and Indel Distance*”  
(Alexsandro O. Alexandrino, Andre R. Oliveira, Ulisses Dias, Zanoni Dias) [8]

# Distância de Rearranjos de Genoma Considerando Regiões Intergênicas

- 2-aproximação para reversões em strings com sinais
- “*Reversal Distance on Genomes with Different Gene Content and Intergenic Regions Information*” (Alexsandro Oliveira Alexandrino, Klairton Lima Brito, Andre Rodrigues Oliveira, Ulisses Dias, Zanoni Dias), *International Conference on Algorithms for Computational Biology* (ALCoB'2021) [1]



# Distância de Rearranjos de Genoma Considerando Regiões Intergênicas

- 4-aproximação para reversões em strings sem sinais
- 4.5-aproximação para transposições e a combinação de reversões e transposições em strings sem sinais
- *“Incorporating Intergenic Regions into Reversal and Transposition Distances with Indels”* (Alexsandro O. Alexandrino, Andre R. Oliveira, Ulisses Dias, Zanoni Dias) [7]

- “*Heuristics for Breakpoint Graph Decomposition with Applications in Genome Rearrangement Problems*” (Pedro Olímpio Pinheiro, Alexandro Oliveira Alexandrino, Andre Rodrigues Oliveira, Cid Carvalho de Souza, Zaroni Dias), *Brazilian Symposium on Bioinformatics* (BSB'2020) [37]
- “*Sorting by Reversals and Transpositions with Proportion Restriction*” (Klairton Lima Brito, Alexandro Oliveira Alexandrino, Andre Rodrigues Oliveira, Ulisses Dias, Zaroni Dias), *Brazilian Symposium on Bioinformatics* (BSB'2020) [13, 14]

- *“Reversal and Transposition Distance of Genomes Considering Flexible Intergenic Regions”* (Klairton Lima Brito, Andre Rodrigues Oliveira, Aleksandro Oliveira Alexandrino, Ulisses Dias, Zanoni Dias), *Latin and American Algorithms, Graphs and Optimization Symposium (LAGOS'2021)* [17]
- *“Reversals Distance Considering Flexible Intergenic Regions Sizes”* (Klairton Lima Brito, Aleksandro Oliveira Alexandrino, Andre Rodrigues Oliveira, Ulisses Dias, Zanoni Dias), *International Conference on Algorithms for Computational Biology (AlCoB'2021)* [15]

- [1] Alexandrino, A. O., Brito, K. L., Oliveira, A. R., Dias, U., and Dias, Z. (2021a). Reversal Distance on Genomes with Different Gene Content and Intergenic Regions Information. In *Proceedings of the 8th International Conference on Algorithms for Computational Biology (AICoB'2021)*, pages 121–133. Springer International Publishing.
- [2] Alexandrino, A. O., Lintzmayer, C. N., and Dias, Z. (2018). Approximation Algorithms for Sorting Permutations by Fragmentation-Weighted Operations. In *Algorithms for Computational Biology*, volume 10849, pages 53–64. Springer International Publishing, Heidelberg, Germany.
- [3] Alexandrino, A. O., Lintzmayer, C. N., and Dias, Z. (2020a). Sorting Permutations by Fragmentation-Weighted Operations. *Journal of Bioinformatics and Computational Biology*, 18(2):2050006.1–2050006.31.

- [4] Alexandrino, A. O., Miranda, G. H. S., Lintzmayer, C. N., and Dias, Z. (2020b). Length-weighted  $\lambda$ -rearrangement Distance. *Journal of Combinatorial Optimization*, pages 1–24.
- [5] Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2020c). On the Complexity of Some Variations of Sorting by Transpositions. *Journal of Universal Computer Science*, 26(9):1076–1094.
- [6] Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2021b). Genome Rearrangement Distance with Reversals, Transpositions, and Indels. *Journal of Computational Biology*, 28(3):235–247.
- [7] Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2021c). Incorporating intergenic regions into reversal and transposition distances with indels. *Journal of Bioinformatics and Computational Biology*, 19(06):2140011.

- [8] Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2021d). Labeled cycle graph for transposition and indel distance. *Journal of Computational Biology*.
- [9] Bader, M., Abouelhoda, M. I., and Ohlebusch, E. (2008). A Fast Algorithm for the Multiple Genome Rearrangement Problem with Weighted Reversals and Transpositions. *BMC Bioinformatics*, 9(1):1–13.
- [10] Bender, M. A., Ge, D., He, S., Hu, H., Pinter, R. Y., Skiena, S. S., and Swidan, F. (2008). Improved Bounds on Sorting by Length-Weighted Reversals. *Journal of Computer and System Sciences*, 74(5):744–774.
- [11] Blanchette, M., Kunisawa, T., and Sankoff, D. (1996). Parametric Genome Rearrangement. *Gene*, 172(1):GC11–GC17.

- [12] Braga, M. D., Willing, E., and Stoye, J. (2011). Double cut and join with insertions and deletions. *Journal of Computational Biology*, 18(9):1167–1184.
- [13] Brito, K. L., Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2020a). Sorting by Reversals and Transpositions with Proportion Restriction. In *Proceedings of the 13th Brazilian Symposium on Bioinformatics (BSB'2020)*, pages 117–128. Springer International Publishing.
- [14] Brito, K. L., Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2021a). Reversals and Transpositions Distance with Proportion Restriction. *Journal of Bioinformatics and Computational Biology*, 19(04):2150013.

- [15] Brito, K. L., Alexandrino, A. O., Oliveira, A. R., Dias, U., and Dias, Z. (2021b). Reversals Distance Considering Flexible Intergenic Regions Sizes. In *Proceedings of the 8th International Conference on Algorithms for Computational Biology (AICoB'2021)*, pages 134–145. Springer International Publishing.
- [16] Brito, K. L., Jean, G., Fertin, G., Oliveira, A. R., Dias, U., and Dias, Z. (2020b). Sorting by Genome Rearrangements on both Gene Order and Intergenic Sizes. *Journal of Computational Biology*, 27(2):156–174.
- [17] Brito, K. L., Oliveira, A. R., Alexandrino, A. O., Dias, U., and Dias, Z. (2021c). Reversal and Transposition Distance of Genomes Considering Flexible Intergenic Regions. In *Proceedings of the XI Latin and American Algorithms, Graphs and Optimization Symposium (LAGOS'2021)*, pages 21–29. Procedia Computer Science, Elsevier.



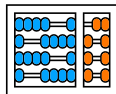
- [18] Bulteau, L., Fertin, G., and Rusu, I. (2012). Sorting by Transpositions is Difficult. *SIAM Journal on Discrete Mathematics*, 26(3):1148–1180.
- [19] Bulteau, L., Fertin, G., and Tannier, E. (2016). Genome Rearrangements with Indels in Intergenes Restrict the Scenario Space. *BMC Bioinformatics*, 17(14):426.
- [20] Caprara, A. (1999). Sorting Permutations by Reversals and Eulerian Cycle Decompositions. *SIAM Journal on Discrete Mathematics*, 12(1):91–110.
- [21] Eriksen, N. (2002).  $(1+\epsilon)$ -Approximation of Sorting by Reversals and Transpositions. *Theoretical Computer Science*, 289(1):517–529.
- [22] 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.

- [23] Galvão, G. R., Lee, O., and Dias, Z. (2015). Sorting Signed Permutations by Short Operations. *Algorithms for Molecular Biology*, 10(1):1–17.
- [24] Gu, Q.-P., Peng, S., and Sudborough, I. H. (1999). A 2-Approximation Algorithm for Genome Rearrangements by Reversals and Transpositions. *Theoretical Computer Science*, 210(2):327–339.
- [25] 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.
- [26] Hartman, T. and Sharan, R. (2005). A 1.5-Approximation Algorithm for Sorting by Transpositions and Transreversals. *Journal of Computer and System Sciences*, 70(3):300–320.
- [27] Jerrum, M. R. (1985). The Complexity of Finding Minimum-length Generator Sequences. *Theoretical Computer Science*, 36(2-3):265–289.

- [28] Lefebvre, J.-F., El-Mabrouk, N., Tillier, E. R. M., and Sankoff, D. (2003). Detection and validation of single gene inversions. *Bioinformatics*, 19(1):i190–i196.
- [29] Lin, G.-H. and Xue, G. (2001). Signed Genome Rearrangement by Reversals and Transpositions: Models and Approximations. *Theoretical Computer Science*, 259(1-2):513–531.
- [30] Lintzmayer, C. N., Fertin, G., and Dias, Z. (2015). Approximation Algorithms for Sorting by Length-Weighted Prefix and Suffix Operations. *Theoretical Computer Science*, 593:26–41.
- [31] Lou, X. and Zhu, D. (2008). A 2.25-Approximation Algorithm for Cut-and-Paste Sorting of Unsigned Circular Permutations. In *Computing and Combinatorics*, volume 5092, pages 331–341, Heidelberg, Germany. Springer International Publishing.

- [32] Miranda, G. H. S., Lintzmayer, C. N., and Dias, Z. (2018). Sorting Permutations by Limited-Size Operations. In *Algorithms for Computational Biology*, volume 10849, pages 76–87. Springer International Publishing, Heidelberg, Germany.
- [33] Nguyen, T. C., Ngo, H. T., and Nguyen, N. B. (2005). Sorting by Restricted-Length-Weighted Reversals. *Genomics Proteomics & Bioinformatics*, 3(2):120–127.
- [34] 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.
- [35] Oliveira, A. R., Jean, G., Fertin, G., Brito, K. L., Bulteau, L., Dias, U., and Dias, Z. (2020a). Sorting Signed Permutations by Intergenic Reversals. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

- [36] Oliveira, A. R., Jean, G., Fertin, G., Brito, K. L., Dias, U., and Dias, Z. (2020b). A 3.5-Approximation Algorithm for Sorting by Intergenic Transpositions. In *Algorithms for Computational Biology*, volume 12099, pages 16–28, Heidelberg, Germany. Springer International Publishing.
- [37] Pinheiro, P. O., Alexandrino, A. O., Oliveira, A. R., de Souza, C. C., and Dias, Z. (2020). Heuristics for Breakpoint Graph Decomposition with Applications in Genome Rearrangement Problems. In *Proceedings of the 13th Brazilian Symposium on Bioinformatics (BSB'2020)*, pages 129–140. Springer International Publishing.
- [38] Vergara, J. P. C. (1998). *Sorting by Bounded Permutations*. PhD thesis, Virginia Polytechnic Institute and State University.
- [39] Willing, E., Stoye, J., and Braga, M. (2020). Computing the inversion-indel distance. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.



# Variações do Problema de Distância de Rearranjos

Alexsandro Oliveira Alexandrino

---

12 de Abril de 2021

**Orientador:** Prof. Dr. Zanoni Dias

**Coorientador:** Prof. Dr. Ulisses Dias

Instituto de Computação - Universidade Estadual de Campinas