

# Proposta de Tese de Doutorado

Algoritmos para o Problema da Ordenação por Rearranjo

Aluno: Gustavo Rodrigues Galvão

Orientador: Zanoni Dias



25 de Abril de 2014

- 1 Introdução
- 2 Objetivos
- 3 Cronograma

# Rearranjo de Genomas

## Problema Motivador

*Estimar a distância evolutiva entre espécies.*

## Proposta de Solução

*Dados dois genomas relativos a duas espécies, calcula-se a distância de rearranjo entre eles, que é igual ao tamanho da menor sequência de mutações globais, chamadas de eventos de rearranjo, que transforma um genoma no outro. Utiliza-se, então, a distância de rearranjo como estimativa para a distância evolutiva entre as espécies.*

# Modelando Genomas

- Premissas quanto aos genomas:
  - Genoma unicromossomal.
  - Mesmo conjunto de genes.
  - Sem repetição de genes.
- Representação clássica: permutações de números inteiros.
  - $\pi = (\pi_1 \pi_2 \dots \pi_n)$ .
- Cada elemento  $\pi_i$  possui um sinal, + ou -, que indica a orientação do gene que ele representa.
  - Omissão do sinal caso não haja informação sobre a orientação.

# Modelando Eventos de Rearranjo

## Reversão

A reversão  $\rho(i, j)$ ,  $1 \leq i < j \leq n$ , aplicada na permutação sem sinal

$$(\pi_1 \ \pi_2 \ \dots \ \pi_{i-1} \ \underline{\pi_i \ \pi_{i+1} \ \dots \ \pi_{j-1} \ \pi_j} \ \pi_{j+1} \ \dots \ \pi_n)$$

a transforma na permutação sem sinal

$$(\pi_1 \ \pi_2 \ \dots \ \pi_{i-1} \ \underline{\pi_j \ \pi_{j-1} \ \dots \ \pi_{i+1} \ \pi_i} \ \pi_{j+1} \ \dots \ \pi_n).$$

Caso a permutação original tivesse sinal, então a reversão a transformaria na permutação com sinal

$$(\pi_1 \ \pi_2 \ \dots \ \pi_{i-1} \ \underline{-\pi_j \ -\pi_{j-1} \ \dots \ -\pi_{i+1} \ -\pi_i} \ \pi_{j+1} \ \dots \ \pi_n).$$

# Modelando Eventos de Rearranjo

## Transposição

A transposição  $\rho(i, j, k)$ ,  $1 \leq i < j < k \leq n + 1$ , aplicada na permutação (com ou sem sinal)

$$(\pi_1 \dots \pi_{i-1} \underline{\pi_i \dots \pi_{j-1}} \underline{\pi_j \dots \pi_{k-1}} \pi_k \dots \pi_n)$$

a transforma na permutação (com ou sem sinal)

$$(\pi_1 \dots \pi_{i-1} \underline{\pi_j \dots \pi_{k-1}} \underline{\pi_i \dots \pi_{j-1}} \pi_k \dots \pi_n).$$

# Problema da Ordenação por Rearranjo

## Definição do Problema

*Dada uma permutação  $\pi$ , encontrar a menor sequência de eventos de rearranjo pertencentes a um modelo de rearranjo que transforma  $\pi$  na permutação identidade  $\iota = (1\ 2\ \dots\ n)$ .*

## Modelo de Rearranjo

*Conjunto de eventos de rearranjo permitidos para transformar uma permutação em outra.*

Modelo de Rearranjo	Complexidade	Melhor Solução Teórica
Reversões com Sinal	Polinomial	Algoritmo exato $O(n^{\frac{3}{2}} \sqrt{\log n})$
Reversões	NP-Difícil	Algoritmo 1.375-aproximado
Transposições	NP-Difícil	Algoritmo 1.375-aproximado
Transposições Curtas	?	Algoritmo $(1+\epsilon)$ -aproximado
Transposições e Transreversões do Tipo A e do Tipo B	?	Algoritmo 1.5-aproximado
Reversões de Prefixo	NP-Difícil	Algoritmo 2-aproximado
Reversões de Prefixo com Sinal	?	Algoritmo 2-aproximado
Reversões Curtas	?	Algoritmo 2-aproximado
Transposições de Blocos	NP-Difícil	Algoritmo 2-aproximado
Transposições de Prefixo	?	Algoritmo 2-aproximado
Reversões com Sinal e Transposições	?	Algoritmo 2-aproximado
Reversões com Sinal, Transposições e Transreversões do Tipo A	?	Algoritmo 2-aproximado
Reversões de Prefixo e Transposições de Prefixo	?	Algoritmo 2-aproximado
Reversões e Transposições	?	Algoritmo $2k$ -aproximado



# Problema da Ordenação por Transposições

- Melhores algoritmos baseiam-se no grafo de ciclos.
- Proposta de ferramentas alternativas ao grafo de ciclos.
- Algoritmos que não baseiam-se no grafo de ciclos:
  - 2.25-aproximação proposta por Walter, Dias e Meidanis<sup>1</sup>.
  - 3-aproximação proposta por Benoît-Gagné e Hamel<sup>2</sup>.
  - Heurística proposta por Guyer, Heath e Vergara<sup>3</sup>.

---

<sup>1</sup>M. E. M. T. Walter, Z. Dias, and J. Meidanis. A new approach for approximating the transposition distance. In *Proceedings of the SPIRE'2000*, pages 199–208, Washington, DC, USA, 2000. IEEE Computer Society.

<sup>2</sup>M. Benoît-Gagné and S. Hamel. A new and faster method of sorting by transpositions. In *Proceedings of the CPM'2007*, volume 4580 of LNCS, pages 131–141, London, Ontario, Canada, 2007. Springer-Verlag.

<sup>3</sup>S. A. Guyer, L. S. Heath, and J. P. C. Vergara. *Subsequence and run heuristics for sorting by transpositions*. Technical Report TR-97-20, Virginia Polytechnic Institute & State University, 1997.

# Resultados

- Resultados obtidos no mestrado:
  - Desmonstração de 3-aproximação para a heurística de Guyer, Heath e Vergara.
  - Dados experimentais para permutações pequenas.
- Novos resultados:
  - Melhoria da complexidade do algoritmo de Benoît-Gagné e Hamel de  $O(n^2)$  para  $O(n \log n)$ .
  - Dados experimentais para permutações grandes.
  - Comparação com os melhores algoritmos conhecidos.
- Submissão de artigo para o J.UCS.

# Desenvolvimento de Heurísticas

- Ferramenta de auditoria para algoritmos de rearranjo de genomas.
- Base de dados de distâncias de rearranjo.
- Utilização das informações da base de dados para desenvolver heurísticas.

# Resultados

- Desenvolvimento de uma heurística de melhoria.
- Testes com 23 algoritmos aproximados ou heurísticos.
  - Reversão (2).
  - Transposição (5).
  - Reversão e Transposição (3).
  - Reversão de Prefixo (2).
  - Transposição de Prefixo (2).
  - Reversão de Prefixo e Sufixo (2).
  - Transposição de Prefixo e Sufixo (2).
  - Reversão e Transposição de Prefixo (3).
  - Reversão e Transposição de Prefixo e Sufixo (2).
- Artigo aceito para publicação no JBCB.

# Desenvolvimento de Algoritmos Aproximados

- Lacuna a ser explorada:
  - Fatores de aproximação relativamente altos.
- Variações envolvidas neste objetivo:
  - Problema da Ordenação por Reversões Curtas.
  - Problema da Ordenação por Blocos.
  - Problema da Ordenação por Reversões e Transposições.

# Problema da Ordenação por Reversões Curtas

- Uma reversão  $\rho(i, j)$  é dita uma  $k$ -reversão se  $j - i + 1 \leq k$ .
  - Uma reversão curta é uma  $k$ -reversão tal que  $k = 3$ .
- Variação introduzida por Heath e Vergara<sup>4</sup>:
  - Propuseram uma 2-aproximação.
  - Apresentaram limitantes para o diâmetro.
- Feng, Meng e Sudborough<sup>5</sup> e Feng, Sudborough e Lu<sup>6</sup> melhoraram limitante superior do diâmetro.
- O problema de ordenar uma permutação apenas com 2-reversões pode ser resolvido em tempo polinomial<sup>7</sup>.

---

<sup>4</sup>L. S. Heath and J. P. C. Vergara. *Sorting by short swaps*. *Journal of Computational Biology*, 10(5):775–789, 2003.

<sup>5</sup>X. Feng, Z. Meng, and I. H. Sudborough. *Improved upper bound for sorting by short swaps*. In *ISPAN'2004*, pages 98–103, Hong Kong, China, 2004. IEEE Computer Society.

<sup>6</sup>X. Feng, I. H. Sudborough, and E. Lu. *A fast algorithm for sorting by short swap*. In *CASB'2006*, pages 62–67, Dallas, Texas, USA, 2006. ACTA Press.

<sup>7</sup>M.R Jerrum. *The complexity of finding minimum-length generator sequences*. *Theoretical Computer Science*, 36:265–289, 1985.

# Problema da Ordenação por Blocos

- Bloco: sequência crescente maximal de inteiros consecutivos.
  - Uma transposição de bloco é uma transposição  $\rho(i, j, k)$  tal que  $\pi_i \pi_{i+1} \dots \pi_{j-1}$  ou  $\pi_j \pi_{j+1} \dots \pi_{k-1}$  é um bloco.
- Variação introduzida por Latifi<sup>8</sup>.
- Demonstrada ser NP-Difícil por Bein *et al.*<sup>9</sup>
- Melhores algoritmos aproximados:
  - 2-aproximação proposta por Mahajan *et al.*<sup>10</sup>
  - 2-aproximação proposta por Bein *et al.*<sup>11</sup>

---

<sup>8</sup> S. Latifi. *How can permutations be used in the evaluation of zoning algorithms?* *International Journal of Pattern Recognition and Artificial Intelligence*, 10(3):223–237, 1996.

<sup>9</sup> W. W. Bein, L. L. Larmore, S. Latifi, and I. H. Sudborough. *Block sorting is hard.* *International Journal of Foundations of Computer Science*, 14(3):425–437, 2003.

<sup>10</sup> M. Mahajan, R. Rama, V. Raman, and S. Vijaykumar. *Approximate block sorting.* *International Journal of Foundations of Computer Science*, 17(2):337–356, 2006.

<sup>11</sup> W. W. Bein, L. L. Larmore, L. Morales, and I. H. Sudborough. *A quadratic time 2-approximation algorithm for block sorting.* *Theoretical Computer Science*, 410(8-10):711–717, 2009.

# Problema da Ordenação por Reversões e Transposições

- Variação introduzida por Walter, Dias e Meidanis<sup>12</sup>:
  - Apresentaram uma 3-aproximação.
- Rahman, Shatabda e Hasan<sup>13</sup> desenvolveram uma  $2k$ -aproximação.

---

<sup>12</sup>M. E. M. T. Walter, Z. Dias, and J. Meidanis. *Reversal and transposition distance of linear chromosomes*. In *SPIRE'1998*, pages 96–102, Santa Cruz, Bolivia, 1998. IEEE Computer Society.

<sup>13</sup>A. Rahman, S. Shatabda, and M. Hasan. *An approximation algorithm for sorting by reversals and transpositions*. *Journal of Discrete Algorithms*, 6(3):449–457, 2008.



# Atividades

- Grupo 1: atividades relacionadas ao cumprimento dos requisitos para a obtenção do grau de doutor.
- Grupo 2: atividades relacionadas ao Objetivo 1.
- Grupo 3: atividades relacionadas ao Objetivo 2.
- Grupo 4: atividades relacionadas ao Objetivo 3.

# Cronograma das Atividades do Grupo 1

2012	2013		2014		2015		2016
ago-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jul
a	a		b				c

- a) Obtenção dos créditos obrigatórios (incluindo os necessários para dispensa do Exame de Qualificação Geral) e realização de estágio docente.
- b) Defesa do Exame de Qualificação Específico de Doutorado.
- c) Escrita e defesa da tese.

## Cronograma das Atividades do Grupo 2

2012	2013		2014		2015		2016
ago-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jul
a	b		c				

- a) Obtenção de novos resultados teóricos relativos aos algoritmos alternativos para o Problema da Ordenação por Transposições.
- b) Comparação experimental entre algoritmos alternativos e os melhores algoritmos conhecidos, que se baseiam no grafo de ciclos.
- c) Publicação de artigo em periódico.

## Cronograma das Atividades do Grupo 3

<b>2012</b>	<b>2013</b>		<b>2014</b>		<b>2015</b>		<b>2016</b>
ago-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jul
	a	a,b	c				

- a) Desenvolvimento de heurísticas.
- b) Avaliação experimental das heurísticas desenvolvidas.
- c) Publicação de artigo em periódico.

## Cronograma das Atividades do Grupo 4

2012	2013		2014		2015		2016
ago-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jun	jul-dez	jan-jul
			a	a,b,d	b,c,d	c,d	d

- a) Desenvolvimento de algoritmos aproximados e heurísticas para o Problema da Ordenação por Reversões Curtas.
- b) Desenvolvimento de algoritmos aproximados e heurísticas para o Problema da Ordenação por Blocos.
- c) Desenvolvimento de algoritmos aproximados e heurísticas para o Problema da Ordenação por Reversões e Transposições.
- d) Publicação de artigos em conferências e periódicos.