

Uma Ferramenta de Auditoria para Algoritmos de Rearranjo de Genomas

Aluno: Gustavo Rodrigues Galvão

Orientador: Zanoni Dias

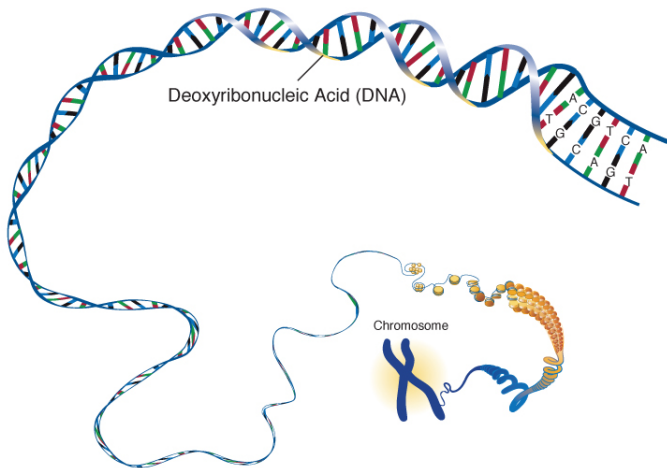
Instituto de Computação, UNICAMP

20 de Dezembro de 2012

Agenda

- 1 Introdução
- 2 Construção da Ferramenta
- 3 Aplicação da Ferramenta
- 4 Considerações Finais

DNA, Gene, Cromossomo e Genoma



Fonte: http://www.genome.gov/glossary/resources/dna_deoxyribonucleic_acid_lg.jpg

Evolução de Genomas

- Imprecisão no processo de replicação do DNA.
 - Mutações locais: remoções, substituições e inserções.
 - Mutações globais: eventos de rearranjo.
- Eventos de rearranjo conservativos.
 - Reversão.
 - Transposição.
- Eventos de rearranjo não conservativos.
 - Remoção.
 - Duplicação.

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 igual ao tamanho da menor sequência 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.
 - $\pi = (\pi_1 \pi_2 \dots \pi_n)$, $1 \leq |\pi_i| \leq n$ e $|\pi_i| \neq |\pi_j| \leftrightarrow i \neq j$.
- Cada elemento π_i possui um sinal, + ou -, que indica a orientação do gene que ele representa.
 - O sinal é omitido caso não exista 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 $t(i, j, k)$, $1 \leq i < j < k \leq n + 1$, aplicada na permutação (com ou sem sinal)

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

a transforma na permutação (com ou sem sinal)

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

Distância de Rearranjo Entre Permutações

Modelo de Rearranjo

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

Distância de Rearranjo

A distância de rearranjo entre duas permutações π e σ é igual ao tamanho da menor sequência de eventos de rearranjo pertencentes a um determinado modelo de rearranjo que transforma π em σ (ou σ em π).

Problema da Ordenação

- Permutação identidade: $\iota = (1\ 2\ \dots\ n)$.
- Ordenar uma permutação significa transformá-la em ι .
 - Problema da Ordenação: encontrar menor sequência de eventos de rearranjo que ordena uma permutação.
- Transformar a permutação π na permutação σ é equivalente a transformar a permutação $\sigma^{-1} \circ \pi$ em ι .
 - Distância de rearranjo de uma permutação.
 - Redução ao Problema da Ordenação.
- Algoritmos de Rearranjo de Genomas.

Estado da Arte de Algumas Variações

Modelo de Rearranjo	NP-Difícil	Melhor Solução Teórica
Reversão com Sinal	Não	Algoritmo exato $O(n^{\frac{3}{2}} \sqrt{\log n})$
Reversão	Sim	Algoritmo 1.375-aproximado
Transposição	Sim	Algoritmo 1.375-aproximado
Reversão de Prefixo	Sim	Algoritmo 2-aproximado
Transposição de Prefixo	?	Algoritmo 2-aproximado
Reversão Curta	?	Algoritmo 2-aproximado
Reversão com Sinal e Transposição	?	Algoritmo 2-aproximado

Uma Ferramenta para Avaliação de Algoritmos

- Dificuldade do Problema da Ordenação.
 - Algoritmos aproximados.
 - Heurísticas.
- Construção de uma ferramenta para avaliação de algoritmos.
 - Apoio à comparação.
 - Apoio à validação.
- Aplicação da ferramenta.
 - Avaliação da justeza do fator de aproximação.

Panorama

- Auditoria de um algoritmo de rearranjo de genomas.
- Etapas da construção da ferramenta.
 - Definição do algoritmo para calcular distâncias de rearranjo.
 - Cálculo das distâncias de rearranjo.
 - Desenvolvimento da ferramenta (GRAAu).

Visão Geral

- Algoritmo de busca em largura.
- Diminuição do uso de memória.
 - Representação de permutações como inteiros.
- Diminuição do tempo de execução.
 - Paralelização.

Implementação

- Algoritmo implementado em C.
 - Utilização da biblioteca *pthread*s.
- Implementação de duas versões.
 - Versão de 32 bits.
 - Versão de 64 bits.
- Eficiência no uso de memória.
 - 12x melhor do que o algoritmo de Dias e Meidanis¹.
 - 7x melhor do que o algoritmo de Walter, Dias e Meidanis².
- Ganho de desempenho com a paralelização.

¹Z. Dias and J. Meidanis. *Sorting by prefix transpositions*. In *Proceedings of the SPIRE'2002*, volume 2476 of LNCS, pages 65–76, Lisbon, Portugal, 2002. Springer-Verlag.

²M. 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.

Visão Geral

- Foram considerados 11 modelos de rearranjo.
 - Permutação sem sinal $\rightarrow n \leq 13$
 - Permutação com sinal $\rightarrow n \leq 10$
- Distribuição das distâncias de rearranjo.
 - Validação de conjecturas a respeito do diâmetro.
 - Proposição de duas novas medidas.
- Base de dados de distâncias de rearranjo.
 - Escrutínio público.
 - Extração de informações.

Visão Geral

- Implementação em Java.
 - Arquitetura cliente-servidor.
 - Utilização de *web services*.
- Estatísticas produzidas.
 - Diâmetro.
 - Distância média.
 - Razão média.
 - Razão máxima.
 - Igualdade.
- Permutações que apresentaram a razão máxima.

Visão Geral

- 16 algoritmos aproximados implementados.

Modelo de Rearranjo	Qtde	Melhor Aprox.
Reversão	2	Não
Transposição	3	Não
Reversão de Prefixo	3	Sim
Reversão de Prefixo com Sinal	3	Sim
Reversão Curta	3	Sim
Transposição de Prefixo	2	Sim

- 12 algoritmos propostos na literatura e 4 propostos por nós.

Contribuições

- Revisão de alguns resultados da literatura.
 - Resultados experimentais de Walter, Dias e Meidanis³.
 - Algoritmo 3-aproximado de Benoît-Gagné e Hamel⁴.
- Demonstração de fatores de aproximação.
 - Fator 2 para o algoritmo de Cohen e Blum⁵.
 - Fator 3 para heurística de Guyer, Heath e Vergara⁶.

³ 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.

⁴ 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.

⁵ D. S. Cohen and M. Blum. On the problem of sorting burnt pancakes. *Discrete Applied Mathematics*, 61(2):105–120, 1995.

⁶ 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.

Visão Geral

- Algoritmos implementados em Java e auditados com o GRAAU.
- Aplicações para as estatísticas produzidas.
 - Comparação.
 - Avaliação da justeza do fator de aproximação.

Contribuições

- Justeza do fator de aproximação.
 - Mostramos que o fator de aproximação de 7 algoritmos é justo.
 - Conjecturamos que o fator de outros 6 é justo.
- Contraposição de duas hipóteses levantadas na literatura.
 - Hipótese de Benoît-Gagné e Hamel⁷.
 - Hipótese de Fischer e Ginzinger⁸.

⁷ 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.

⁸ J. Fischer and S. W. Ginzinger. A 2-approximation algorithm for sorting by prefix reversals. In *Proceedings of the ESA'2005*, volume 3669 of LNCS, pages 415–425, Mallorca, Spain, 2005. Springer-Verlag.

Conclusão

- Sumário.
 - Construção da ferramenta.
 - Aplicação da ferramenta.
- Críticas.
- Expectativas.

Publicações

- Foram 8 publicações ao todo.
 - 1 resumo estendido no ACM-SAC 2011.
 - 1 resumo estendido e 2 artigos completos no BSB 2011.
 - 2 artigos completos no BICoB 2012.
 - 1 resumo estendido e 1 artigo completo no BSB 2012.
- Artigos para periódicos.

Trabalhos Futuros

- Manutenção do GRAAu.
- Estudar mais profundamente as medidas que propusemos.
- Validar conjecturas a respeito da justeza do fator de aproximação dos algoritmos.
- Desenvolver algoritmos aproximados melhores.