

Plano de Estudos

Um Estudo Sobre Ordenação Por Reversão

Aluna: Andréa Tieme Nakasato

Orientadora: Profa. Dra. Yoshiko Wakabayashi

Área de Concentração: Otimização Combinatória

— São Paulo, 25 de abril de 2005—

UM ESTUDO SOBRE ORDENAÇÃO POR REVERSÃO

Resumo

O problema de ordenação por reversão (*sorting by reversals*) é de interesse em Biologia Computacional para o tratamento de problemas relativos a rearranjo de genomas. Dados dois genomas representados como permutações de mesmos elementos (isto é, genes), o problema consiste em encontrar a seqüência mais parsimoniosa de reversões que transforma um genoma em outro. Depois que o primeiro algoritmo polinomial para o problema foi apresentado [9], muitos trabalhos foram feitos a fim de melhorar e simplificar o algoritmo [5, 1, 2]. O nosso objetivo é estudar os artigos que tratam desse problema e fazer uma resenha atualizada sobre o tema.

1 Introdução

No fim dos anos 80, Jeffrey Palmer e seus colegas compararam os genomas mitocondriais de *Brassica oleracea* (repolho) e de *Brassica campestris* (nabo) e descobriram que os genomas dessas espécies têm conjunto de genes idênticos, mas estes diferem na ordem em que ocorrem. Este e muitos outros estudos provaram que esta característica de rearranjo de genomas é uma forma comum de evolução molecular [4, 8, 9].

A análise de rearranjo de genomas permite comparar dados moleculares de espécies que divergem há muito tempo. Entretanto, muito pouco se sabe sobre rearranjos que produziram as variedades existentes das arquiteturas genômicas. A razão para encontrar uma série mínima de eventos de rearranjo é justificada pela *hipótese da parsimônia*, na qual supõe-se que a Natureza sempre encontra caminhos que realizam de um mínimo de mudanças. Portanto, se desejarmos investigar como um organismo de uma espécie transformou-se em um organismo de outra espécie (ou vice-versa), devemos tentar encontrar uma série mínima de eventos de rearranjo que possivelmente tenham realizado esta transformação.

2 Reversão

Rearranjos de genomas, através de reversões, podem alterar a ordem dos genes em um genoma, e também a direção de transcrição. Identificamos cada gene com um número

inteiro em $\{1, \dots, n\}$, acrescido de um sinal de mais (+) ou menos (-) para indicar sua direção. A ordem será representada por uma permutação π de $\{\pm 1, \dots, \pm n\}$ tal que $\pi[-i] = -\pi[i]$, onde $\pi[i]$ denota o i -ésimo elemento em π . A partir daqui, indicaremos o sinal de um elemento em uma permutação apenas quando for menos.

A *reversão* do intervalo $[i, j] \subseteq [1, n]$ ($i < j$) é a permutação

$$\rho = 1, \dots, i, -j, \dots, -(i+1), j+1, \dots, n.$$

Note que $\pi\rho$ é a permutação obtida de π revertendo a ordem e invertendo os sinais dos elementos no intervalo. Se ρ_1, \dots, ρ_k é uma seqüência de reversões, dizemos que ela ordena uma permutação π se $\pi\rho_1 \dots \rho_k = Id$ (onde Id é a permutação *identidade* $1, \dots, n$). O comprimento da menor seqüência de reversões que ordena π é chamado de *distância genômica* ou *distância de reversão* de π e é denotado por $d(\pi)$.

O *problema de ordenação por reversão* é encontrar a menor seqüência de reversões que ordena π . Esse problema tem sido bastante investigado, sendo vasta a literatura a respeito [7, 6, 2]. Hannenhall e Pevzner [10] apresentaram o primeiro algoritmo polinomial para esse problema que consome tempo $O(n^4)$. A partir daí muitos melhoramentos foram feitos. Um algoritmo mais rápido que consome tempo $O(n^2)$ é o de Kaplan, Shamir e Tarjan [5]. Bader, Moret e Yan [1] descreveram um algoritmo linear para computar $d(\pi)$ apenas, mas este algoritmo não encontra a seqüência de reversões.

3 Aplicações

A grande motivação no estudo desse problema é a análise de rearranjos de genomas em cromossomos, organelas de plantas, vírus, etc. Tal análise possibilita descobrir relações entre os genomas de duas espécies diferentes e, com isso visualizar a evolução entre elas.

4 Atividades a serem desenvolvidas e cronograma

Inicialmente estudaremos os artigos [6, 2, 3]. À medida que os estudos forem avançando, pretendemos investigar outros artigos, procurando fazer um trabalho bem abrangente e atualizado sobre o tema.

O cronograma que prevemos para 3 meses é o seguinte.

ATIVIDADE	1o. mês.	2o. mês.	3o. mês.
Pesquisa de textos/artigos rel. proj.	✓	✓	✓
Leitura de textos/artigos rel. proj.	✓	✓	✓
Escrita da Monografia		✓	✓

5 Metodologia de estudo e análise dos resultados

Os estudos serão conduzidos individualmente pelo aluno, sob a supervisão da orientadora. O acompanhamento dos estudos será feito através de reuniões regulares, nas quais o aluno deverá expor os tópicos considerados mais complexos, ou os avanços conseguidos em suas investigações.

Referências

- [1] Moret B.M.E. Bader D. A. and Yan M. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. pages 365–376. Proceedings of the 7th Workshop on Algorithms and Data Structures, 2001.
- [2] Tannier E., Bergeron A., and Sagot M-F. Advances on sorting by reversals. Technical report, Jan 2005.
- [3] Tannier E. and Sagot M-F. Sorting by reversals in subquadratic time. volume 3109, pages 1–13. Proceedings of the 15th Annual Symposium on Combinatorial Pattern Matching (CPM), Jul 2004.
- [4] Nadia El-Mabrouk. *Genome Rearrangement by Reversals and Insertions/Deletions of Contiguous Segments*, volume 1848, page 222. Lecture Notes in Computer Science, Jan 2000.
- [5] Shamir R. Kaplan H. and Tarjan R. E. Faster and simpler algorithm for sorting signed permutations by reversals. *SIAM Journal on Computing*, 29:880–892, 1999.
- [6] Sagot M-F. and Tannier E. Perfect sorting by reversals. Technical report, Feb 2005.

- [7] Berman P. and Hannenhalli S. Fast sorting by reversals. In *Lecture Notes in Computer Science (LNCS)*, number 1075, pages 168–185. In *Combinatorial Pattern Matching, Proceedings of the 7th Annual Symposium (CPM'96)*, 1996.
- [8] Pevzner P. and Tesler G. Genome rearrangements in mammalian evolution: Lessons from human and mouse genomes. *Genome Research*, 13:37–45, Jan 2003.
- [9] Hannenhalli S. and Pevzner P. Towards a computational theory of genome rearrangements. *Lecture Notes in Computer Science*, 1000:184–202, 1995.
- [10] Hannenhalli S. and Pevzner P. Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). pages 178–189. *Proceedings of the 27th ACM Symposium on Theory of Computing*, 1995.