

Estimação da Distância de Reversão de Genomas Baseada em Técnicas de *Machine Learning*

Nathalia Menini, Sergio Arnosti e Jorge da Silva

Instituto de Computação da Unicamp,
Av. Albert Einstein, 1251 - Cidade Universitária, Campinas - SP
{nathmenini,serza.arnosti,jorge.inatel}@gmail.com

Resumo FAZER

Keywords: fazer.

1 Introdução

Comparar dois genomas é uma tarefa fundamental para se estudar as relações e a evolução entre os genes [5,2]. Nesse caminho, a Biologia Computacional tem se apresentado como uma ótima aliada para os pesquisadores da área, tornando possível descobrir relações entre os genes que, para a percepção do ser humano, poderia ser muito difícil. [2]

O Rearranjo de Genomas é uma mutação que ocorre nos genomas mitocondrias [1]. A ordem dos genes no genoma está em constante rearranjo. Desse modo, através da estimação da distância de rearranjo entre dois genes, a relação entre eles pode ser estimada [4].

Quando um fragmento do filamento de DNA é revertido na replica final, temos o que chamamos de *reversão*, que é uma das mutações mais comumente vista em genomas. Por outro lado, se dois fragmentos de DNA trocam de posições durante o processo de replicação (mas não sofrem reversão), temos a mutação chamada de *transposição* [2]. Neste trabalho, o nosso foco será apenas na mutação conhecida por reversão.

A área da Filogenia, que estuda a história evolutiva das relações entre espécies, depende fortemente do Princípio da Parsimônia. Basicamente, dado um conjunto de possíveis explicações para um fato, a explicação mais simples é a mais provável de estar certa. Como as mutações são relativamente raras de acontecer, quando os pesquisadores tentam construir uma árvore filogenética, eles tentam fazer com que as espécies tenham o menor número de antecessores possíveis, ou seja, é muito mais provável que duas espécies que possuem uma determinada característica tenham evoluído do mesmo ancestral comum que desenvolveu essa característica, em vez de acreditar que a característica evoluiu duas vezes, de espécies diferentes [2].

Considere a permutação identidade $\iota = (1, 2, \dots, n)$ de tamanho n como o genoma original. Dada qualquer permutação π de tamanho n , o problema consiste

em construir um modelo capaz de calcular a distância de reversão de ι e π , em que são permitidos apenas operações de reversão.

Uma outra possível representação de genomas é utilizando a *orientação* dos genes. Nesse caso, nós representamos o genoma com os símbolos $+$ ou $-$ para cada gene, por exemplo, $(+1, -2, \dots, -n)$. Na literatura, essa abordagem é conhecida como *permutações com orientação*. Se nós não temos informações sobre a orientação dos genes, os sinais são omitidos e a abordagem passa a ser chamada de *permutação sem orientação*. Neste trabalho, trabalharemos apenas com as permutações sem orientação.

Entretanto, determinar o número mínimo de reversões necessárias para um genoma se transformar no outro, no contexto de permutação sem orientação, não é um problema fácil de se resolver, uma vez que foi provado ser um problema NP-difícil [6]. Muitas abordagens já foram propostas para atacar esse problema como, por exemplo, Kececioglu and Sankoff [7], propuseram o algoritmo de 2-aproximação que remove *breakpoints* por reversão (o conceito de *breakpoint* será descrito posteriormente) e Bafna and Pevzner [8] criaram o algoritmo de 1.75-aproximação utilizando grafos de reversão. O atual estado da arte foi desenvolvido por Barman et al. [9], que construiu um algoritmo de 1.375-aproximação.

2 Revisão Bibliográfica

3 Implementação

4 Resultados e Avaliação

5 Conclusões

Referências

1. Russell, P. J.: iGenetics. Benjamin Cummings (2002)
2. da Silva, M., Oliveira, A., Dias, Z.: Machine Learning Applied to Sorting Permutations by Reversal and Transpositions. Technical report, Unicamp (2017)
3. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
4. Pevzner, P.: Computational Molecular Biology An Algorithmic Approach. MIT Press (2001)
5. Lou, X.W., Zhu, D.M.: Sorting Unsigned Permutations by Weighted Reversals, Transpositions, and Transversals. Journal of Computer Science and Technology (2009)
6. Caprara, A.: Sorting by Reversals is Difficult. Proceedings of the First International Conference on Computational Molecular Biology (1997)
7. Kececioglu, J., Sankoff, D.: Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. Algorithmica (1995)

8. Bafna, V., Pevzner, P.: Genome rearrangements and sorting by reversals. *SIAM Journal on Computing* (1996)
9. Berman, P., Hannenhalli, S., Karpinski, M.: 1.375-approximation algorithm for sorting by reversals. In *10th Annual European Symposium on Algorithms* (2002)