

Aluno: Glaucio Teixeira Souza – Instituto de Física – UFRGS

Orientadora: Rita Maria Cunha de Almeida – Instituto de Física – UFRGS

Colaboradores: José Luiz Rybarczyk Filho – Instituto de Física – UFRGS

Rodrigo J. S. Dalmolin – Depto de Bioquímica – UFRGS

Introdução

O genoma pode ser definido como toda a informação hereditária de um organismo e esta informação está codificada no DNA. Cada vez que uma célula se divide, todas as moléculas de DNA devem ser copiadas para as células resultantes, de modo que estas possuam uma cópia completa do genoma e assim, ter a capacidade de desenvolver e desempenhar suas funções específicas. É neste processo de cópia que podem ocorrer alterações nos nucleotídeos, as quais, ao longo do tempo, serão a base para a evolução do genoma.

Motivação

A evolução pode ser considerada um processo de ramificação, sendo uma árvore filogenética a representação geralmente utilizada para uma filogenia, isto é, para representar a história evolucionária inferida para um grupo de organismos. Construindo uma árvore filogenética para um grupo ortólogo, definido como um conjunto de genes de diferentes espécies existentes que possuam um gene ancestral comum, podemos inferir como se deu a evolução para um determinado gene do grupo.

O genoma ao evoluir geralmente adquire novos genes, e a forma como são adquiridos parece determinar muitas das características das redes genômicas. Essa aquisição se dá, basicamente, por duplicação de genes, de cromossomos ou de todo o genoma, e pela formação de novos genes.

Segundo Mirkin[1], a reconstrução do cenário evolucionário para um determinado grupo de genes ortólogos pode ser formulada como a busca pelo cenário evolutivo mais parcimonioso, isto é, o cenário com o menor número de eventos para um dado conjunto de ortólogos em uma árvore filogenética. Nesse contexto, eventos representam o surgimento ou perda de genes, que tenham ocorrido ao longo da evolução de tal forma que resultem no padrão filético observado atualmente, isto é, o padrão de presença e/ou ausência do gene para as espécies atuais.

Considerações biológicas sugerem que é mais provável que ocorra a perda de genes existentes do que o surgimento de novos genes. Tendo isto em mente, para construir um cenário evolutivo mais realista percebe-se a necessidade de diferenciar os eventos de aquisição e perda de genes, introduzindo um sistema de custo evolutivo que penaliza o surgimento de novos genes. Assim, o cenário mais parcimonioso é o que minimiza o custo total, $S = \lambda + \gamma$, onde λ é o número de perdas e γ é o número de surgimentos e g é a penalidade de surgimento. O valor mínimo para esse custo é denominado valor de inconsistência para um dado gene. Os custos relativos aos eventos evolucionários consideram duas unidades de custo para surgimento do gene ($g=2$), e uma unidade de custo para perda [2,3].

Objetivos

Este trabalho busca sistematizar a determinação da raiz ancestral, o gene mais ancestral, para um gene individual presente em um dado grupo de genes ortólogos, a partir da análise da árvore filogenética de consenso utilizando o método da análise de parcimônia.

Metodologia e desenvolvimento

A metodologia de trabalho consistiu na pesquisa bibliográfica sobre o estado da arte dos modelos e na proposição de um algoritmo para análise de parcimônia de árvores filogenéticas. O algoritmo desenvolvido para a inferência da raiz ancestral, o último ancestral comum que possui o gene estudado, busca o cenário evolutivo mais parcimonioso ao mapear todas as configurações possíveis de surgimento e perda do gene de modo a encontrar a configuração que minimize a seguinte equação:

$$S = V_{raiz} + \sum_{i=0}^N \gamma V_i (V_{i+1} - V_i)^2 + \lambda V_{i+1} (V_{i+1} - V_i)^2$$

Onde V_i representa o valor de presença (1) ou ausência (0) do gene para o organismo em questão (nó da árvore), γ é o custo do surgimento do gene ($\gamma=2$), λ é o custo de perda ($\lambda=1$) e N é o número de nós representados na árvore filogenética, isto é, o número total de organismos desde os mais ancestrais até os atuais, contemplados na árvore filogenética de consenso.

Resultados

A partir do algoritmo desenvolvido, foi implementada uma ferramenta computacional desenvolvida na linguagem de programação Java. Esta ferramenta foi capaz de reproduzir os cenários evolutivos calculados manualmente e utilizados no artigo “Evolutionary origins of human apoptosis and genomestability gene networks”[5]. Na figura ao lado, em vermelho, a localização do gene ancestral do gene MSH5 do grupo de ortólogos 0221 de eucariotos. Este gene possui um valor de inconsistência igual a 9.

Essa ferramenta possibilitará uma rápida análise para uma grande quantidade de grupos de ortólogos individuais possibilitando a inferência da ancestralidade dos genes, propiciando a validação de medidas experimentais e fornecendo dados para a simulação do crescimento de genomas (realizada pelo Grupo de Modelos Teóricos e Computacionais do IF UFRGS) que possam ser utilizados para direcionar o desenvolvimento dos modelos do grupo.

Referências

- [1] Mirkin, B.G., Fenner, T.I., Galperin, M.Y. and Kooning, E.V. (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.*, 3.
- [2] Snel, B., Bork, P. and Huynen, M.A. (2002) Genomes in flux: the 110 evolution of archaeal and proteobacterial gene content. *Genome Res.*, 12, 17–25.
- [3] Kunin, V. and Ouzounis, C.A. (2003) The balance of driving forces during genome evolution in prokaryotes. *Genome Res.*, 13, 1589–1594.
- [4] Jensen, L.J. et al. STRING 8 - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416 (2009).
- [5] Castro, M.A.A. et al. (2008) Evolutionary origins of human apoptosis and genomestability gene networks. *Nucleic Acids Res.*, 36, 6269–6283.

