

Ferramentas para comparação de genomas

Giovani de Sousa Leite e Prof. Doutor Luciano Antonio Digiampietri

Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, SP, Brasil.

giovani.leite@usp.br

Objetivos

Muitas das comparações de genomas são baseadas em apenas um gene. Além dos problemas relacionados a anotações incorretas de genes, muitas vezes não é possível diferenciar espécies muito próximas utilizando esse método. Por isso, neste projeto foram especificadas e implementadas ferramentas para comparação de genomas utilizando um método mais robusto, baseado em todos os genes dos genomas em comparação, a fim de complementar as ferramentas já existentes.

Métodos e Procedimentos

A comparação entre os genomas desenvolvida é realizada em duas etapas. Na primeira são identificados os genes homólogos entre os genomas, sob as restrições estipuladas pelo usuário. O resultado dessa etapa é uma matriz na qual as colunas representam os genomas e as linhas as famílias de genes (agrupamentos dos homólogos). A partir dessa matriz, na segunda etapa é calculada a distância entre os genomas. A distância significa a contagem das diferenças da presença ou não de genes homólogos entre os genomas, ou seja, quantas vezes um gene de uma família está presente em um dos genomas e não está no outro. O resultado desta etapa é uma matriz quadrada em que cada célula contém a distância entre o genoma representado pela coluna e o genoma representado pela linha. Utilizando o pacote *ape* do R é possível montar a árvore filogenética, baseada nessa matriz de distâncias, apresentada na figura 1.

Resultados

As ferramentas foram testadas com quinze genomas de bactérias do gênero *Xanthomonas*, disponíveis no site do Centro Nacional de Informações Biotecnológicas (NCBI). Por considerar todos os genes de um grupo de genomas em comparação, o método aqui

proposto minimiza os problemas consequentes de erros na anotação, sendo mais estável quando comparado ao método que leva em conta apenas um gene.

Além disso, com essa estratégia foi possível diferenciar cepas de espécies bastante próximas, como as três *Xanthomonas oryzae* pv. *oryzae*, diferentemente do resultado obtido a partir da ferramenta PhyML (análise baseada apenas no gene *16S rRNA-processing protein RimM*).

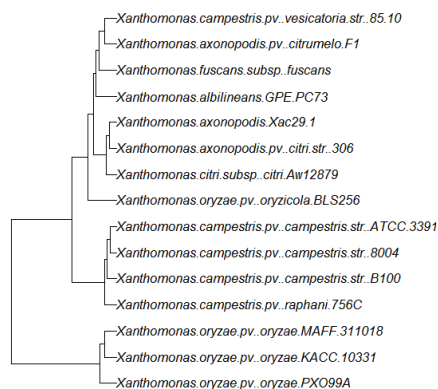


Figura 1: Árvore filogenética produzida

Conclusões

As ferramentas apresentadas nesse trabalho fornecem não só dados relativos a comparação de genomas utilizando de uma estratégia mais robusta, como dados úteis para a identificação da filogenia de novas espécies.

Referências

R. P. Ryan, F.-J. Vorholter, N. Potnis, J. B. Jones, M.-A. V. Sluys, A. J. Bogdanove, and J. M. Dow. Pathogenomics of xanthomonas: understanding bacterium-plant interactions. Nature Reviews Microbiology, 9(5):344–355, 2011.

Palavras-chave: Genomas; Algoritmos; Bioinformática.

Apoio: MEC/Sisu.