

Desenvolvimento de *Pipeline* para Investigação de Seleção Natural em Regiões Genômicas Codificantes.

Fernanda Trindade¹, Henrique Vieira Figueiró¹, Eduardo Eizirik¹.

¹Pontifícia Universidade Católica do Rio Grande do Sul – Campus Porto Alegre.
Av. Ipiranga, 6681 - Partenon - Porto Alegre/RS - CEP: 90619-900.

Resumo

A bioinformática surgiu como uma área de estudo que visa fornecer praticidade para a resolução de problemas biológicos e possibilidade de lidar com questões cuja escala inviabiliza resolução manual. Estas questões são extremamente variadas e envolvem diversos estudos em diferentes áreas do conhecimento. Um campo onde ferramentas de bioinformática são amplamente utilizadas é a comparação entre genomas, que se beneficia da elaboração de *scripts* e *pipelines* para tornar a análise mais rápida e eficiente, principalmente por se tratar de conjuntos de dados muito grandes. *Scripts* consistem de um conjunto de comandos cujo objetivo é automatizar uma ação, enquanto *pipeline* se refere a uma técnica que permite a execução integrada de múltiplos processos de maneira mais rápida e eficaz. Já existem disponíveis vários *softwares* necessários para estudos genômicos, porém a seleção dos mais adequados, a construção de bases de dados para cada um deles, bem como a definição e otimização deste processo são fatores ainda pouco explorados. O presente trabalho tem por objetivo desenvolver um *pipeline* para investigação de seleção natural afetando regiões genômicas codificantes, caracterizadas através da técnica de RNA-seq. Este foi desenvolvido inicialmente para análises evolutivas no transcriptoma de onça-pintada (*Panthera onca*). O *pipeline* foi estabelecido em ambiente shell do sistema Unix para o uso de três *softwares* – BLAST, MACSE e PAML – com a adição de *scripts* escritos em Perl para manipulação de arquivos intermediários. Primeiramente, para obter um conjunto de genes ortólogos, é executada uma busca de sequências codificantes em uma base de dados previamente estabelecida, homólogas àquelas da espécie alvo – etapa realizada pelo programa BLAST 2.2.25+. Posteriormente, um alinhamento múltiplo que leva em consideração o quadro de leitura é realizado para cada conjunto de ortólogos – etapa executada pelo programa MACSE. O último *software* usado é o YN00 do pacote PAML que, com base no conjunto de sequências alinhadas, calcula a razão entre substituições nucleotídicas sinônimas e não-sinônimas, que é uma abordagem consolidada para identificação de seleção natural. O *pipeline* já se encontra em uso e vem sendo rodado com diferentes combinações de parâmetros com o objetivo de definir aqueles adequados aos transcritos da onça-pintada. Apesar de já se encontrar plenamente funcional, modificações ainda devem ser feitas ao *pipeline* no futuro a fim de torná-lo mais ágil e prático. Os resultados obtidos até o momento indicam a eficiência deste para realização de análises nesse campo, e sugerem que a mesma poderá ser aplicada a outros organismos no futuro.

Palavras-chave

Bioinformática; Evolução molecular; Genômica; *Pipeline*.