DESENVOLVIMENTO DE FERRAMENTA AUTOMATIZADA PARA CLASSIFICAÇÃO DE ELEMENTOS TRANSPONÍVEIS

Caroline Heloíse de Oliveira – Universidade Estadual de Ponta Grossa – UEPG – carol42.helo@gmail.com
Adriano Ferrasa – Universidade Estadual de Ponta Grossa – UEPG – ferrasa@uepg.br

Resumo: Elementos Transponíveis (TEs) representam grande importância no estudo de doenças em plantas, microorganismos e mamíferos. Sua identificação e classificação apresenta desafios devido a sua grande diversidade. Neste sentido, algoritmos de Redes Neurais Artificiais (RNAs) podem ser utilizados na identificação de padrões em conjuntos de dados biológicos. Assim, busca-se a criação de uma ferramenta capaz de realizar a classificação automatizada e eficiente de TEs.

Palavras-chave: Bioinformática. Redes Neurais Artificiais. Elementos Repetitivos.

1. Introdução

Elementos Transponíveis (TEs) são regiões do genoma que podem se mover para outras localizações, estando presentes no genoma de eucariotos (Beck *et al.*, 2010). Sua ação é uma fonte de variabilidade genética, podendo estar relacionada ao desenvolvimento de doenças nestes organismos (Feschotte, 2008; Van De Lagemaat *et al.*, 2003). A classificação de TEs é de suma importância para ampliar o entendimento de sua dinâmica e também no desenvolvimento de doenças que possam afetar diferentes organismos. Contudo, o problema de classificação destes continua sendo um desafio devido à sua grande diversidade – TEs são classificados em 29 tipos que podem ser organizados de forma hierárquica em classes, subclasses, ordens e superfamílias (Wicker *et al.*, 2007). Dessa forma, pretende-se desenvolver uma ferramenta capaz de identificar e classificar TEs a nível de superfamília. Para alcançar este objetivo, será utilizado RNAs, que são algoritmos de aprendizado de máquina capazes de identificar padrões em conjuntos de dados, incluindo dados biológicos de diferentes organismos.

2. Materiais e Métodos

Para treinamento e testes de RNAs, os dados das sequências de TEs serão primeiramente selecionados e filtrados, removendo duplicatas, padronizando o comprimento das sequências e tratando dados faltantes ou incompletos. Após, uma arquitetura utilizando classificadores hierárquicos como nLLCPN (*non-Leaf Local Classifier per Parent Node*) e LCPNB (*Classifier per Parent Node and Branch*) serão utilizados. A avaliação de desempenho da classificação hierárquica de TEs será avaliada usando métricas padrão.

3. Resultados

A ferramenta de classificação hierárquica automatizada poderá ser utilizada para identificar e classificar a subfamília de um determinado TE de forma eficiente e com acurácia satisfatória, facilitando estudos de sua atividade genômica. Para estudos futuros, recomenda-se explorar a utilização de outros algoritmos de aprendizado de máquina, bem como outros modelos de representação de dados de treinamento.

4. Conclusão

O desenvolvimento deste projeto, fornecerá um recurso automatizado para a obtenção de dados que podem ser utilizados no estudo da evolução biológica de TEs e no entendimento de potenciais doenças que podem ser geradas por estes.

Referências

BECK, C. R. et al. LINE-1 retrotransposition activity in human genomes. **Cell**, v. 141, n. 7, p. 1159–1170, 2010.

FESCHOTTE, C. Transposable elements and the evolution of regulatory networks. **Nat Rev Genet**, v. 9, p. 397–405, 2008.

VAN DE LAGEMAAT, L. N. et al. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Trends Genet**, 2003.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nat Rev Genet** v. 8, p. 973–982, 2007.