

## **Algoritmo para construção de árvore filogenética em genomas completos de eucariontes utilizando a álgebra linear por vetores de frequência de tri-peptídios e entradas InterPro**

Lucas Felipe Silva (Autor), Matheus Allef Cruz (Co-Autor), Dalila Dominique Duarte Rocha (Co-Autor), Lara Maria Silva Miranda (Co-Autor), Thiago do Carmo Librelon Rocha (Co-Autor), Braulio Roberto Gonçalves Marinho Couto (Orientador), Marcos Augusto Dos Santos (Co-Orientador)

No estudo da evolução das espécies, é imprescindível a utilização de árvores filogenéticas para verificar a relação entre os mesmos. Entretanto, usando métodos filogenéticos tradicionais ocorrem erros, por exemplo, de alinhamentos incorretos ou número restrito de genes. Além disto, métodos de alinhamentos de sequências completas de genoma são impraticáveis, por possuírem enormes dados para processamento. Os métodos de álgebra linear abrem possibilidades em aplicações biométricas, utilizando a representação de proteínas como vetores no espaço multidimensional. O banco de dados InterPro integra modelos preditivos ou "assinaturas" de proteínas, que são usados para análises genéticas. A presente pesquisa busca responder à seguinte questão: Genomas analisados por métodos de álgebra linear, por representação vetorial a frequência de tri-peptídeos e entradas de InterPro podem gerar relações filogenéticas válidas do ponto de vista biológico? Utilizando genomas completos de 317 espécies de eucariontes, retiradas do banco de dados Uniprot, analisou-se a filogenia por meio de duas técnicas: Metodologia clássica (alinhamento par-a-par) testadas nos programas: MEGA, ClustalW, Clustal Omega, MUSCLE, BioEdit, CLC Sequence, e o uso de álgebra linear com vetores de frequência de tri-peptídeos (vetores de frequências de tri-peptídeos dos genomas) e entradas do InterPro (usando vetores de representação de genomas com a presença (1) ou ausência (0)), através do software MATLAB. O método clássico não gerou resultado válido em nenhum dos programas testados, devido a problemas computacionais pela grande quantidade de dados (genomas avaliados com mais de 13 milhões de aminoácidos). Com o uso da álgebra linear construiu-se árvores filogenéticas que apresentaram resultados similares através do dendograma. Os resultados obtidos demonstram que genomas podem ser avaliados por meio de técnicas de álgebra linear, esta premissa e de suma importância no tratamento de dados.

Instituição de Ensino: Centro Universitário de Belo Horizonte