

## UTILIZAÇÃO DE COEFICIENTES DE CLUSTERIZAÇÃO PARA ANÁLISE DE SEQUÊNCIAS DE DNA

Rogério Jose Panis Filho (voluntário), Günther J. L. Gerhardt - Deptº de Física e Química/Centro de Ciências Exatas e Tecnologia/UCS - [rjpanisf@ucs.br](mailto:rjpanisf@ucs.br)

O DNA, código da vida dos seres vivos, é formado por quatro tipos de bases (A, T, G, C), e devido a sua importância para a ciência, vem sendo estudado exaustivamente nos últimos anos. Mas para isso, se faz necessário a utilização de várias ferramentas. Neste trabalho propomos uma ferramenta alternativa para a análise da sequência do DNA, baseada em conceitos gráficos teóricos, investigando através de uma rede de tripletes (conjunto de três bases ao longo da sequência). Nesta rede, tripletes da sequência de DNA são vértices, sendo que dois vértices são conectados se eles ocorrem justapostos no genoma. Caracterizamos a topologia desta rede pela medida do coeficiente de clusterização. Testamos nossa metodologia por dois métodos: pelo conteúdo CG e pela periodicidade 3-bp (pares de bases) da sequência de DNA. Realizamos os testes construindo redes aleatórias, com conteúdo CG variável e uma periodicidade 3-bp imposta. Um grupo teste de organismos foi construído e investigamos a metodologia na luz de redes aleatórias construídas. Concluímos que o coeficiente de clusterização é uma ferramenta valiosa, nos fornecendo informação que não estão diretamente contidas na periodicidade 3-bp, nem na variação do conteúdo GC.

Palavras-chave: genoma, clusterização, periodicidade

Apoio: UCS