



CARACTERIZAÇÃO DE SEQUÊNCIAS PROMOTORAS RELACIONADAS AO SIGMA 28 PARA APRIMORAMENTO DA FERRAMENTA BACPP

Gabriel Dall Alba (BIC-UCS), Scheila de Avila e Silva (Orientador(a))

Sequências promotoras (promotores) são elementos regulatórios relacionados no processo de transcrição gênica, a qual ocorre através de uma interação entre a enzima RNA polimerase e uma sequência promotora. Estes elementos possuem propriedades estruturais (i.e. como estabilidade e curvatura) que podem servir como característica discriminante entre regiões promotoras e regiões codificantes. Devido à sua aplicabilidade no reconhecimento de padrões, técnicas de aprendizado de máquina (como Redes Neurais Artificiais (ANN)) tem sido constantemente empregadas na predição de promotores. Com base nisso, o presente trabalho tem como objetivo estabelecer o perfil que melhor caracteriza sequências promotoras procariontes relacionadas ao σ_{28} utilizando a estabilidade da dupla-fita de DNA como parâmetro. Para isso, foi desenvolvida uma análise inspirada na metodologia da ferramenta BacPP, uma ferramenta de predição de promotores baseada em ANNs e codificação de dados em valores de estabilidade. Ao todo, 231 sequências foram obtidas, com um tamanho padrão (80 nucleotídeos (nt)): (i) 144 como exemplos positivos (sequências promotoras relacionadas ao σ_{28}), extraídas do banco de dados RegulonDB e (ii) 87 como exemplos negativos (sequências não promotoras), geradas aleatoriamente de acordo com metodologia descrita na literatura. Todas as sequências foram codificadas em valores de estabilidade (V_s) conforme metodologia previamente estabelecida e comparadas com valores-padrão (V_r) obtidos por treinamento de uma ANN. A diferença (V_{Dif}) entre V_r e V_s está sendo analisada com o intuito de validar o ponto de corte que classifique corretamente sequências promotoras e não promotoras. Até o presente momento, os testes ainda estão em andamento. Com estes resultados, pretende-se aplicá-los no aprimoramento da ferramenta BacPP, auxiliando também na resolução de problemáticas relacionadas à predição de promotores, como a geração de elevados números de falsos positivos ou a ocorrência de baixos valores de sensibilidade.

Palavras-chave: Bioinformática, Promotores, Aprendizado de Máquina

Apoio: UCS