

XXIV Encontro de Jovens Pesquisadores VI Mostra Acadêmica de Inovação e Tecnologia



Implementação de valores de estabilidade do DNA como parâmetro de classificação da ferramenta BacPP

Bolsista: Gustavo Sganzerla Martinez (BIC-UCS)

Orientador: Prof. Dr. Daniel Luis Notari

Coorientadora: Prof. Dra. Scheila de Ávila e Silva

INTRODUÇÃO

Uma região promotora é uma sequência de DNA que antecede a região gênica e, ao ser reconhecida pela enzima RNA polimerase, desencadeia o processo de transcrição gênica. Os promotores, possuem uma sequência de nucleotídeos características e com um certo grau de conservação em diferentes bactérias. Adicionalmente, os promotores apresentam características físicas e estruturais que os diferem dos genes, como por exemplo, a estabilidade da dupla fita. A ferramenta Bacterial Promoter Prediction (BacPP) tem como função identificar e prever regiões promotoras de bactérias Gram - negativas utilizando as informações de nucleotídeos. No entanto, apesar da taxa de acerto do BacPP ser comparável a literatura, a quantidade de falsos positivos é elevado ao classificar promotores distantes filogeneticamente da bactérias Escherichia Coli..

OBJETIVOS

O objetivo deste trabalho é aprimorar a ferramenta BacPP inserindo os valores de estabilidade da dupla de DNA como um parâmetro de classificação, fazendo com que esta ferramenta identifique menos falsos positivos e tenha sua exatidão aumentada.

MATERIAL E MÉTODOS

1 Banco de dados e exemplos

Os exemplos de regiões promotoras foram obtidos através do banco de dados de domínio público Regulon DB. As informações estão disponíveis no site: <http://regulondb.ccg.unam.mx/index.html>.

2 Ferramentas

- Python foi a linguagem de programação selecionada para desenvolver o novo algoritmo de classificação.

3 Preparação de dados

As sequências analisadas possuem 81 nucleotídeos representados pelas letras A, T, C, G. Cada letra foi convertida em um valor de estabilidade extraído de uma rede neural. Os promotores foram tidos como os exemplos verdadeiros deste estudo e os exemplos falsos foram sequências aleatórias.

Os dados foram analisados através de cálculo de média móvel. Esta análise estatística consiste em analisar pontos nos dados, calculando uma série de média de subconjuntos diferentes do conjunto original. Optou-se por esta medida devido aos promotores apresentarem variação da posição do nucleotídeo devido a inserção ou deleção de pares ao longo do processo evolutivo

4 Desenvolvimento do algoritmo

4.1 Região discriminante

Ao analisar o resultado da análise de estabilidade dos promotores em uma rede neural que aprendeu e formou padrões sobre os dados foi encontrado uma região discriminante. Esta região apresentou elevações nos valores de estabilidade dos exemplos:

- Associados a produção de cílios e flagelos (grupo 1);
- Associados aos genes de fixação de nitrogênio (grupo 2);

Imagem 1 – Estabilidade de promotores associados a produção de cílios e flagelos

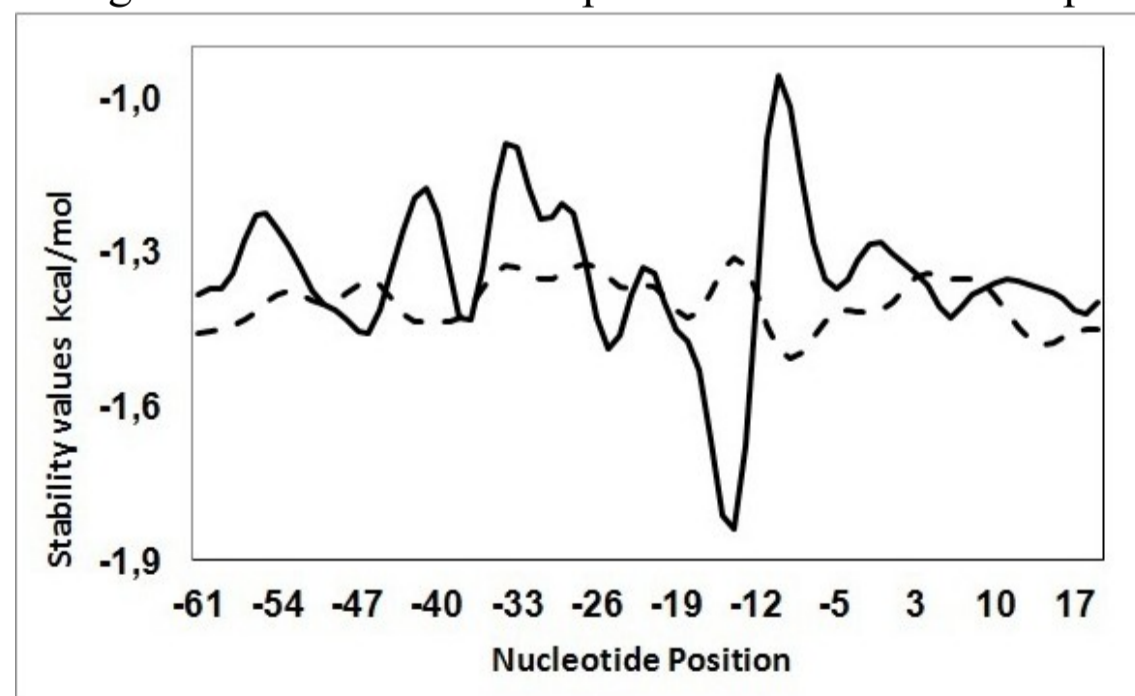
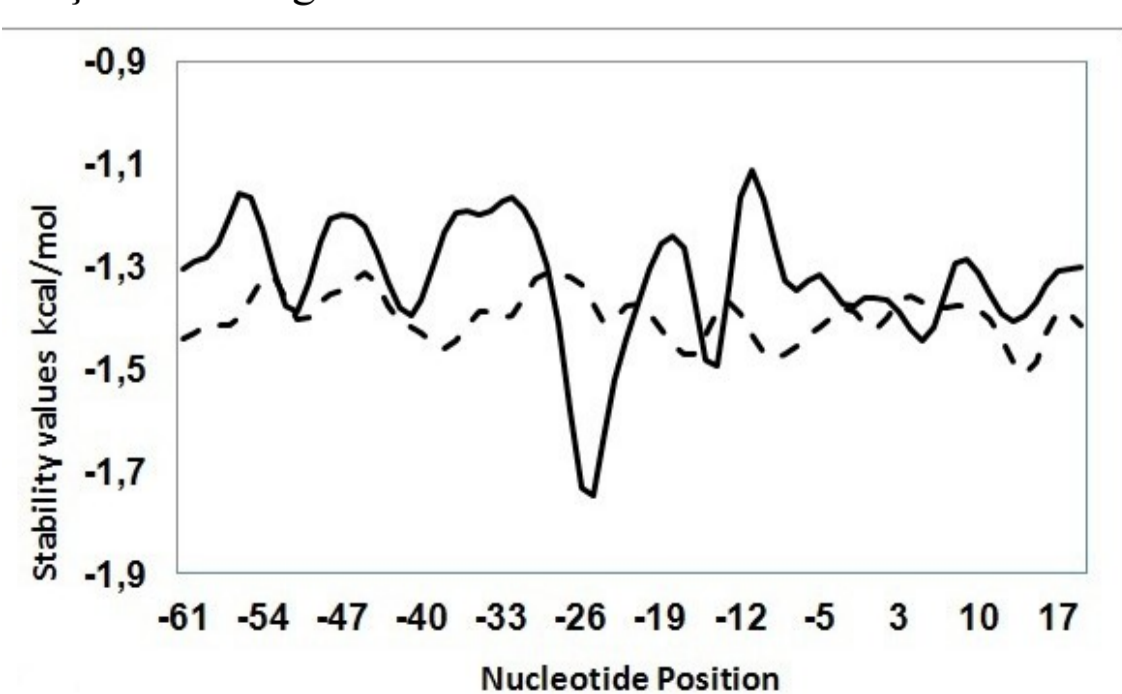


Imagem 2 – Estabilidade de promotores associados a genes de fixação de nitrogênio

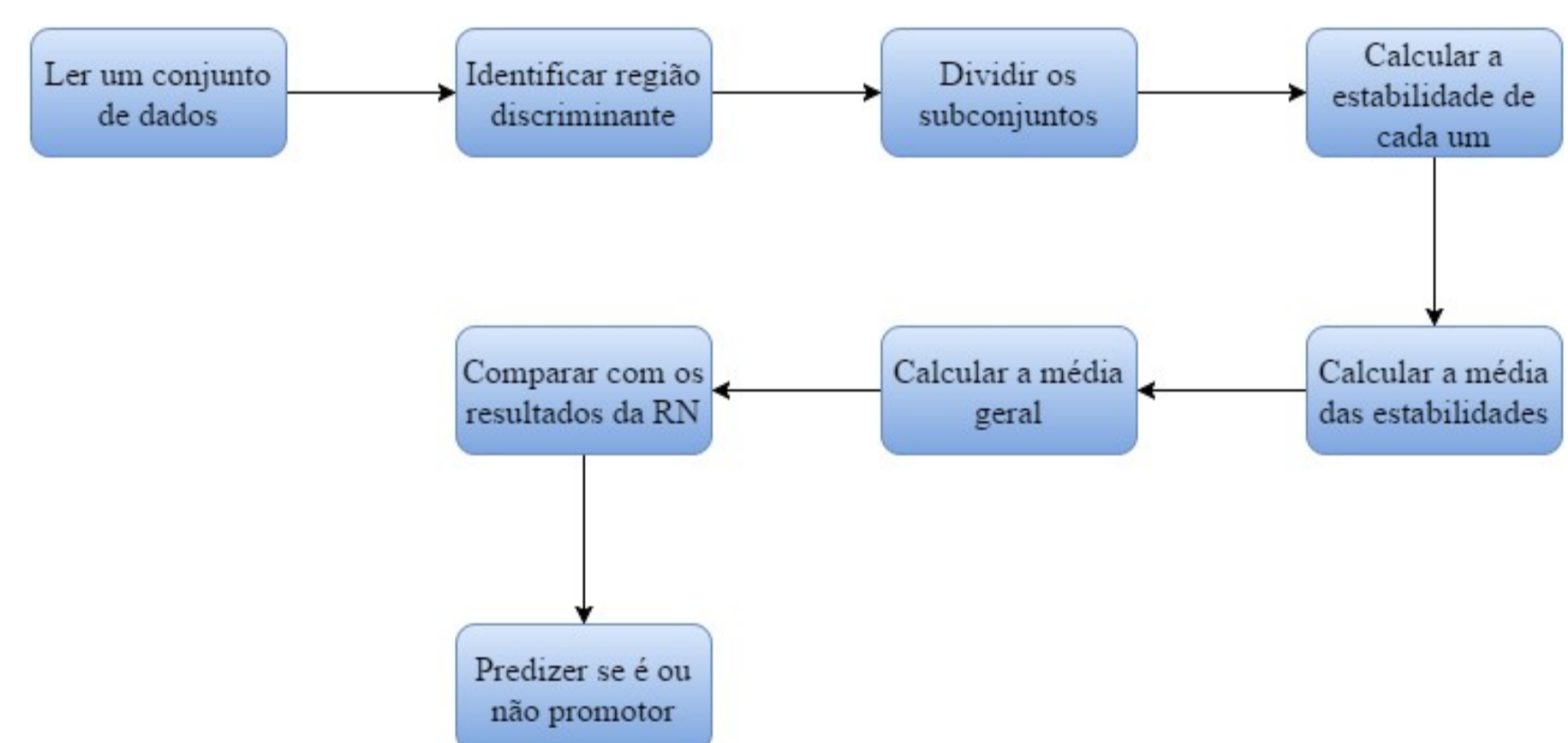


Esta região discriminante foi tomada como base para identificar promotores, onde aquele exemplo que apresenta a elevação nas posições -26 ou -14, trata-se de um promotor do grupo 1 ou grupo 2, respectivamente.

4.2 Algoritmo da Região Discriminante

Foi então desenvolvido um algoritmo baseado nas regras extraídas da RNA, onde para o grupo 2, os valores de estabilidade foram bastante acentuados na posição de nucleotídeo -25. Este algoritmo leva em conta os cálculos da média móvel. Os passos do algoritmo estão descritos pela imagem 3.

Imagem 3 – Diagrama de atividades do algoritmo da região discriminante



RESULTADOS E DISCUSSÃO

Ao analisar um conjunto de dados de 93 exemplos de promotores pertencentes ao grupo 2 na ferramenta BacPP e no novo algoritmo desenvolvido, foram encontrados os seguintes resultados mostrados pela tabela 1. Onde dos 93, apenas 19 (18%) foram reconhecidos pela versão atual da ferramenta. Ao submeter este mesmo conjunto de dados no algoritmo desenvolvido neste trabalho, a taxa de acerto foi de 54 (58%) dos promotores corretamente validados corretamente.

Tabela 1 – Comparativo entre as duas ferramentas

| | Taxa de Acerto (%) |
|-----------------------------------|--------------------|
| Algoritmo da região discriminante | 54 (58%) |
| BacPP | 19 (18%) |

Ao aplicar a medida de classificação com promotores do grupo 1, a ferramenta não teve um desempenho satisfatório, tendo uma taxa de acerto entre 15 e 25%. Estes valores são mais baixos do que a taxa de acerto atual da ferramenta e estudos futuros serão desenvolvidos buscando identificar um padrão para este grupo.

Por fim, ao realizar esta segunda abordagem classificatória, os resultados mostraram-se satisfatórios, podendo ser implementado como uma segunda análise tornando a ferramenta BacPP mais robusta ao classificar regiões promotoras deste grupo específico.

REFERÊNCIAS BIBLIOGRÁFICAS

de Avila e Silva, Scheila; FORTE, FRANCIELE ; T.S. SARTOR, IVAINE ; ANDRIGHETTI, TAHILA ; J.L. GERHARDT, GÜNTHER ; LONGARAY DELAMARE, ANA PAULA ; Echeverrigaray, Sergio ; ECHEVERRIGARAY, Sergio . DNA duplex stability as discriminative characteristic for Escherichia coli σ 54-and σ 28-dependent promoter sequences. Biologicals (London. Print), p. 22-28,2013 .

de Avila e Silva, Scheila; ECHEVERROGARAY, Sergio. Bacterial Promoter Features Description and Their Application on E. coli in silico Prediction and Recognition Approaches. In: Horacio Pérez-Sánchez. (Org.). Bioinformatics. 1ed. Rijeka: Intech, 2012, v. 1, p. 241-260.

Krebs, J. E.; Goldstein, E. S., Kilpatrick, S. T. Lewin's GENES XI, Jones & Bartlett Publ., 2014.

SONG, K. Recognition of prokaryotic promoters based on a novel variable-window Z-curve method, Nucleic Acids Research, vol. 40, n. 3, p. 963-971, 2012.