

PESQUISA MOVIMENTA INOVAÇÃO. INOVAÇÃO MOVIMENTA O FUTURO.

XXVIII ENCONTRO DE JOVENS PESQUISADORES E
X MOSTRA ACADÊMICA DE INOVAÇÃO E TECNOLOGIA

07 e 08 de OUTUBRO de 2020
UCS CAMPUS-SEDE - CAXIAS DO SUL



UCS
UNIVERSIDADE
DE CAXIAS DO SUL
PESSOAS EM
MOVIMENTO

BIC-UCS

Implementação e validação de algoritmo para predição de promotores de *E. coli* utilizando informações de estabilidade do DNA

Autores: Lucas Webber Molin, Scheila de Avila e Silva, Daniel Luis Notari

CBBL
COMPUTATIONAL BIOLOGY AND
BIOINFORMATICS LABORATORY

INTRODUÇÃO/OBJETIVO

A transcrição é a primeira etapa do processo de regulação da expressão gênica. Essa resulta na síntese de RNA (ácido ribonucleico), por meio de uma fita modelo de DNA (ácido desoxirribonucleico). Para isso, é necessário a ligação da enzima RNAP (RNA polimerase) a um elemento promotor presente anteriormente ao início de um gene. Assim, o controle a nível transcricional é fundamental para os organismos, visto que determina quais genes serão expressos ou reprimidos [1]. Em bactérias, esse sistema é regulado por subunidades protéicas denominadas fatores sigma (σ), responsáveis pela ligação da RNAP ao DNA promotor [2]. Características físico-químicas das regiões promotoras estão envolvidas com o reconhecimento dessas sequências de DNA. Desse modo, podem ser utilizadas como parâmetros para auxiliar na predição computacional desses elementos [3]. Nesse contexto, o objetivo do presente trabalho foi utilizar a característica física de estabilidade do DNA da bactéria *Escherichia coli* na implementação e validação de um algoritmo para predição de promotores regulados pelo fator σ_{28} , que posteriormente será utilizado como classificador de segunda etapa para o BacPP: uma ferramenta *web* para predição de promotores [4].

EXPERIMENTAL

A implementação do algoritmo foi baseada no trabalho de conclusão de curso do egresso Gabriel Dall'Alba do curso de Ciências Biológicas da Universidade de Caxias do Sul (UCS) [5], utilizando a linguagem de programação Python e o ambiente de desenvolvimento integrado (IDE) PyCharm. As etapas estão descritas no esquema da Figura 1:

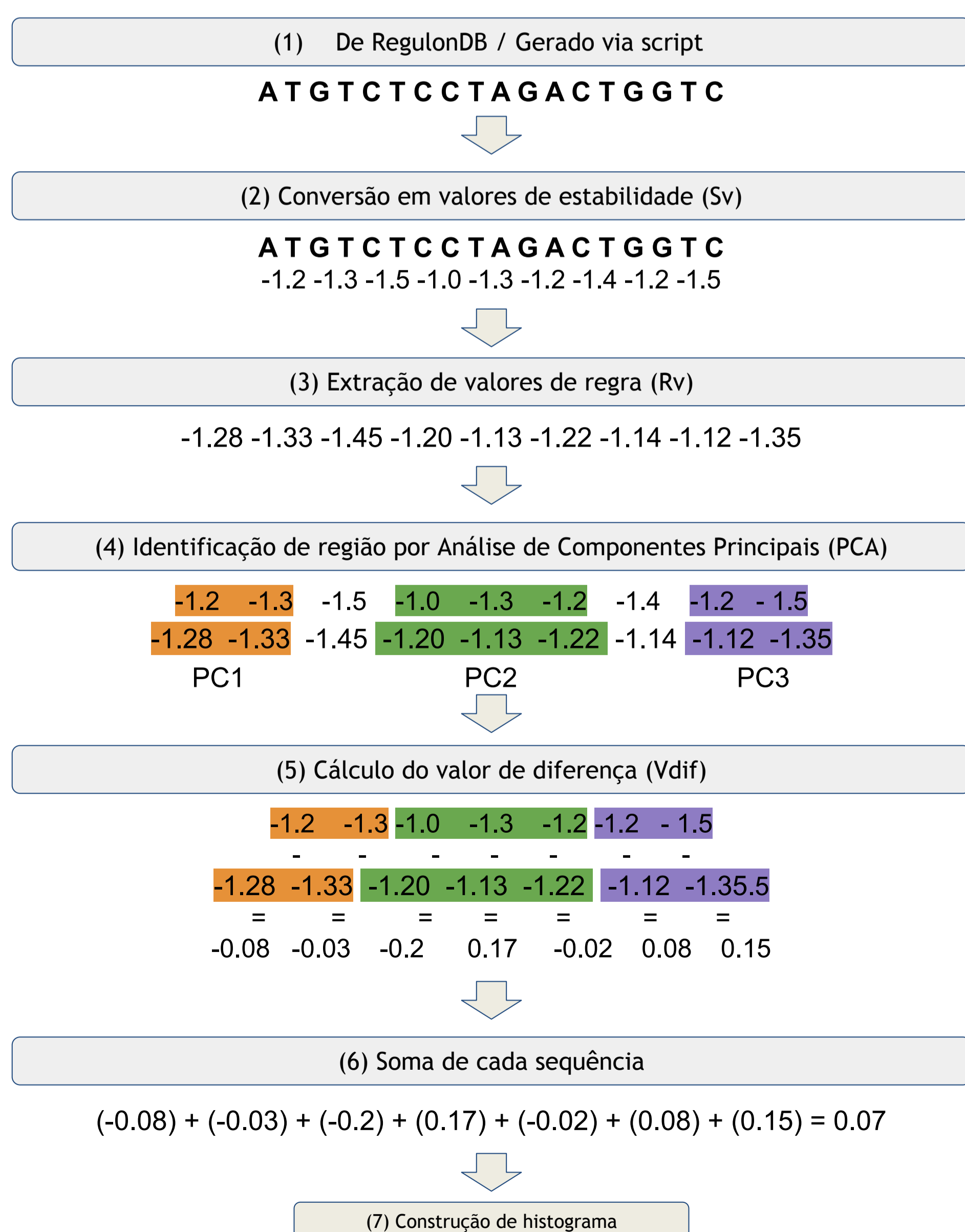


Figura 1 - Esquema simplificado do funcionamento do algoritmo.

RESULTADOS E DISCUSSÃO

Como resultado deste trabalho, obteve-se a implementação do algoritmo proposto, cujo resultado de predição para o fator σ_{28} , dentre amostras do banco de dados RegulonDB, está apresentado na Tabela 1 em comparação com o resultado da ferramenta BacPP:

Amostra Fator σ	BacPP (Positivo)	BacPP (Negativo)	Algoritmo (Positivo)	Algoritmo (Negativo)
σ_{24}	2	518	260	260
σ_{28}	29	115	52	92
σ_{32}	1	310	126	185
σ_{38}	1	232	142	91
σ_{54}	1	98	38	61
σ_{70}	29	1956	1331	654
σ_{28} embaralhado	0	144	60	84
σ_{28} falso	2	142	77	67

Tabela 1 - Resultado da classificação de amostras provenientes do banco de dados RegulonDB.

Comparando-se o resultado do algoritmo com os resultados da ferramenta, para amostras de fator σ_{28} , obteve-se a seguinte relação apresentada na Tabela 2:

Relação	Quantidade
Acertos em comum	16
Erros em comum	79
Acertos exclusivos do algoritmo	36
Acertos exclusivos da ferramenta	13

Tabela 2 - Relação entre resultados do algoritmo e da ferramenta BacPP, para um total de 144 amostras de fator σ_{28} provenientes do banco de dados RegulonDB.

Conforme apresentado na Tabela 2, calcula-se que há uma discordância de aproximadamente 44,8% para os acertos e de 31,3% para os erros do algoritmo em relação, respectivamente, aos acertos e erros da ferramenta.

CONCLUSÕES

Com o presente estudo foi possível concluir que há uma baixa discordância entre os resultados do algoritmo implementado e da ferramenta BacPP. Para aumentar o grau de assertividade da predição dessa ferramenta, futuramente serão realizados estudos adicionais com o intuito de incorporar este algoritmo a ela como um classificador de segunda etapa.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ALBERTS, Bruce; BRAY, Dennis; HOPKIN, Karen; JOHNSON, Alexander; LEWIS, Julian; RAFF, Martin; ROBERTS, Keith; WALTER, Peter. **Fundamentos da Biologia Celular**. Porto Alegre: Artmed, 2011.
- [2] SILVA, Scheila de Avila e. **Redes Neurais Artificiais aplicadas na caracterização e predição de regiões promotoras**. 2006. 144 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2006.
- [3] SILVA, Scheila de Avila e. **Redes neurais artificiais aplicadas no reconhecimento de regiões promotoras em bactérias Gram-negativas**. 2011. 108 f. Tese (Doutorado) - Curso de Biotecnologia, Universidade de Caxias do Sul, Caxias do Sul, 2011.
- [4] SILVA, Scheila de Avila e; ECHEVERRIGARAY, Sergio; GERHARDT, Günther J.L.. BacPP: bacterial promoter prediction: a tool for accurate sigma-factor specific assignment in enterobacteria. **Journal Of Theoretical Biology**, [S.L.], v. 287, p. 92-99, out. 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.jtbi.2011.07.017>.
- [5] DALL'ALBA, Gabriel. **Development of an algorithm for promoter prediction of *Escherichia coli* using information of the DNA duplex stability for σ_{28} -related sequences**. 2018. 18 f. TCC (Graduação) - Curso de Ciências Biológicas, Universidade de Caxias do Sul, Caxias do Sul, 2018.