



# BacPP: a web-based tool for Gram-negative bacterial promoter prediction

S. de Avila e Silva<sup>1,2</sup>, D.L. Notari<sup>3</sup>, F.A. Neis<sup>3</sup>, H.G. Ribeiro<sup>3</sup> and S. Echeverrigaray<sup>1</sup>

<sup>1</sup>Instituto de Biotecnologia, Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil

<sup>2</sup>Universidade de Caxias do Sul, Campus Universitário de Vacaria, Vacaria, RS, Brasil

<sup>3</sup>Centro de Ciências Exatas e da Tecnologia, Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil

Corresponding author: S. de Avila e Silva

E-mail: [sasilva6@ucs.br](mailto:sasilva6@ucs.br)

Genet. Mol. Res. 15 (2): gmr.15027973

Received November 4, 2015

Accepted December 9, 2015

Published April 4, 2016

DOI <http://dx.doi.org/10.4238/gmr.15027973>

**ABSTRACT.** Bacterial Promoter Prediction (BacPP) is a tool used to predict given sequences as promoters of Gram-negative bacteria according to the  $\sigma$  factor that recognizes it. The first version of BacPP was implemented in Python language in a desktop version without a friendly interface. For this reason, a web version of BacPP is now available with the purpose of improving its usability and availability. The present paper describes the implementation of the web version of this tool, focusing on its software architecture and user functionalities. The software is available at [www.bacpp.bioinfocps.com/home](http://www.bacpp.bioinfocps.com/home).

**Key words:** Promoter prediction; Bioinformatics tool; Gene regulation

## INTRODUCTION

The regulation of gene expression plays an important role in the proper response to changing environmental conditions. This is a critical process for single-celled organisms, such as bacteria, and one of the genetic elements that play a role in the bacterial gene regulation process is the promoter region (Feklístov et al., 2014; Narlikar, 2014).

The promoter region is located upstream of a given coding region and its recognition by the RNA polymerase (RNAP) enzyme starts the gene expression process. The RNAP presents a subunit denoted sigma ( $\sigma$ ) factor and, in bacteria, there are several  $\sigma$  factors ( $\sigma^{24}$ ,  $\sigma^{28}$ ,  $\sigma^{32}$ ,  $\sigma^{38}$ ,  $\sigma^{54}$ ,  $\sigma^{70}$ ) that are each responsible for initiating the expression of a set of genes related to a specific cellular function (Gruber and Gross, 2003; Feklístov et al., 2014). The major subunit is  $\sigma^{70}$ , which is known as a housekeeping transcriptional regulator since it is associated with a large number of genes and biological pathways. Besides  $\sigma^{70}$  factor, the other  $\sigma$  factors are associated with only a few genes and they allow for specificity in biological processes. For instance, genes with promoters recognized by  $\sigma^{24}$  and  $\sigma^{32}$  respond to drastic temperature changes. While  $\sigma^{28}$  is associated with mobility genes,  $\sigma^{54}$  is associated with nitrogen metabolism genes and  $\sigma^{38}$  is associated with genes involved in the starvation response (Gruber and Gross, 2003; de Avila e Silva and Echeverrigaray, 2012; Feklístov et al., 2014).

In a global view, promoters present some degree of conservation in their nucleotide content. The canonical definition of the  $\sigma^{70}$  promoter sequence is two consensual motifs, located in the -10 and -35 regions upstream of the first nucleotide transcribed (+1). The content of these regions is usually TATAAT and TTGACA, respectively (Lisser and Margalit, 1993). However, these motifs are not highly conserved among either the promoters recognized by the  $\sigma^{70}$  factor or other  $\sigma$  factors (Feklístov et al., 2014). Therefore, the classification of a given sequence, according to the  $\sigma$  factor that recognizes it, is very important and remains as a challenge in bioinformatics (Song, 2012; de Avila e Silva et al., 2014). There are many approaches to improve bacterial promoter prediction (Burden et al., 2005; Rangannan and Bansal, 2011; Song, 2012; Zhou et al., 2013; de Avila e Silva et al., 2014), but few of them are available on the web or present the  $\sigma$  factor classification.

Aiming at contributing to solving this problem in life sciences research, the present paper describes the web version of the Bacterial Promoter Prediction (BacPP) tool. This tool classifies an input sequence as a promoter or non-promoter of Gram-negative bacteria. In addition to the classification, BacPP provides the probability that each sequence (or fragment of the sequence) will be recognized by a given  $\sigma$  factor (de Avila e Silva et al., 2011).

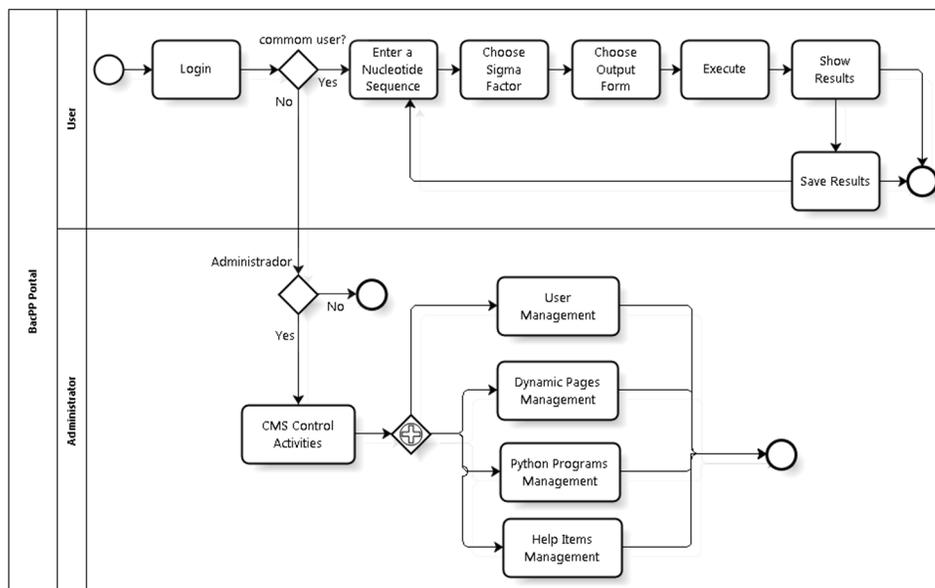
## MATERIAL AND METHODS

BacPP is as a free tool which provides the probability of a Gram-negative promoter sequence to be recognized by a given  $\sigma$  factor. The analytical approach of BacPP was originated by a process of weighting rules extracted from neural networks trained with promoter sequences known to respond to a specific  $\sigma$  factor. Its desktop version was implemented in Python programming language. The complete description of the BacPP approach is found in de Avila e Silva et al. (2011).

In the web version of BacPP (<http://bacpp.bioinfocps.com/home>), the graphical interface was developed by using the PHP (hypertext preprocessor) web programming language. The communication between PHP and Python programs was made by using WSGI (web server gateway interface), which is a standard pattern communication between HTTP server and Python applications (PEP333, 2014). The information about portal contents and administrator area was stored in a MySQL database (MySQL, 2014).

## RESULTS

The web BacPP interface is viewable in all major web browsers and operating systems, including Safari, Internet Explorer, Firefox and Chrome using Linux, Windows, or Apple iOS. Its non-registered user interface is organized in the following levels: i) the initial page that presents the aims of the tool; ii) the help page with some frequently asked questions; and iii) the contact page that allows the user communication with the BacPP administrator. The BacPP tool page is only accessible after user registration, which is classified into two types of profiles: common and administrator. A workflow outlining BacPP user functionalities is provided in Figure 1.



**Figure 1.** A graphical view of the architecture of the BacPP web site.

The login for common users can be obtained either by doing a new registration or with the same login as the IntergenicDB database (<http://intergenicdb.bioinfocps.com/>) (Notari et al., 2014). To carry out the promoter analysis, the user must do the following: i) enter a nucleotide sequence into the form field or upload a file in fasta format; ii) choose the  $\sigma$  factor (one or more than one); iii) choose the output format (screen display or file); and iv) click the run button. After processing, the result will appear according to the user's choice from step three.

The administration area, called the content management system, is restricted to the manager of the portal. In this area, it is possible to: i) create, alter, and remove user information; ii) create dynamic pages (Home, BacPP, Help, and Content tabs); iii) establish the communication between BacPP web tabs and BacPP Python programs; and iv) create and remove items in the help tab.

## DISCUSSION

The main goal of the BacPP web version is to offer a user friendly tool. The second goal

is to assist both gene regulation research and the progress of state of the art bioinformatics. For future implementations, we are committed to the improvement of prediction by reducing the number of false positives. This will be achieved with the incorporation of structural properties of promoter sequences, such as duplex stability, curvature, and bendability of DNA. Additionally, a Gram-positive version of BacPP is currently under development.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We thank the members of our teams for the discussion and comments, and the University of Caxias do Sul for the support.

## REFERENCES

- Burden S, Lin YX and Zhang R (2005). Improving promoter prediction for the NNPP2.2 algorithm: a case study using Escherichia coli DNA sequences. *Bioinformatics* 21: 601-607. <http://dx.doi.org/10.1093/bioinformatics/bti047>
- de Avila e Silva S and Echeverrigaray S (2012). Bacterial promoter features description and their application on E. coli in silico prediction and recognition approaches. In: *Bioinformatics* (Pérez-Sánchez H, eds). Published by InTech, Rijeka, Croatia, 241-260.
- de Avila e Silva S, Echeverrigaray S and Gerhardt GJ (2011). BacPP: bacterial promoter prediction--a tool for accurate sigma-factor specific assignment in enterobacteria. *J. Theor. Biol.* 287: 92-99. <http://dx.doi.org/10.1016/j.jtbi.2011.07.017>
- de Avila e Silva S, Forte F, T S Sartor I, Andrighetti T, et al. (2014). DNA duplex stability as discriminative characteristic for Escherichia coli  $\sigma^{(54)}$ - and  $\sigma^{(28)}$ - dependent promoter sequences. *Biologicals* 42: 22-28. <http://dx.doi.org/10.1016/j.biologicals.2013.10.001>
- Feklistov A, Sharon BD, Darst SA and Gross CA (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* 68: 357-376. <http://dx.doi.org/10.1146/annurev-micro-092412-155737>
- Gruber TM and Gross CA (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* 57: 441-466. <http://dx.doi.org/10.1146/annurev.micro.57.030502.090913>
- Lisser S and Margalit H (1993). Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res.* 21: 1507-1516. <http://dx.doi.org/10.1093/nar/21.7.1507>
- MySQL (2014). Open source database. Available at [http://www.mysql.com/]. Accessed November 20, 2014.
- Narlikar L (2014). Multiple novel promoter-architectures revealed by decoding the hidden heterogeneity within the genome. *Nucleic Acids Res.* 42: 12388-12403. <http://dx.doi.org/10.1093/nar/gku924>
- Notari DL, Molin A, Davanzo V, Picoletto D, et al. (2014). IntergenicDB: a database for intergenic sequences. *Bioinformatics* 10: 381-383. <http://dx.doi.org/10.6026/97320630010381>
- PEP333 (2014). Python web server gateway interface v1.0. Available at: <http://www.python.org/dev/peps/pep-0333/>. Accessed November 20, 2014.
- Rangannan V and Bansal M (2011). PromBase: a web resource for various genomic features and predicted promoters in prokaryotic genomes. *BMC Res. Notes* 4: 257. <http://dx.doi.org/10.1186/1756-0500-4-257>
- Song K (2012). Recognition of prokaryotic promoters based on a novel variable-window Z-curve method. *Nucleic Acids Res.* 40: 963-971. <http://dx.doi.org/10.1093/nar/gkr795>
- Zhou X, Li Z, Dai Z and Zou X (2013). Predicting promoters by pseudo-trinucleotide compositions based on discrete wavelets transform. *J. Theor. Biol.* 319: 1-7. <http://dx.doi.org/10.1016/j.jtbi.2012.11.024>