DOI 10.18699/BGRSSB-2018-139

Genome-wide prediction of transcription factor binding sites in cassava via phylogenetic footprinting between plants in Euphorbiaceae family

S. Kalapanulak[1, 2]*, C. Rangsiwutisak[3], T. Saithong[1, 2]

[1] *Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, Thailand*

[2] *Systems Biology and Bioinformatics Research Laboratory, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, Thailand*

[3] *Bioinformatics and Systems Biology Program, School of Bioresources and Technology, and School of Information Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, Thailand*

** e-mail: saowalak.kal@kmutt.ac.th*

Key words: cassava, regulatory elements, transcription factor binding site, phylogenetic footprinting

Motivation and Aim: Discovering the list of all transcription factor biding sites (TFBSs) on promoter regions in the genome is vital for a complete understanding of transcriptional regulation inside the cell. In order to unravel the systems regulation, several computation approaches have been applied to predict TFBSs and their transcription factors (TFs), for example, TFBS scan based on sequence similarity between known TFBSs of model organisms and promoter sequences of interested organisms. To overcome the limitation of TFBS information, phylogenetic footprinting approach was proposed under the hypothesis that the regulatory elements will be conserved across the related species via evolutionary conservation [1, 2].

Methods and Algorithms: Therefore, in this work, the phylogenetic footprinting approach was applied to identify all putative transcription factor binding sites (TFBSs) in cassava, the emphasized plant for food and energy security in the 21st century. Firstly, the related plants in the same Euphorbiaceae family as cassava were selected, i. e. physic nut and castor bean. The 10,890 orthologous groups between cassava and the other two related plants were identified via bi-directional BLASTp. The upstream sequences of each orthologous group were retrieved from the three plant genomes in the range up to 2,000 bps from translation start site without the overlapping coding sequences with previous gene for identifying TFBSs via Multiple Em for Motif Elicitation (MEME) and Analysis of Motif Enrichment (AME) tool.

Results: Finally, 12,925 candidate TFBSs were discovered from 7,769 cassava genes functioning as several biological components such as enzymes and regulatory proteins.

Conclusion: These results will be useful for proposing regulatory elements found in non-coding DNA, and further hypothesizing a transcriptional regulation in cassava plant.

Acknowledgements: Supported by Thailand research cluster (NSTDA, TRF, ARDA, STI, HSRI and NRCT) with grant ID: P-17-51609 and King Mongkut's University of Technology Thonburi, Thailand for the travelling grant in BGRS\SB-2018 conference.

References

1. Katara P., Grover A., Sharma V. (2012) Phylogenetic footprinting: a boost for microbial regulatory genomics. Protoplasma. 249(4):901-907.
2. Hu J. et al. (2014) Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae Nannochloropsis. Sci Rep. 26(4):5454.