DOI 10.18699/WIBSB-2018-16

Computer analysis of the influence of the presence of transcription factors on plant genome evolution

A.I. Dergilev[1, 2, 3], A.V. Tsukanov[2, 4], Y.L. Orlov[1, 2]*

[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Université Grenoble Alpes, Grenoble, France*
[4] *Perm National Research Polytechnic University, Perm, Russia*
* *e-mail: orlov@bionet.nsc.ru*

Key words: bioinformatics, full-genome analysis, binding sites, PWM, big data

Motivation and Aim: The great evolution of plants led to increasing during plant complexification: 5,700 species of green algae combined in 360 genera exist today. With changing of complexification of plants the specific transcription factors (TF) in plant genomes also have been changing with time as well as TF binding sites [1]. The question to figure out is: "How the presence of a TF changes the genome sequence?" What did evolved first? Does the transcription factor change first and then this created binding sites or are there already binding sites in the genome of the ancient one and the transcription factor evolved the binds of these one?

The idea is to look for TFBS in genomes with or without the TF and test whether enrichment is detectable overall [2]. The work requires full-genome analysis of the families of certain plants as well as highlighting the best results (best TFBS scores among all).

Methods and Algorithms: Although many algorithms for recognizing TFBS exist, tools for using the DNA binding models they generate are relatively scarce and their use is limited among the biologist community by the lack of flexible and user-friendly tools. We use a suite of tools "Morpheus" to analyse transcription factor binding sites (TFBS) on DNA sequences [3]. As input, the program uses a set of sequences in FASTA format and a PWM (Position Weight Matrix) and Python language to write scripts. We use library Matplotlib to get a graphic interpretation of our results (histograms for best TFBS representing at density and arrangement in their vicinity (probability of detecting others, relative position [like ER-IR-DR or 10N distance for ARFBS, high density for LFY or MADS also 10N]).

Availability: "Morpheus" is available at http://biodev.cea.fr/morpheus/

References

1. Rensing S.A. et al. (2008) "The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science. 319:64.
2. Wasserman W.W., Sandelin A. (2004) Applied bioinformatics for the identification of regulatory elements. Nat. Rev. Genet. 5:276-287. PMID 15131651. DOI 10.1038/nrg13152.
3. Minguet E.G., Segard S., Charavay C., Parcy F. (2015) MORPHEUS, a Web tool for transcription factor binding analysis using position weight matrices with dependency. PLoS One. 10(8):e0135586. DOI 10.1371/journal.pone.0135586.