DOI 10.18699/PlantGen2019-042

Clusters of transcription factor binding sites in plant genomes

Dergilev A.I.[1, 2]*, Babenko R.O.[2], Galieva A.G.[2], Orlov Y.L.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *e-mail: arturd1993@yandex.ru*

Analysis of transcription regulation in plants presents important problem in molecular biology challenging growing volume of sequencing data. The computer analysis of structure of transcripts and the genome organization of crop plants is important for biotechnology and agrobiology. ChIP-Seq technology allows to detect interactions between DNA and proteins and allow to analyze gene expression regulation in genome-wide scale for large genomes of eukaryotes including crop plants. In course of evolution of plants the specific transcription factors (TF) in plant genomes also have been changing with time as well as TF binding sites. Thus, comparison of genome-wide distribution of TF binding sites is important. It is critical to annotate clusters of binding sites of different transcription factors that may function as enhancers in complex fashion. We developed R scripts and computer tools for TFBS analysis, including following steps: Look for TFBS in genomes with or without the TF and test whether enrichment is detectable overall (plot TFBS density as threshold score); Refine analysis by checking the location of bound regions (intergenic/gene/promoter/intron); Refine by checking whether the enrichment increases with the size of the TF family; Search for best TFBS and look at density and arrangement in their vicinity (probability of detecting others, relative position). Then compare regions from a TF-containing genome and a relative plant genome depleted by such sites. We have developed set of computer programs and have integrated them to the program complex for transcription factor binding sites analysis. Construction of clusters of transcription factor binding sites allowed reconstruct gene regulatory networks. We continue work on integration of ChIP-seq transcription factor binding sites data in plant genomes using available data sources.
Acknowledgements: The work was supported by RFBR (18-04-00483) and ICG SB RAS budget project 0324-2019-0040. Calculation was done at Shared Facility Center Bioinformatics, Siberian Supercomputer Center.