

INTEGRATIVE COMPUTER ANALYSIS OF ANTISENSE TRANSCRIPTS AND miRNA TARGETS IN PLANT GENOMES

**Orlov Y.L.^{1*}, O. Dobrovolskaya¹, C.H. Yuan², D.A. Afonnikov¹, Y. Zhu³,
M. Chen^{2**}**

¹Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

²Zhejiang University, Hangzhou, China

³East Agricultural University, Harbin, China

*e-mail: orlov@bionet.nsc.ru, **e-mail: mchen@zju.edu.cn

Non-coding RNA, including small interfering RNAs (siRNAs), are important components of gene expression in eukaryotes, forming a regulatory network. miRNAs are expressed through nucleolytic maturation of hairpin precursors transcribed by RNA Polymerase II or III. Such transcripts are involved in post-transcriptional gene regulation in plants, fungi and animals. miRNAs bind to target RNA transcripts and guide their cleavage (mostly for plants) or act to prevent translation. siRNAs act via a similar mechanism of cleavage of their target genes, but they also can direct genomic DNA methylation and chromatin remodeling. It is estimated that large fraction, up to 30% of all human genes also may be post-transcriptionally regulated by miRNAs. For plant genomes numbers could be higher depending on quality of sequencing and genome annotation. Due to availability of genome and mRNA sequences genome-wide searches for sense-antisense transcripts have been reported, but few plant sense-antisense transcript pairs have been studied. Integration of these data in specialized databases is challenging problem of computer genomics. We have developed set of computer programs to define antisense transcripts and miRNA genes based on available sequencing data. We have analyzed data from PlantNATsDB (Plant Natural Antisense Transcripts DataBase) which is a platform for annotating and discovering Natural Antisense Transcripts (NAT) by integrating various data sources [1]. NATs can be grouped into two categories, cis-NATs and trans-NATs. Cis-NAT pairs are transcribed from opposing DNA strands at the same genomic locus and have a variety of orientations and differing lengths of overlap between the perfect sequence complementary regions, whereas trans-NAT pairs are transcribed from different loci and form partial complementarity. The database contains at the moment 69 plant species. The database provides an integrative, interactive and information-rich web graphical interface to display multidimensional data. Available information for the transcription factors (TF) for each species was retrieved from the Plant Transcription Factor Database. We have compared gene structure for wheat, rice and related plant genomes. The phenomenon of antisense transcription and miRNA interference need further annotation in new sequenced genomes. GO annotation and high-throughput small RNA sequencing data currently available will be integrated to investigate the biological function of such transcripts.

The work is supported in part by RFBR 12-04-00897; Integration projects SB RAS 21, 39, 130, Presidium RAS Programs 6.8, 30.29, No.28.

1. Chen et al. (2012) *Nucleic Acids Res*, 40(Database issue): D1187-93.