

GREENC 2.0: A COMPREHENSIVE DATABASE OF PLANT LONG NON-CODING RNAS

DI MARSICO M.*, PAYTUVI GALLART A.*, SANSEVERINO W.*, AIESE CIGLIANO R.*

*) Sequentia Biotech SL, Carrer del Dr. Trueta, 179, 08005 Barcelona, 08018 Barcelona, Spain

GreeNC2, lncRNAs, bioinformatics

Long non-coding RNAs (lncRNAs) used to be considered as transcriptional noise in the past decades, but lately, this class of molecules has gained increasing attention in epigenetic research and it is now recognized to have an important role in mediating the transmission and the expression of genetic information. LncRNAs are RNA molecules longer than 200 nucleotides with no protein-coding ability, despite this, they are involved in fundamental biological processes, and their activities are complex and diverse. In fact, lncRNAs could help in the regulation of protein modification but also chromatin remodeling, RNA metabolism, transcription, DNA methylation, and many other processes. Due to their activity in a wide number of pathways, lncRNA are very well studied in human and clinical applications. In plants, many lncRNAs have been characterized in model organisms as *Arabidopsis thaliana*, *Zea mays*, and *Solanum lycopersicum*. For instance, *Arabidopsis* lncRNA APOLO (AUXIN-REGULATED PROMOTER LOOP) regulates the expression of un-related distant auxin-responsive genes during the lateral root development by modulating local chromatin conformation. *Zea mays* PILNCR1 is involved in the plant adaptation to phosphate deficiency. LncRNA1459, detected in *Solanum lycopersicum*, has been shown to be involved in the fruit ripening process. In order to help the scientific community to study plant lncRNA sequences and functions, we developed and published the GreeNC database in 2015. Since then, the database has been accessed more than 250000 times and it has become a reference for the plant scientific community working on lncRNAs. In the last years tens of new plant species have been sequenced and for many of the species shown in GreeNC new or updated reference genomes have been published. For this reason we present GreeNC 2.0, a new update where

lncRNAs from 78 new species were added and 16 species were updated, resulting in the annotation of more than 495000 lncRNAs. Moreover, sequence clustering was applied providing information about sequence conservation and gene families. With this additional information researchers will be able to detect whether candidate lncRNAs belong to gene families and if they are conserved across species. The current version of the database is available at:
http://greenc.sequentiabiotech.com/wiki2/Main_Page.