

ASTER-REP, A DATABASE OF ASTERACEAE SEQUENCES FOR STUDYING THE STRUCTURE AND FUNCTION OF TRANSPOSABLE ELEMENTS

VENTIMIGLIA M.*, BOSI E.**, VASARELLI L.***, CAVALLINI A.*, MASCAGNI F.*

*) Department of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, 56124 Pisa (Italy)

**) Department of Earth, Environmental and Life Sciences, University of Genoa, Genova, Corso Europa, 26, 16132 Genova (Italy)

***) CNR, Istituto di Informatica e Telematica, Via Giuseppe Moruzzi, 1, 56124 Pisa (Italy)

transposable elements, Asteraceae, genomic data, database

The annotation of repeat sequences is considered essential for any study of eukaryotic genomes since they generally represent the most abundant part of the genome, and because their detection promotes the completeness required for high-quality genome annotation.

The ideal strategy to identify full-length transposable elements (TEs) is to use structure-based search software.

In this study, the discovery of TEs was carried out in a pool of species belonging to the Asteraceae family, in particular, *Helianthus annuus*, *Lactuca sativa*, *Cynara cardunculus* var. *scolymus*, *Artemisia annua*, *Carthamus tinctorius*, and *Chrysanthemum seticuspe* were chosen. All TEs were retrieved using structure-based software, by detecting all of their fundamental features, i.e., existing in a high copy number, possessing typical sequence motifs, and/or encoding for specific domains. Based on the most current classification, TEs were identified for the following five TE orders: LTR, SINE, TIR, MITE, and Helitron. A total of 334,747 full-length TEs were identified and included in a database, called ASTER-REP, set up on a Linux-Apache-MySQL-PHP (LAMP) system. ASTER-REP use has been made simple and intuitive: the user can choose the species to which the desired sequences belong, and consequently select the class, order, superfamily, and lineage at which these elements are classified. The result of the query is then visible and downloadable in FASTA and GFF formats.

ASTER-REP is so far the most comprehensive directory for full-length TEs, concerning a specific plant family. ASTER-REP will significantly improve transposon annotation not just for the species already present in the database, but also for other Asteraceae genomes, and it will be a useful tool for studies on transposon diversity and dynamics, helping to infer the process of evolution that occurred during the separation of Asteraceae species. Also, the availability of ASTER-REP will allow the scientific community to keep in step with the continuous generation of genomic data.