EVOLUTIONARY PATTERNS OF REPETITIVE DNA IN NON-MODEL SPECIES REVEALED BY GENOME SKIMMING AND GRAPH BASED CLUSTERING

MASCAGNI F.*, GIORDANI T.*, CECCARELLI M.**, CAVALINI A.*, NATALI L.*

*) Dept. of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, 56124 Pisa (Italy)
**) Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia (Italy)

genome evolution, genome skimming, Helianthus, LTR-retrotransposons, repetitive DNA

In the last years, genome skimming, the sequencing of genomic DNA at low-coverage, emerged as a successful approach for the identification of the repetitive component in different plant species, even in the absence of genome sequence. In fact, although the repetitive DNA is often discarded without consideration of any potential use, its key role in the evolution of species is increasingly accepted. Repetitive elements drive genome divergence and recombination through rapid DNA remodelling and the change of host regulatory gene networks, representing a valuable source of information for evolutionary analyses.

Here we report on the identification of DNA repeats and their fine characterization (with regard to abundance, variability and evolution) using methods suitable even for non-model species. We performed a comparative analysis of the repetitive genome fraction in sunflowers (i.e. species belonging to the Helianthus genus), a genus that recently has been studied to understand the ecological genetics of speciation and adaptation, but still uncharacterized about retrotransposon-related variability among species. In our study, we focused on the existent variability of long terminal repeat retrotransposons (at superfamily, lineage and sublineage levels) across ten species and one subspecies of Helianthus.

After determining the relative genome size of each species, genomic DNA was isolated and subjected to Illumina sequencing. Then, different assembling and graph-based clustering approaches allowed exploring the repetitive portion of all genomes.

On average, repetitive DNA in Helianthus species represented more than 75% of the genome, being composed mostly by long terminal repeat retrotransposons. In addition, the prevalence of Gypsy over Copia superfamily was observed and, among lineages, Chromovirus was by far the most represented. Although nearly all the same sublineages are present in all species, we found considerable variability in the abundance of diverse retrotransposon lineages and sublineages.

This large variability might indicate that different events of amplification or loss related to these elements occurred following species separation. Moreover, in some cases, retrotransposons drove large genome size variation in species, (as H. agrestis), whose genome size doubled without changes in ploidy level. Although the occurrence of differently abundant repeats among genomes, our analyses showed the influence of species habit (i.e., annual or perennial) rather than taxonomic relationships between species in determining the accumulation (or the reduction) of specific superfamilies or lineages. A significant separation between annual and perennial species was ascertained for the most abundant lineages of both Gypsy (Chromovirus) and Copia superfamily (Maximus/SIRE), which might be related to the different life cycle length of species. Such
separation was not significant for other retrotransposon lineages, suggesting that habit affected the abundance of specific types of repeats.