

## GENOMIC CHARACTERIZATION OF *HELIANTHUS ANNUUS* AND RELATED SPECIES

A. CAVALLINI\*, L. NATALI\*, T. GIORDANI\*, V. FERRILLO\*, I. JURMAN\*\*,  
A. ZUCCOLO\*\*, N. VITACOLONNA\*\*\*, V. SARRI\*\*\*\*, M. CECCARELLI\*\*\*\*,  
P.G. CIONINI\*\*\*\*, M. MORGANTE\*\*

\*) Dept. of Crop Plant Biology, Genetics Section, Via del Borghetto 80, 56124 Pisa, Italy –  
acavalli@agr.unipi.it

\*\*) Dept. of Crop and Environmental Sciences, Via delle Scienze 208, 33100 Udine, Italy

\*\*\*) Dept. of Mathematics and Informatics, Via delle Scienze 208, 33100 Udine, Italy

\*\*\*\*) Dept. of Cellular and Environmental Biology, Via Elce di Sotto, 06123 Perugia, Italy

*Asteraceae, genome, Helianthus, repetitive DNA, sunflower*

The genome sizes of many plant species differ as a result of variable amounts of repetitive DNA. Grasses are by far the group of plants where most information has been collected for species with large genomes, such as maize, barley, wheat. Very little information is available on genome composition and organization in the Compositae family, which includes important crop species. The availability of information on genome composition, especially on its repetitive component, generates a new array of enabling technologies for genome analysis that facilitates both fundamental and applied research. Improved knowledge of genome composition increases the efficiency of the discovery and development of molecular markers, leads to improved strategies for genetic and physical mapping of genomes and is a prerequisite for sequencing projects both of ESTs and of genomic regions, as well as for positional cloning projects.

With the aim of characterizing the structure and organization of the genome of sunflower (*Helianthus annuus* L.), randomly sheared (nebulised) genomic DNA was used to construct a library, representing a random sample of the genome. The sunflower genome is large (nearly 3,300 Mbp per haploid genome); more than 1,700 sequences were obtained, corresponding to nearly 1.35% of the genome. A clustering analysis was carried out to assess the amount of repeated sequences in the library. Following this analysis and Blastn and Blastx searches, the library resulted composed by 75.2% repeated sequences, of which 28.1% are retrotransposons, 1.3% rDNA, 0.7% putative MITES, 0.7% putative DNA transposons and 44.4% are unclassified.

Clones containing DNA sequences belonging to different classes of putative repeats were subsequently selected and characterized with respect to their redundancy, genomic organization, and in situ localisation. The most repeated sequence in the sunflower genome (27,000 copies per haploid genome) resulted the contig 61, a repeated sequence of unknown nature. LTR-retrotransposons represent a large portion of the sunflower genome. The copy number per haploid genome of the most redundant *gypsy* families ranged from 5,000 to 23,900, that of *copia* families from 1,000 to 5,100, and there are no predominant retrotransposon families. Among sequences containing direct repeats, four clones resulted medium repeated, from 2,500 to 7,800 copies per genome. Other repetitive DNA sequences (DNA transposons, LINES, MITES) show a redundancy of less than 1,000 copies per genome.

The sunflower library was spotted onto membranes and hybridised to genomic DNAs of ten *Helianthus* species (4 annuals and 6 perennials) and of twelve non-*Helianthus* Asteraceae species. It was observed that redundant sequences of sunflower are usually redundant also in different *Helianthus* species, while they are very differently redundant in other Asteraceae, suggesting that *Helianthus* genome structure was realized already in the progenitor of this genus, before *Helianthus* species separation. Moreover, large differences occur also between *Helianthus* and *Viguiera* or *Tithonia*, i.e. two genera very close to *Helianthus*, indicating that dramatic changes in the genome structure have occurred in the *Helianthus* progenitor and contributed to the formation of this genus. The majority of retroelements amplified before *Helianthus* speciation, though retroelement activity can be inferred from our data also after speciation.