Proceedings of the Joint Meeting AGI-SIBV-SIGA
Assisi, Italy – 19/22 September, 2011
ISBN **978-88-904570-2-9**

Poster Communication Abstract – 2A.44

TOWARDS THE CONSTRUCTION OF OLEAREP: A DATABASE OF REPEATED SEQUENCES OF *OLEA EUROPAEA* L.

BARGHINI E.*, COSSU R.M.*, GIORDANI T.**, CATTONARO F.***,
MORGANTE M.***,****, NATALI L.*, CAVALLINI A.*

*) Dept. of Crop Plant Biology, University of Pisa, Via del Borghetto 80, 56124 Pisa (Italy)
**) SSSUP S. Anna, P.zza Martiri della Libertà 33, 56127 Pisa (Italy)
***) Institute of Applied Genomics, Via Linussio, 33100 Udine (Italy)
****) Dept. of Crop and Environmental Sciences, University of Udine, Via delle Scienze, 33100 Udine (Italy)

Repeated sequences, retrotransposons, genome sequencing, next generation sequencing

Improved knowledge of genome composition, especially of its repetitive component, generates information that is of importance in both theoretical and applied research, such as for improving strategies for genetic and physical mapping of genomes and for the discovery and development of molecular markers. Moreover, knowledge of genome composition is a prerequisite for the annotation steps in sequencing projects.

Despite the importance of olive as a crop, studies on structural genomics of *Olea europaea* are rare. In particular, the repetitive component of olive genome has been studied mostly at cytogenetic level, evidencing the occurrence of tandem repeats in centromeres and telomeres. Concerning transposable elements, only a few sequences have been isolated and characterized unitil now.

With the aim of producing a database of repeated sequences to be used for annotation in sequencing projects, we have sequenced 1.2x genomic DNA of olive, cv. Leccino, using the Illumina NGS technique.

We have analysed 1.8 Gb (25,000,000 of 75 nt-long Illumina reads) of sequences of genomic DNA, corresponding to 1.2x coverture. These sequences were assembled using mainly the CLC Bio Workbench 5.0 software following different procedures. In a first assembly, using default parameters, we obtained 105,507 contigs (N_{50} = 252). Of these, 3474 contigs were longer than 500 bp and were annotated against NCBI databases and databases of repeated sequences. On the whole, we identified 581 LTR retrotransposon fragments (350 *Copia*-like, 231 *Gypsy*-like), 28 non-LTR retrotransposon fragments, 14 fragments of unidentified retrotransposons, 76 DNA transposons fragments, 22 putative helitron fragments. 1312 contigs did not show any similarity to known sequences.

In other experiments, the pool of Illumina reads was splitted into 8 or 16 portions covering 0.15x or 0.075x each and assembled separately; the resulting contigs were assembled at their turn using CAP3 assembler obtaining 3254 and 3114 supercontigs. All supercontigs and contigs were masked against the previously obtained *Olea* database. This allowed to identify 306 supercontigs, specific to these assembly procedures and representing mostly retrotransposons and repeated sequences.

We are now sequencing and annotating a number of BAC clones from a library BAC of *O. europaea* cv. Leccino, to identify complete repeated sequences and implement the database.

This database will first applied to the annotation process of the olive genome, that is being subjected to complete sequencing.

Research work supported by MIPAAF, Project OLEA-Olive Genomics and Breeding.