

## **DE NOVO ASSEMBLY AND GENOME STRUCTURE ANALYSIS IN THE POPULUS GENUS**

GIACOMELLO S.\*\*\*, ZAINA G.\*, VEZZI F.\*\*, SCALABRIN S.\*\*, FELICE N.\*,  
CATTONARO F.\*\*\*\*\*, MORGANTE M.\*\*\*

\*) Dipartimento di Scienze Agrarie e Ambientali, Università di Udine, Via delle Scienze 208,  
33100 Udine (Italy)

\*\*) Istituto di Genomica Applicata, Parco Tecnologico ‘L. Danieli’, Via Linussio 51, 33100 Udine  
(Italy)

\*\*\*) IGA – Technology Services, Parco Tecnologico ‘L. Danieli’, Via Linussio 51, 33100 Udine  
(Italy)

*Next-generation sequencing technologies, genome assembly, Populus nigra, Populus trichocarpa*

*De novo* sequencing of a genome is today accessible and affordable thanks to the advent of the next-generation sequencing technology that has made sequence data production accurate, cheap and fast. Since the release of this new technology, many genome sequences have been published but comparative and structural genomics analyses are still a challenging issue. Given the huge amount of sequence data available, nowadays a special effort has to be given to the latter analyses which are crucial to better understand both the evolution and the composition of the different genomes.

Exploiting the Illumina technology and a *de novo* assembly approach, the present work aims to obtain the genome sequence of an Italian genotype of *Populus nigra*, the native European poplar species which is very important for wood and paper industry. We sequenced the poplar tree at high coverage (90X) using different kinds of libraries in order to solve repetitions and allow the contig scaffolding: technical and critical aspects will be discussed. Then, we focused on two different softwares to perform the *de novo* assembly: performance comparisons and results will be provided. On the selected assembly (length 318 Mb and N50 4487 bp), we developed an analysis pipeline to characterize the contig content in terms of genes and repetitive elements, and sequence novelty compared to *P. trichocarpa*, the American poplar species sequenced using the Sanger method. A test experiment was run on almost 300 randomly-selected contigs in order to validate the pipeline which, due to the low error rate detected, proved to be efficient and accurate in comparing and characterizing the genome sequences concerned. We think our pipeline can be applied to the comparative genome analysis of different closely related organisms.

The comparative approach between the two *Populus* species will be exploited to introduce the concept of the pan-genome, which includes core genomic features common to both species and a dispensable genome composed of non-shared DNA elements that can be individual- or population-specific and important for explaining plasticity, phenotypic variation and heterotic behavior.