



The annotation of the Poplar genome is one of many tasks that will be done by the IPGC (International *Populus* Genome Consortium) in order to provide the scientific community with the genome of a new plant model organism.

The IPGC is a cooperation between the U.S. Department of Energy (DOE) Joint Genome Institute (JGI, USA), INRA – Nancy (France), Umeå Plant Science Centre (Sweden), National Center Genome Research (USA), Oak Ridge National Laboratory Plant Genomics Group (USA), Université Laval (Canada) and Ghent University VIB-PSB (Belgium) in which the tasks of sequencing, providing full length cDNAs and gene prediction are distributed.



The poplar tree is a perennial plant with great economic value and increasing scientific importance. Poplar is rapidly becoming the model organism for the tree biotechnology: the fact that it is a perennial tree makes it an excellent organism for research on bud development, dormancy and wood-formation.

Poplar genome:

- *Populus balsamifera* ssp. *Trichocarpa*
- 19 chromosomes
- estimated genome size ± 500 Mbp

Sequencing:

- sequencing was done by the JGI (USA)
- whole-genome shotgun strategy
- $\sim 8x$ coverage

EST/ cDNA resources:

- 4166 full length inserts, Umeå Sweden
- 405 cDNAs, ORNL USA
- 905 full length inserts, UBC Canada

2nd assembly:

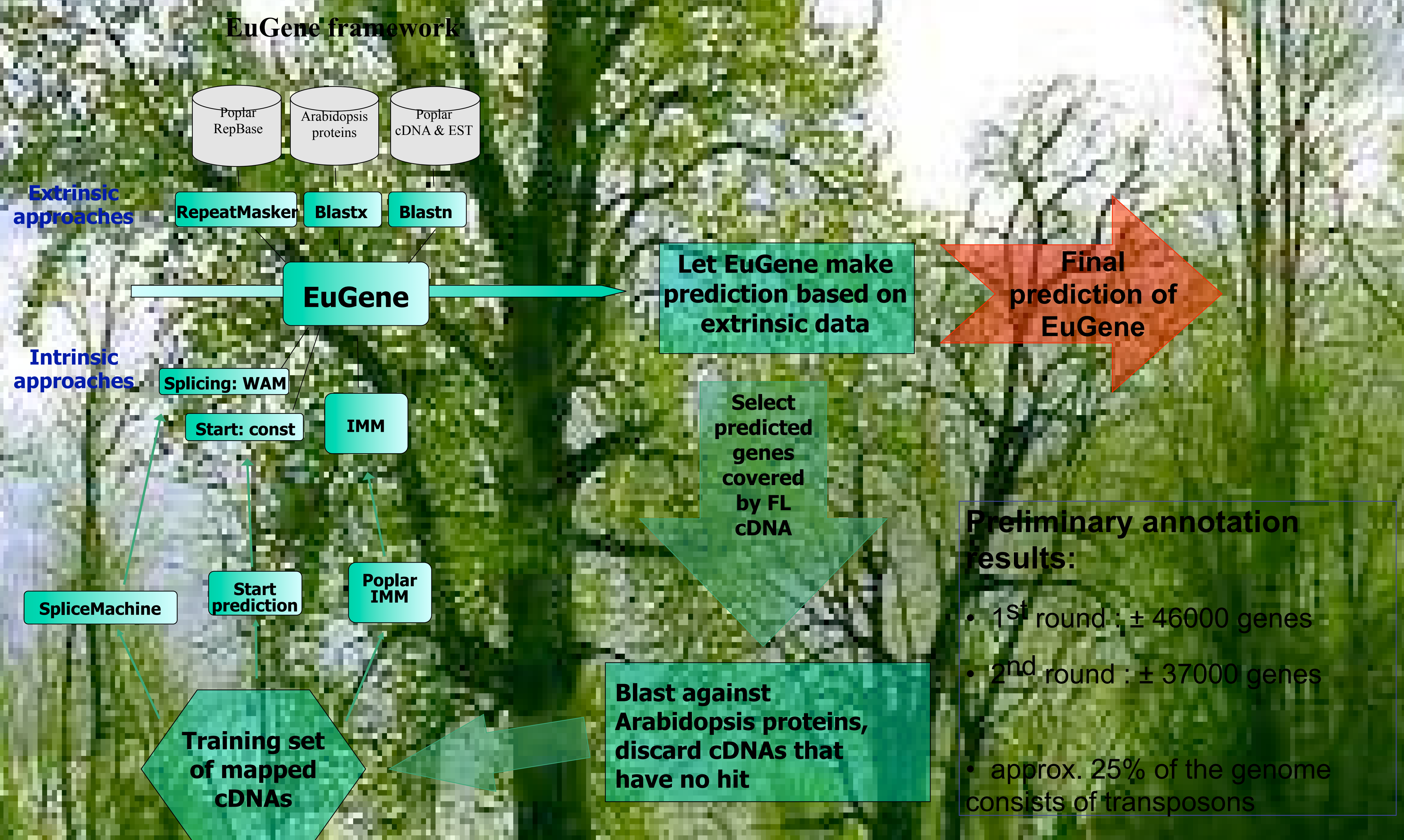
- 2447 scaffolds
- biggest 9.97Mbp, smallest 2.5Kbp
- total size current assembly : 409 Mbp



Creating a training set: semi-automated pipeline.

Because manually checking full length (FL) cDNAs is a very time consuming task we've created an semi-automated pipeline by using our annotation platform EuGene.

The EuGene platform is particular due to it's ability to use extrinsic data as input to enhance its predictions. By forcing the prediction platform to strictly follow extrinsic data, we can let our genome annotation platform do the checking of all the mapped FLs.



Perspectives:

Poplar is an ideal organism for comparative studies with *Arabidopsis thaliana* enabling one to have a look at regulatory sequences that could explain differences in phenotypes, synteny between genomes and lineage-specific duplications. Also a poplar-driven reannotation of the *Arabidopsis* genome and many other comparative in silico research topics can now be exploited with the availability of a second dicot plant genome.