

Toward a better understanding of plant genomes structure: combining NGS and optical mapping technology to improve the sunflower assembly

Céline CHANTRY-DARMON¹, Stéphane CAUET¹, Céline JEZIORSKI¹, Nathalie RODDE¹, William MARANDE¹, Arnaud BELLEC¹, Sonia VAUTRIN¹, Nicolas LANGLADE², Stéphane MUNOS², Jérôme GOUZY² and Hélène BERGES¹

¹French Plant Genomic Resource Center, INRA - CNRGV, France
²Laboratory of Plant-Microbe Interactions, INRA - LIPM, France

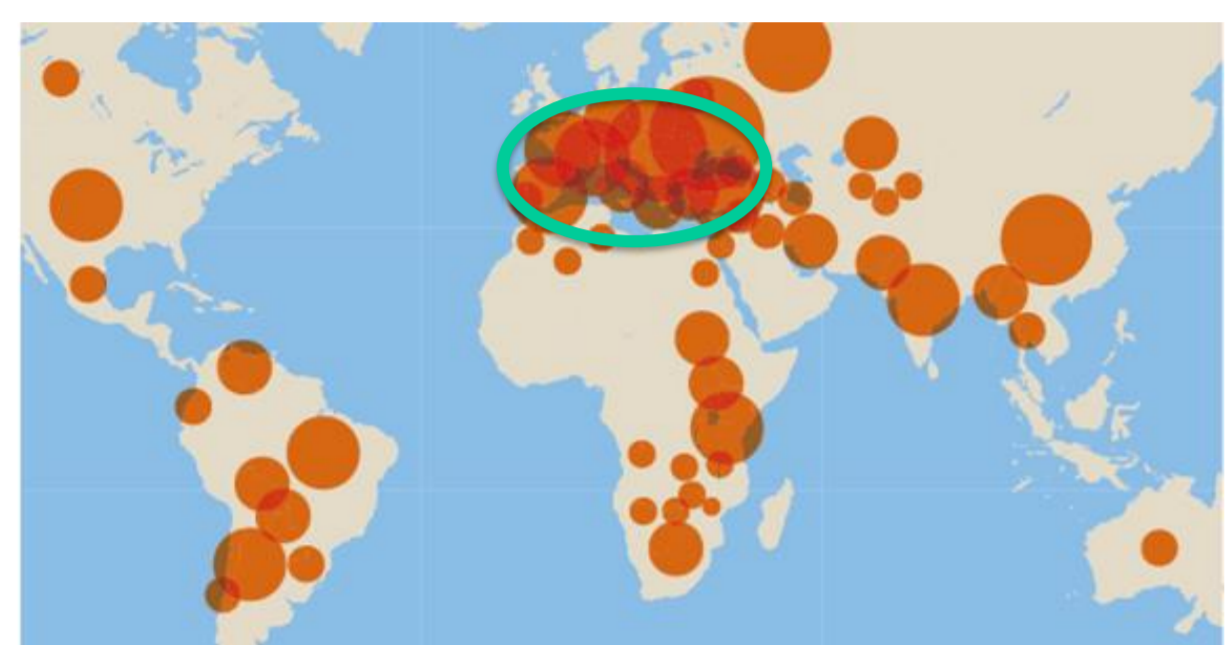
Agricultural research must deal with major issues on various scales, in the context of the climate change, the population growth, the need to face limited energy resources. In this context, better understanding how plants evolve and adapt is a major goal. Increasing plant genomes knowledge is essential to tackle the various scientific questions that arise from that. Despite the revolution of the Next Generation Sequencing technologies including the technologies allowing long reads, the study of plant genomes remains challenging due to their high complexity: size, polyploidy and the high percentage of repetitive elements. In order to have a better understanding of complex genomes the CNRGV has invested in an Irys Instrument from BioNano Genomics which allows the rapid construction of physical whole genome maps.

Economical and scientific context

An important crop in Europe

39 Million tons of seed produced worldwide

80% in Europe



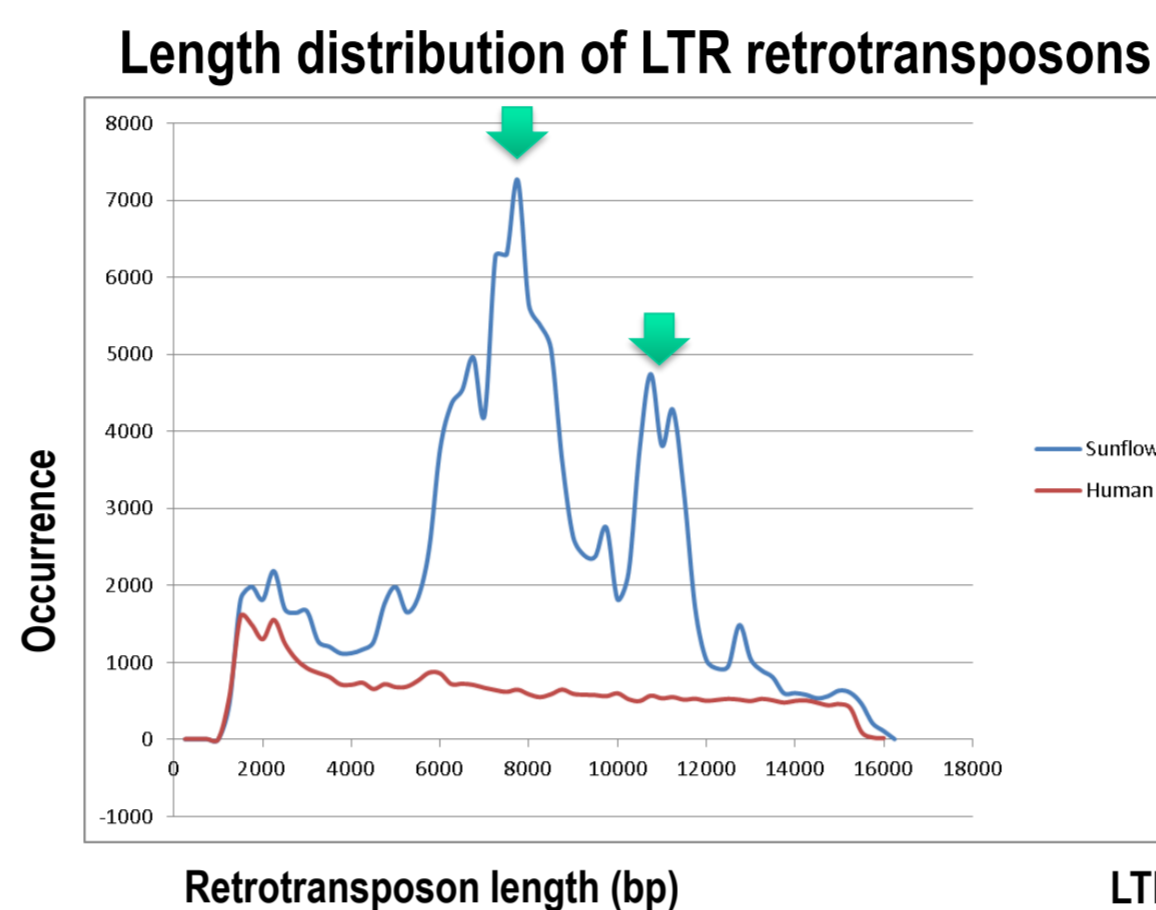
30 Million hectares worldwide

71% in Europe

The global production of sunflower seeds has to increase to meet growing demand (human food, animal feed, green chemistry...)

➤ Societal challenge

Sunflower genome contains long repeats sequences



Repeats = 33% of the sunflower genome

Repeats = 8% of the Human genome

Two major repeats in the sunflower genome : 8 kb and 11.5 kb

➤ The repeats make the assembling very difficult



- *Helianthus annuus*
- 3.6 Gb
- 2n = 34 chromosomes

PacBio Genome Assembly

- More than 100X PacBio sequences
- XRQ sunflower line
- 12 318 contigs
- N50 : 524 kb
- 81% of the genome inside contigs

Building Whole Genome Optical Map

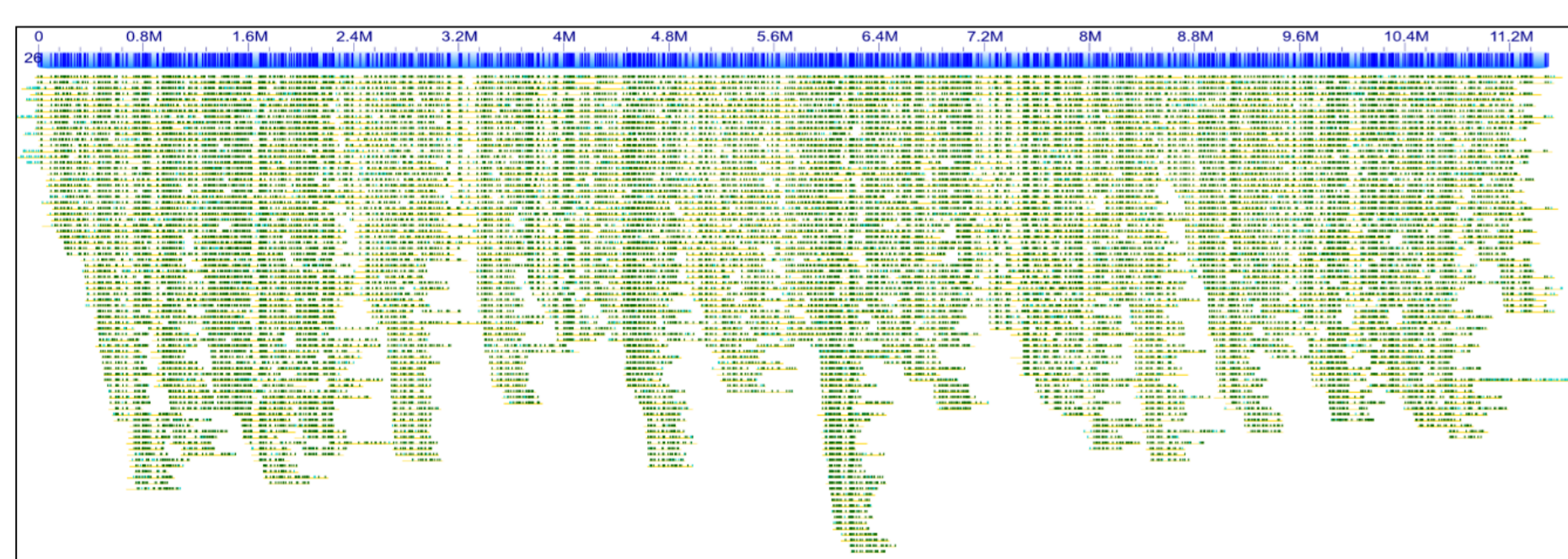
Bionano Workflow



- Direct visualization of long DNA molecules (>100 kb)
- Provides real physical distance information

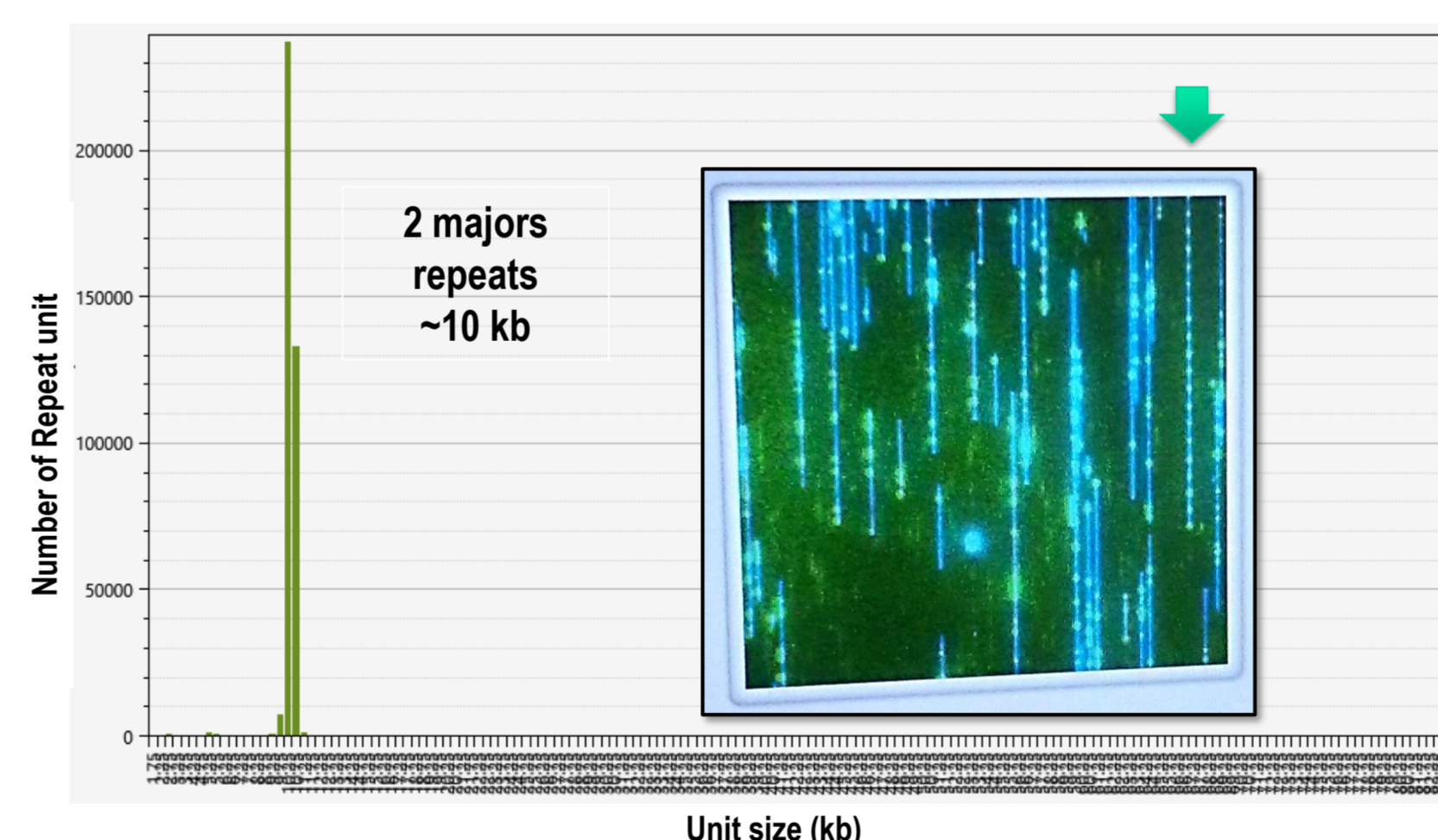
Improvement of the Sunflower Genome Assembly Quality

- HMW DNA Extraction of fresh young dark treated leaves
- 2 nicking enzymes : BspQ1 (176X) & BssS1 (167X)
- Molecules N50 : BspQ1 (206 kb) & BssS1 (187 kb)
- Molecules from 150 kb to 2.3 Mb

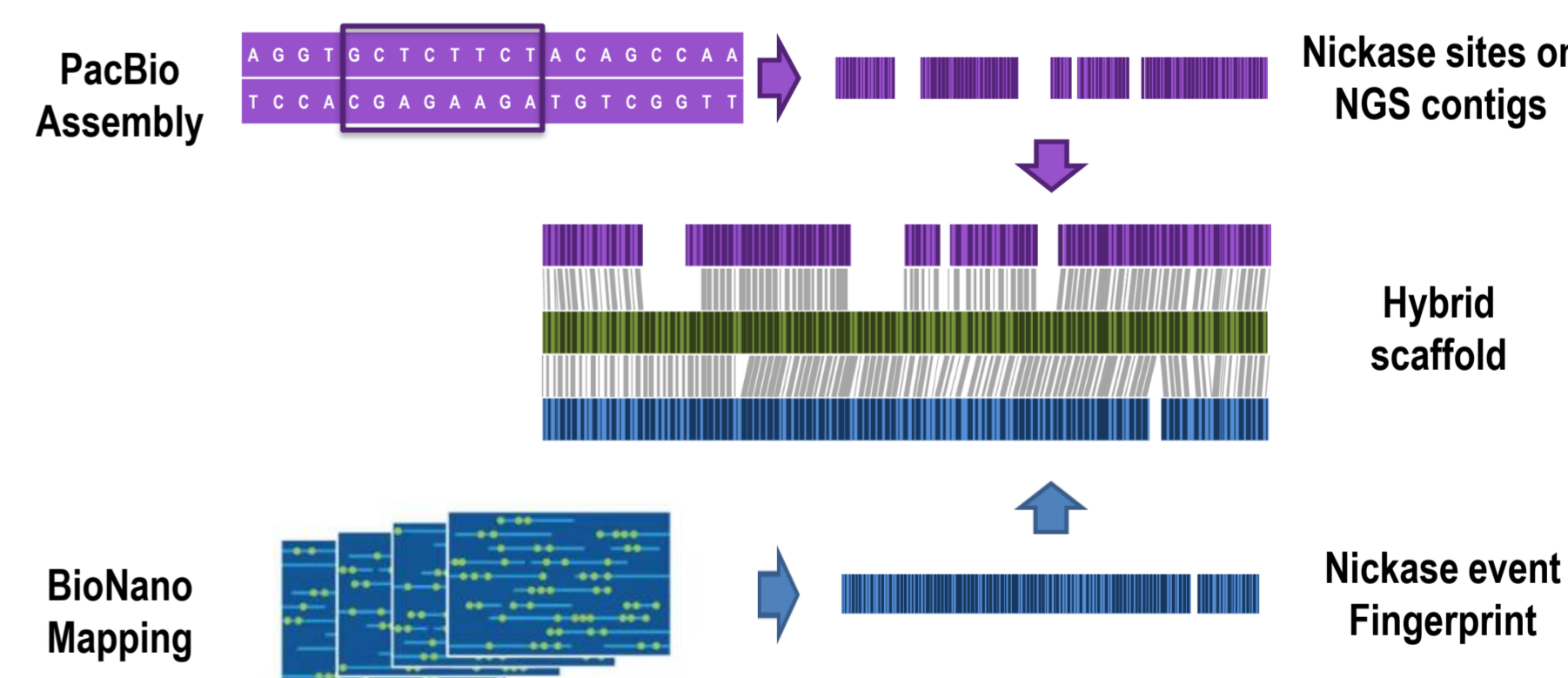


Example of a BioNano map with the molecules alignment

Repeats visualization



Hybrid scaffolding



	PacBio Assembly	BioNano BspQ1 Assembly	Hybrid scaffold BspQ1	BioNano BssS1 Assembly	Hybrid scaffold 2 Step
Count	12318	2228	1430	4287	1069
Median length (Mb)	0.120	0.999	1.442	0.551	1.914
N50 length (Mb)	0.524	1.979	2.87	0.968	4.166
Max length (Mb)	3.35	11.49	17.45	5.322	24.670
Total length (Mb)	2930	3191	2922	3112	2960
% genome coverage	81%	88%	81%	86%	82%

This preliminary optical mapping results are enhancing the sunflower genome with more than 7 fold improvement of the N50 length. The two-step hybrid scaffolding strategy using two nicking enzymes improves significantly the resulting N50. Therefore, this technology offers new possibilities to contribute to the understanding of genomes. It will not only improve the assembly of genomes obtained from new sequencing technologies but also will allow to explore diversity within a species by highlighting structural variations (insertion, deletion, translocation, copy number variation (CNV)).