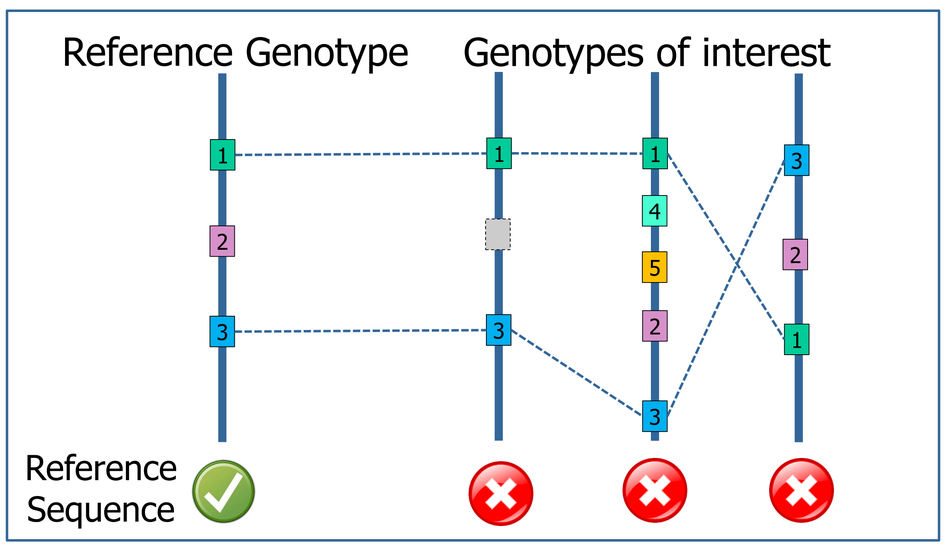
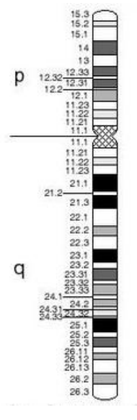
# Combining old genomic strategies with new technologies to decipher the complex structure of plant genomes

PLANT GENOMIC CENTER

Nathalie RODDE, Carine SATGE, Céline CHANTRY-DARMON, Margaux-Alison FUSTIER, Stephane CAUET, Caroline CALLOT, William MARANDE, Sandrine ARRIBAT, Sonia VAUTRIN, Arnaud BELLEC and Hélène BERGES

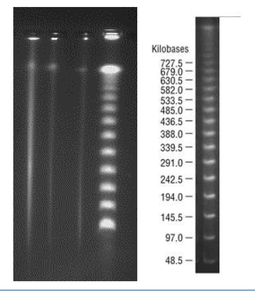
Among living organisms, plants display a high level of genomes complexity due to their large size, variations in ploidy levels and high percentage and variability of repetitive elements. Despite the Next Generation Sequencing revolution including the recent long read technologies, it remains challenging to obtain high quality assemblies at the genome scale. In order to be efficient, when addressing a scientific question, it is important to choose the relevant strategy consistent with the raised topic: exhaustive information on whole genome is not always required while reliable and quality information of the region of interest is crucial and necessary.



Genetic & sequence data

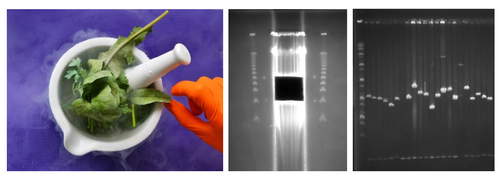
Pulse field gel electrophoresis of sunflower high molecular weight DNA. Starting material : 2g of frozen leaves. Dilution 1 (1), 1/3 (2) and 1/6 (3).



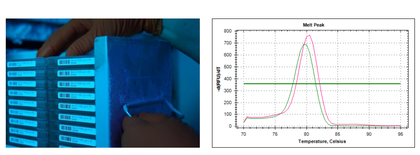

**High Molecular Weight DNA extraction**

NGS data

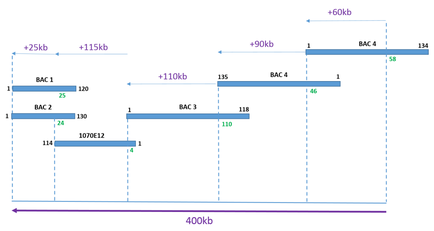
## BAC library construction to decipher plant genome complexity and target genomic region of interest



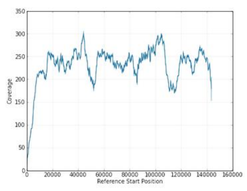
- BAC library construction and characterization



- Screening of the transformants with specific markers
- Identification of positive BAC clones
- BAC clone characterization (BES, insert size)



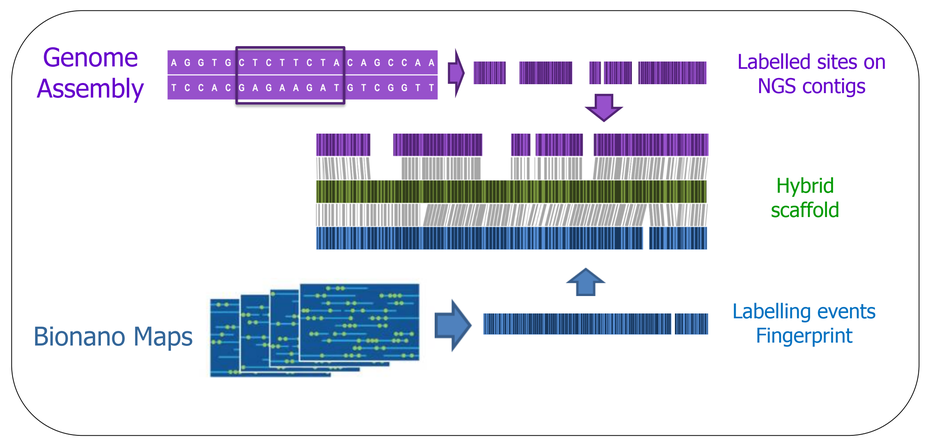

- Assembly of the MTP of the targeted region
- Assembly of the MTP of the targeted region




- BAC-Pool Sequencing (PacBio, barcoded libraries)
- 1 contig per BAC (> 60X coverage)

## Optical Map production using the Saphyr system from Bionano Genomics

Hybrid scaffolding strategy :

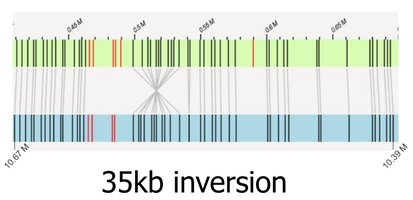



## 1 - Improvement of the sunflower XRQ genome assembly

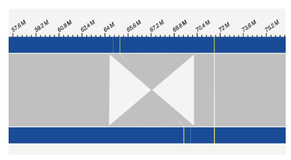
| | Statistic | Original BNG | Original sequence | Sequence used in hybrid scaffold | Hybrid scaffold | Hybrid A + leftover unscaffolded sequence |
|---|---|---|---|---|---|---|
| DLE-1 | Number of maps | 69 | 11676 | 8738 | 25 | 5317 |
| | N50 (Mb) | 175.21 | 0.52 | 0.46 | 176.33 | 175.95 |
| | Total length (Mb) | 3057.67 | 2926.51 | 2792.45 (95.42%) | 3000.44 | 3134.36 |

2359 cuts were made on 1167 sequences during chimera detection.
Cuts can be due to chimera or allelic difference.

## 2 – Large structural variations detection

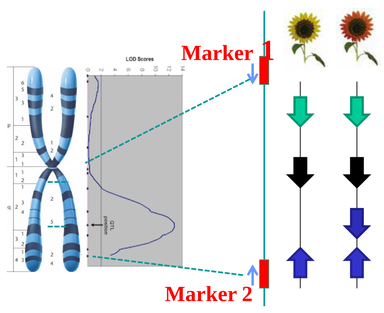


35kb inversion



5,2Mb inversion

## CRISPR-CATCH targeting strategy




We adapted a sequence capture approach to plant genomes, called CATCH (Cas9-Assisted Targeting of Chromosomal segments, Jiang et al., 2015).
This system uses two RNA-guided Cas9 enzymes to capture genomic regions up to 700 kb.
Thank to long read sequencing technologies, it is possible to obtain a precise genomic information of a region of interest. This method allows to compare structural variations between several genotypes for a genomic region of interest.

Although the quality of reference genomes are significantly increasing, it remains difficult to obtain a relevant information for the whole diversity of a specie. The French Plant Genomic Resources Center (CNRGV) develop several strategies combining large fragment genomic DNA libraries, CRISPR-CATCH targeting strategy and optical mapping technology combined with NGS to obtain very high quality sequences. Our aim is to improve reference genomes through hybrid scaffolding, explore the large structural variations comparing optical maps and finally obtain a high quality sequence at the targeted locus using a CRISP-CATCH or a cloning strategy. I collaboration with research teams, we provide dedicated innovative and efficient genomic tools to better characterize plant biodiversity and understand how plants adapt to their environment through the analysis of their genomes and the intra/inter species variability. The complementarity of these strategies allows the production of reliable sequence information, which is essential to link a genotype to a phenotype.

INRA SCIENCE & IMPACT

CNRGV
24 Chemin de Borde Rouge
Auzeville - CS 52627
31326 Castanet tolosan cedex
Tél: +33 5 61 28 52 53 / Fax: +33 5 61 28 55 64

infocnrgv@.inra.fr
@ CNRGV