

# Development of a CRISPR-Cas9 large DNA fragment targeting technique for plant genomes

Carine SATGÉ<sup>1</sup>, Céline CHANTRY DARMON<sup>1</sup>, Margaux-Alison FUSTIER<sup>1</sup>, Sandra MOREAU<sup>3</sup>, Caroline CALLOT<sup>1</sup>, Stéphane CAUET<sup>1</sup>, Adrien CASTINEL<sup>4</sup>, Nicolas MILON<sup>2</sup>, Arnaud BELLEC<sup>1</sup>, William MARANDE<sup>1</sup>, Pauline DURIEZ<sup>3</sup>, Stéphane MUNOS<sup>3</sup>, Aurélien BANCAUD<sup>2</sup> and Hélène BERGES<sup>1</sup>

<sup>1</sup> French Plant Genomic Center (CNRGV) – INRA, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

<sup>2</sup> LAAS, CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France

<sup>3</sup> Laboratoire Plantes Micro-organismes INRA/CNRS, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

<sup>4</sup> GET Plage, INRA/CNRS, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

## 1- Introduction

In the context of plant improvement and plant adaptation, genomic exploration is a strategy to better understand plant genomes. However, this exploration remains challenging due to the complexity of plant genomes in terms of size, repetitive elements content and various levels of ploidy. Moreover, because of a high intra-species variability, a quality reference sequence is not enough to obtain a precise and reliable information of a genomic region linked to a trait of interest in a specific genotype. New strategies for efficiently targeting large regions of interest in complex genomes are really needed to be able to link a phenotype to a genotype.

Here, we report a new sequence capture approach for large DNA fragments in eukaryote genomes based on the programmable endonuclease function of the CRISPR/Cas9 system. We adapted the first steps of the CATCH method (Cas9-Assisted Targeting of Chromosomal segments from Jiang et al., 2015 and Jiang and Zhu, 2016) to capture and to sequence specific loci from 30 kb to 700 kb from the model plant *Medicago truncatula* and a 120 kb region from a larger and more complex genome, the sunflower *Helianthus annuus*.

## 2- Complexity Of Plant Genomes

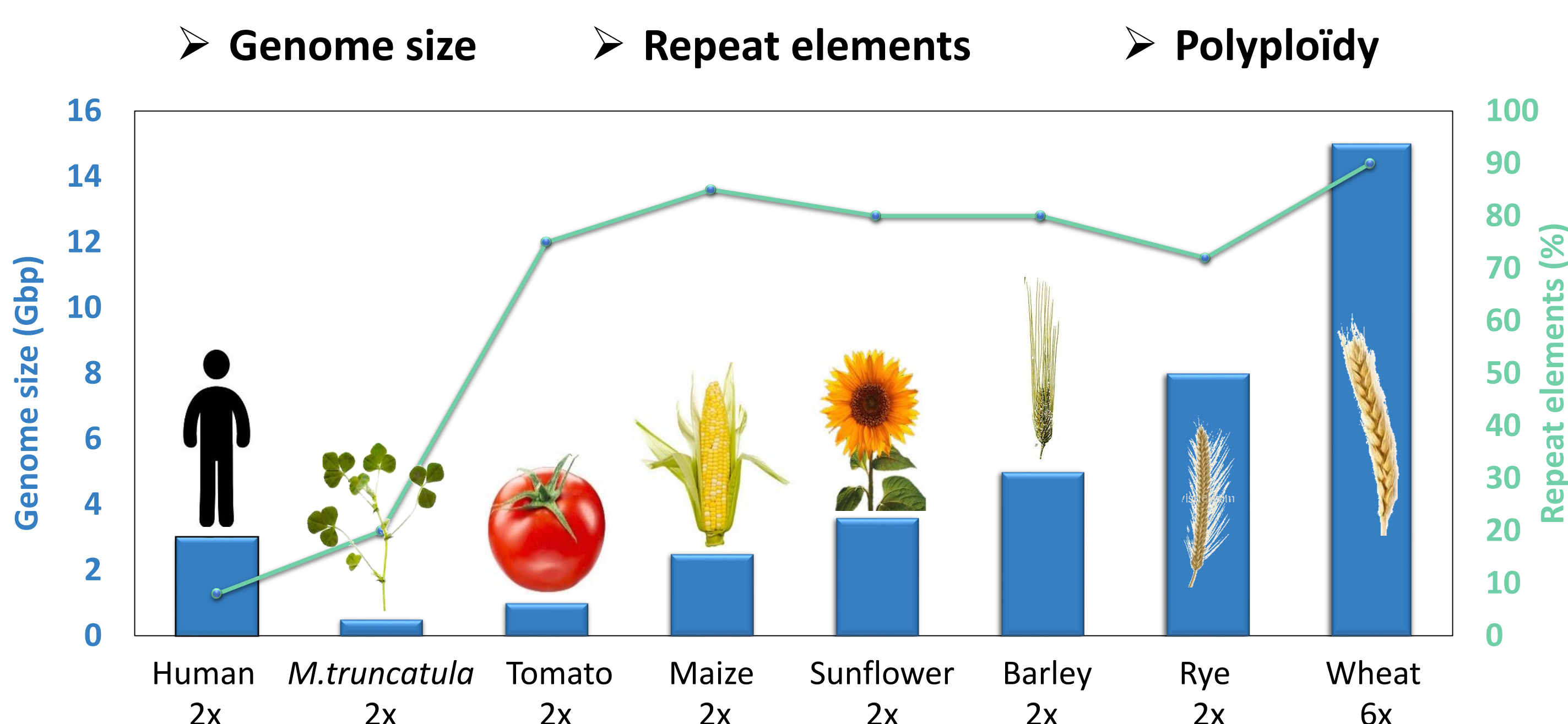


Figure 1. Plants have complex genomes that often reach several gigabases with a high amount of repeat elements and various levels of ploidy.

## 3- The CRISPR-Cas9 System To Target A Genomic Region Of Interest

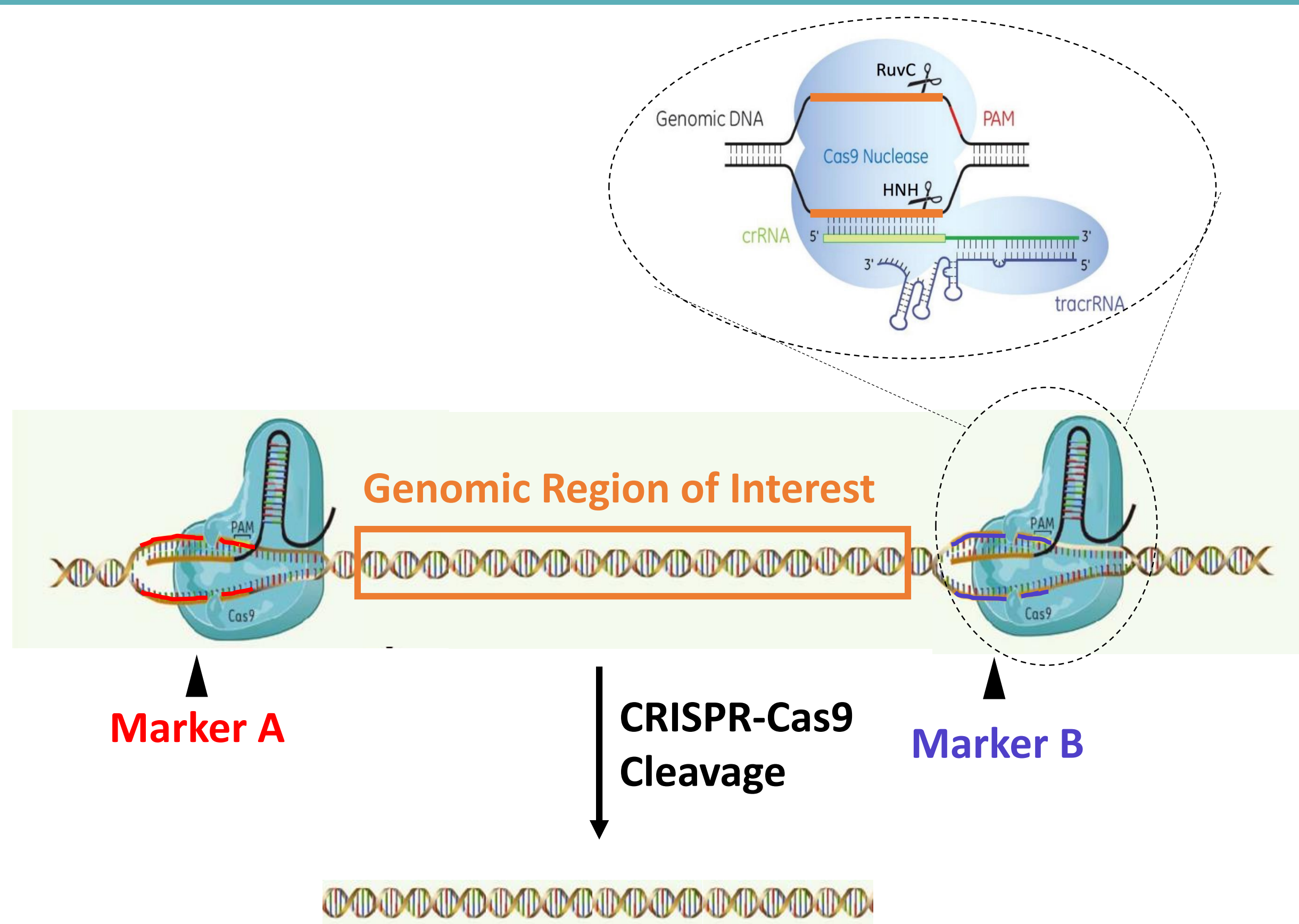


Figure 2. The Cas9 is an RNA-guided endonuclease that creates site-specific cleavage of double stranded DNA. We design small guide RNAs by using the marker sequences flanking the region of interest.

## 4- CATCH Technique: General Workflow

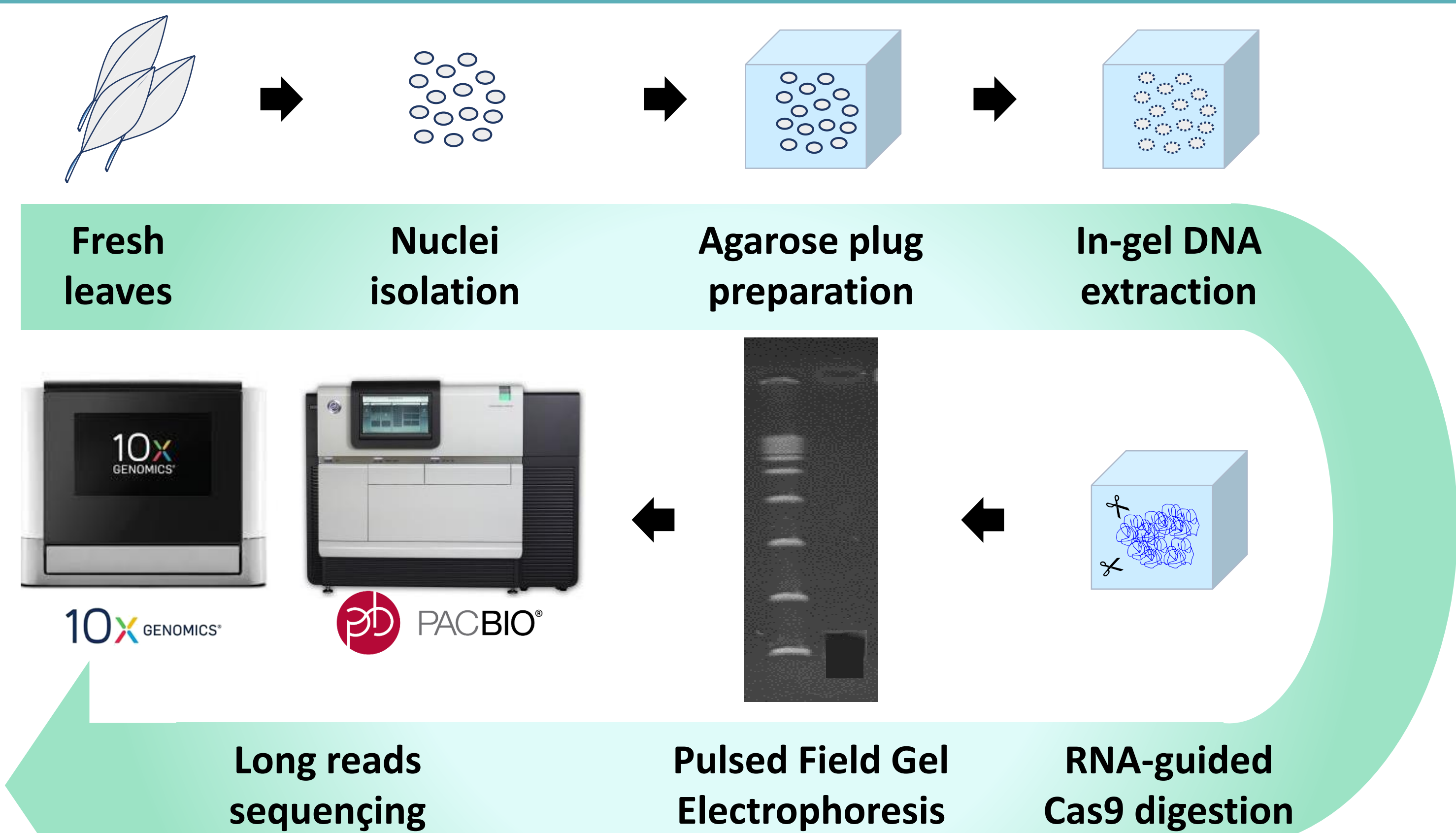
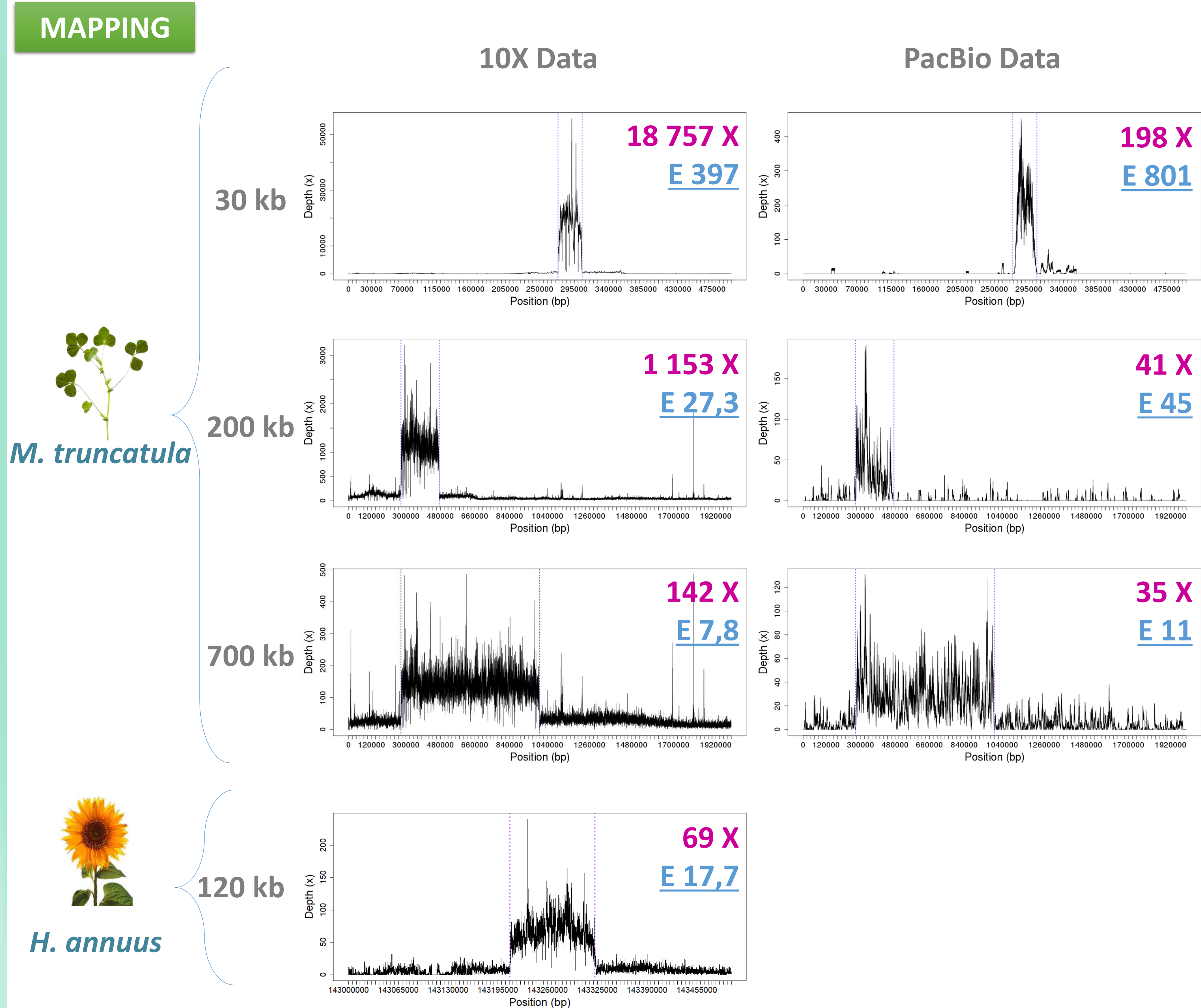


Figure 3. Cell nuclei are isolated from leaves and embedded in agarose plugs. Cell nuclei are lysed to release DNA. Genomic DNA was digested using RNA-guided Cas9 and separated by PFGE. The band corresponding to the target DNA is excised and the DNA is extracted from the gel. The target DNA is analyzed by using long read technologies.

## 5- Sequencing Analysis Of Genomic Region Of Interest From Model And Crop Plants



## ASSEMBLING

Sample ID	Total length (bp)	Contig number	% Identity
<i>M. truncatula</i> _ 30 kb	34 865	1	99
<i>M. truncatula</i> _ 200 kb		Work in progress	
<i>M. truncatula</i> _ 700 kb		Work in progress	

Figure 4. Sequencing analysis of the enriched target regions. Pacbio reads were mapped with minimap2 and 10X reads were mapped with Long Ranger align against the sequence of the v4.0 *Medicago truncatula* genome. The dash lines symbolize the sgRNA positions, the purple numbers represent the mean depth sequencing and the blue numbers indicate the enrichment for the region of interest. Based on these results, we captured genomic regions up to 700 kb from *M. truncatula* genome and a 120 kb genomic region for the complex plant genome, *H. annuus*. For the Pacbio 30kb sample, reads were overlapped to themselves using minimap2 then processed by home-made cleaning informatics tools and then assembled *de novo* with Canu in one contig spanning the 30kb targeted region.

## 6- Conclusion/Perspectives

### - Target large genomic region of interest from complex plant genomes

- Provide an accurate and reliable genomic information for the region of interest
- Allow a rapid comparison of a region of interest between two genotypes

### - Hybrid scaffolding with 10X data and PacBio data in progress to assemble genomic regions bigger than 30 kb

- Correction of PacBio reads with 10X data
- Assembly of corrected PacBio reads
- Hybrid scaffolding with 10X data

