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

Carine SATGÉ[1], Margaux-Alison FUSTIER[1], Caroline CALLOT[1], Stéphane CAUET[1], William MARANDE[1], Arnaud BELLEC[1], Pauline DURIEZ[2], Stéphane MUÑOS[2] and Sonia VAUTRIN[1*]

[1] *French Plant Genomic Center (CNRGV) – INRA, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France*

[2] *Laboratory of Plant Microbe Interactions INRA/CNRS, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France*

** To whom correspondence should be addressed. Email: sonia.vautrin@inrae.fr*

Abstract

To accelerate plant breeding, **high quality assemblies are essential**. They help to understand genome structure and to identify genes involved in agronomic traits. However, the exploration of plant genomes remains challenging. Very high genome sizes, repetitive element contents and the polyploidy nature of some plants (especially many of the crop species) make **plant genomes very complex**. 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 investigate the potential of the **CRISPR/Cas9 system to target a 110 kbp genomic region** involved in a plant parasitic resistance[1] from the **sunflower *Helianthus annuus* genome**. We improved and adapted the first steps of the **CATCH method**[2] (Cas9-Assisted Targeting of CHromosomal segments). Then, we sequenced the targeted region with two PacBio long reads sequencing protocols (with or without amplification). The PacBio 'Low-Input Protocol' allowed the enrichment of the genomic region with **high quality assembly**. Thus, we propose a CRISPR/Cas9 based method **amplification-free**, with a simplified bio-informatical pipeline and a potential for **multiplexing** to sequence large genomic region of interest from plant genomes.

The CATCH Method To Target A Genomic Region Of Interest

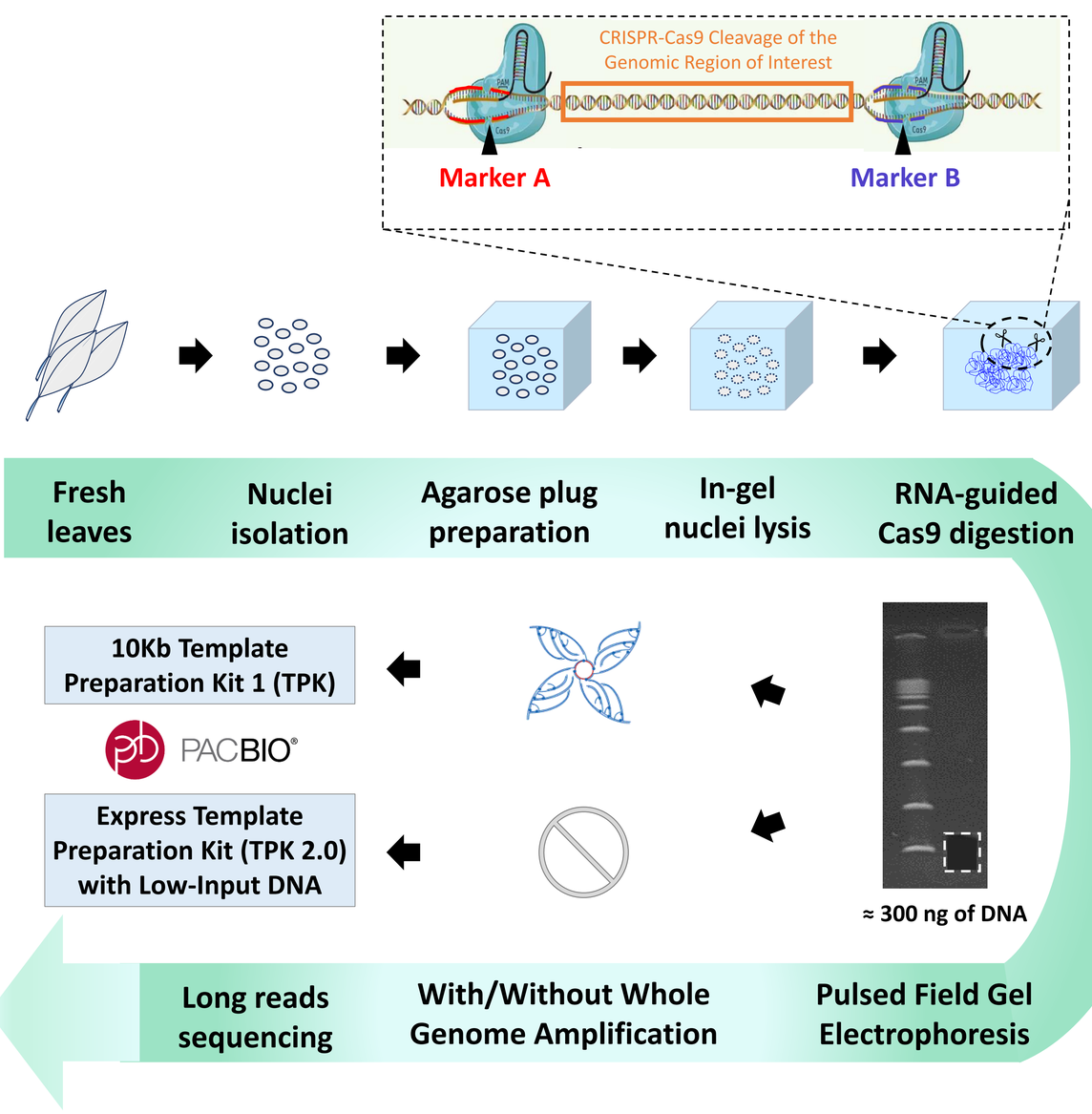

Figure 1. The CATCH Method Workflow. Cell nuclei are isolated from leaves, embedded in agarose plugs and lysed to release DNA. Genomic DNA is then digested using RNA-guided Cas9 and separated by PFGE. The band corresponding to the target DNA is excised and the DNA is extracted from the gel[3]. Two library preparation protocols are tested: either the input DNA is amplified by whole genome amplification and the library is prepared with the 10kb Template Preparation Kit 1[4] or the input DNA is not amplified and the Express Template Preparation Kit 2.0 with Low-Input DNA[5] is used to perform the library.

Library Preparation Workflow

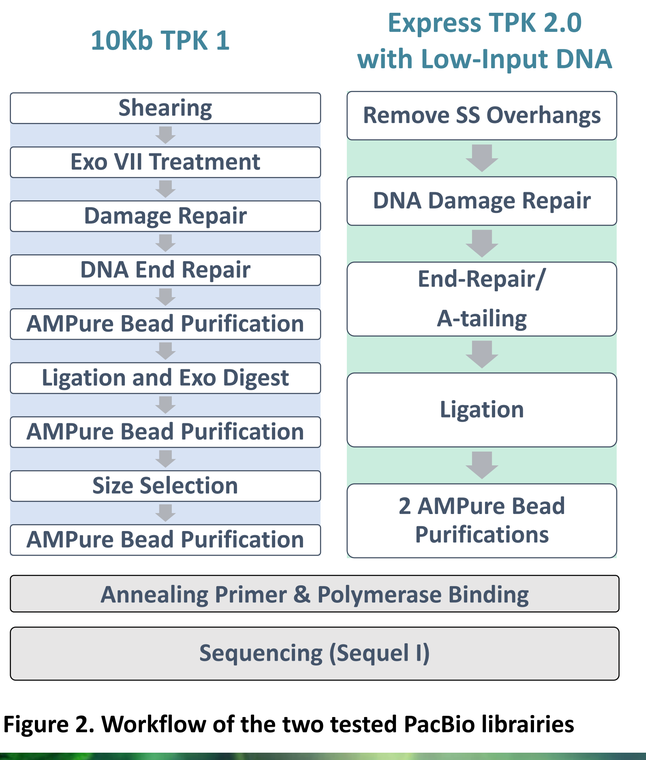

Figure 2. Workflow of the two tested PacBio librairies

10Kb TPK 1 Library Preparation requirements:

- **3-5 µg** of starting **sheared** genomic DNA
- Shearing method: **Megaruptor 2**
- **Sizing step: BluePippin**

Express TPK 2.0 with Low-Input DNA Library Preparation requirements:

- Required minimum gDNA > **150 ng**
- Input gDNA quality required: Majority of gDNA > 20 kb
- gDNA shearing/sizing method: **not required**

Sequencing A 110 kbp Genomic Region From Sunflower Genome

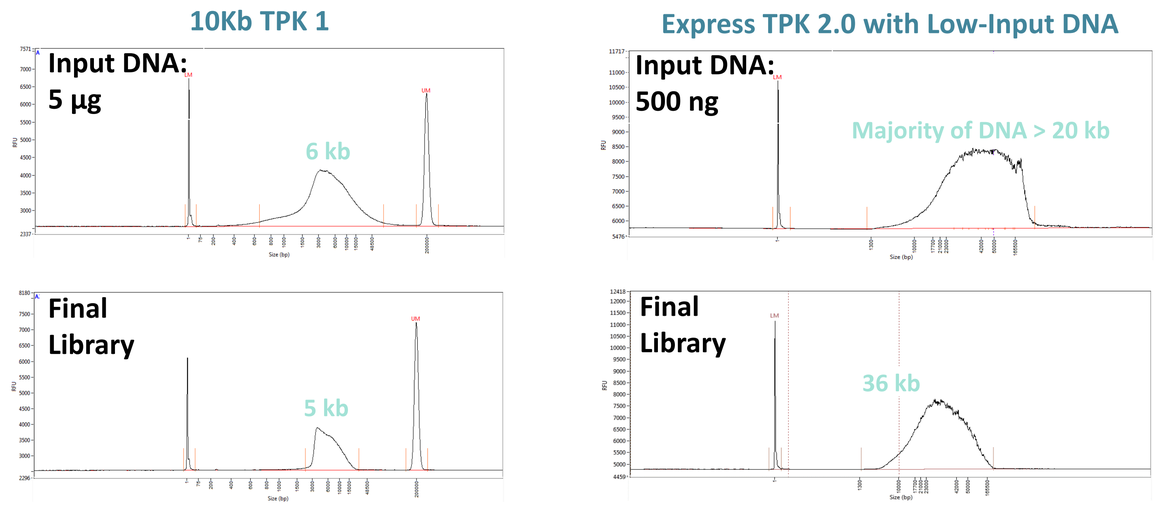

Figure 3. Fragment Analyser (left) and Femto Pulse (right) QC of *Helianthus annuus* input DNA and resulting libraries with 10Kb TPK 1 or Express TPK 2.0 with Low-Input DNA. Numbers indicate the mean fragment length.

Library Prep Protocol	Total Yield (Gb)	Unique Mol. Yield (Gb)	N50 Polymerase Read Length	N50 Subread Length	P0 %	P1 %	P2 %	Chimeric Reads (%)
10Kb TPK 1	23.4	NA	33,729	5,761	10.3	70.7	19.0	80
Express TPK 2.0 with Low-Input DNA	2.7	1.8	23,254	6,789	62.9	26.2	10.9	<1

Table 1. Run statistics for the Sequel I SMRT Cell run (v6.0 chemistry, 10 hours run).

10kb TPK 1 - Before cleaning		10kb TPK 1 - After correcting	
Number of reads	Average length (b)	Number of reads	Average length (b)
3,977,966	**5,894**	25,695,887	**845**

Table 2. Effect of correcting palindromes from the data obtained with whole genome amplification coupled to 10Kb TPK library. To remove chimeric reads, 10Kb TPK 1 filtered reads were mapped against their corresponding reference (v2 *H.annuus XRQ* genome sequence) with minimap2 (option: -cx map-pb –secondary=n) to generate PAF format output. This step allows the splitting of chimeric reads. Non-chimeric corresponding fastq sequences were generated using homemade python script. → **Correcting palindromes decreases read length.**

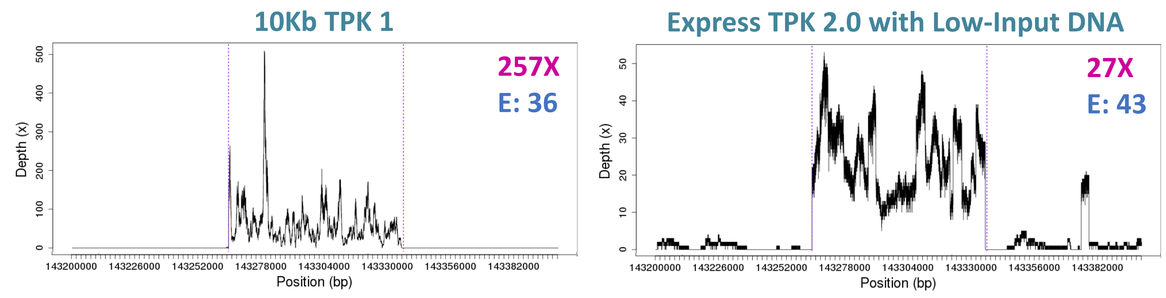

Figure 4. Mapping PacBio reads against the *H. annuus XRQ v2* reference genome. Pacbio CLR reads were mapped with minimap2 and align against the reference genome. The dash lines symbolize the sgRNA positions, the purple numbers represent the mean depth sequencing and the blue numbers indicate the enrichment for the region of interest. → **The Expressed TPK 2.0 with Low-Input DNA allows a total coverage of the targeted genomic region.**

	Total lenght (bp)	Contig number	% Identity
10Kb TPK 1	Assembly failed		
Express TPK 2.0 Low-Input DNA	110,876	1	99.8

Table 3. Assembly statistics of the PacBio targeted genomic region from *H. annuus XRQ.* Assembly was run with Canu 1.9 (default parameters). → **The Expressed TPK 2.0 with Low-Input DNA enables assembling of the 110 kbp genomic region of interest.**

Conclusion/Perspectives

- **CATCH method coupled to Express TPK 2.0 with Low-Input DNA library preparation is one solution to target large genomic region of interest from complex plant genomes**
  - → Need **Low-input amount DNA** (500 ng for a 3.6 Gb genome)
  - → Provide an **accurate and reliable genomic information** for the region of interest
  - → Allow a **rapid comparison** of a region of interest between several genotypes
- **Pooling multiple samples in a same PacBio SMRT Cell run in progress**

→ **To learn more about how to solve plant genome complexity, see our poster** *'PO0095: Combining Various Genomic Strategies With New Technologies to Decipher the Complex Structure of Plant Genomes'*

References

[1] P. Duriez et al., "A Receptor-like Kinase Enhances Sunflower Resistance to *Orobanche Cumana", Nat. Plants,* vol.5, pp.1211-1215, 2019

[2] W. Jiang *et al.,* "Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters," *Nat. Commun.*, vol. 6, pp. 1–8, 2015.

[3] N. Milon *et al.*, "µLAS technology for DNA isolation coupled to Cas9-assisted targeting for sequencing and assembly of a 30 kb region in plant genome," *Nucleic Acids Res.*, vol. 47, no. 15, pp. 8050–8060, 2019.

[4] https://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-10-kb-Template-Preparation-and-Sequencing.pdf

[5] https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-SMRTbell-Libraries-Using-Express-Template-Prep-Kit-2.0-With-Low-DNA-Input.pdf