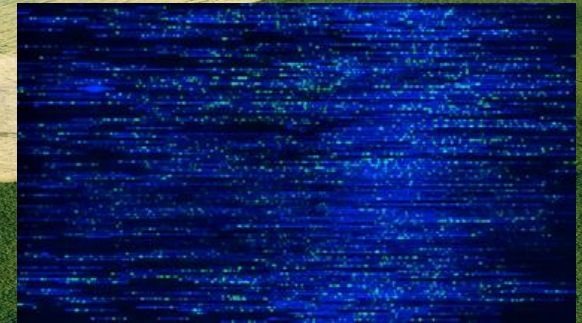


**Toward a better understanding of plant  
genomes structure :  
Combining NGS, optical mapping technology  
and CRISPR-CATCH approach**





INRA  
SCIENCE & IMPACT



PLANT GENOMIC CENTER

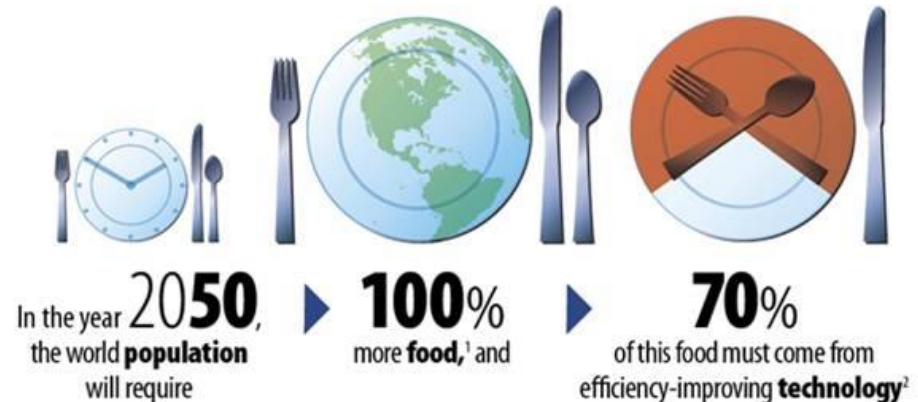
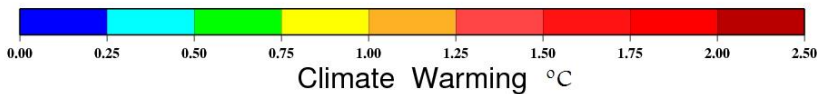
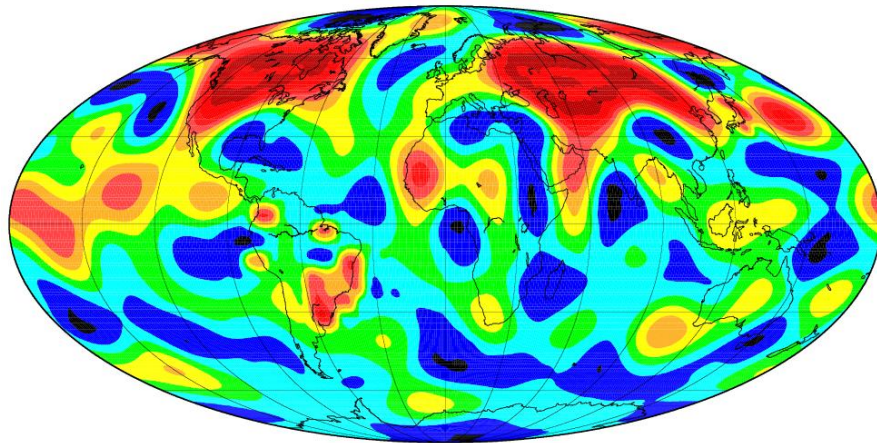


40 Plant species  
406 Genomic libraries  
20 Millions samples  
283 Collaborators

# Global warming effects, Population growth, Erosion of genetic progress, Consumer expectations

## INNOVATION :

Genome's exploration is one of the strategic approaches for better understanding plant evolution and plant improvement and adaptation



# The importance of a high quality reference genome sequence



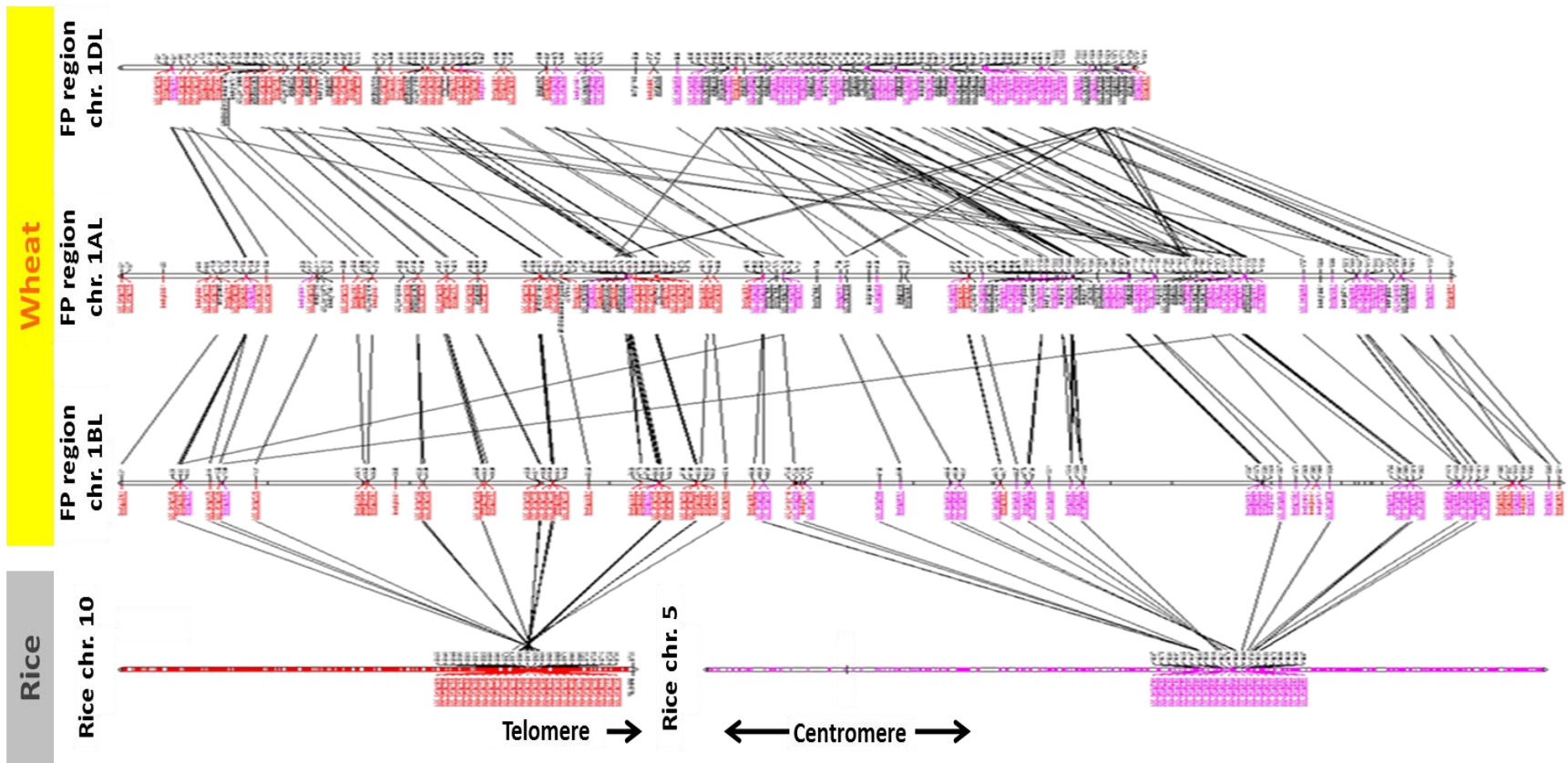
A. BELLEC



K. Eversole

Explore the genome at a large scale

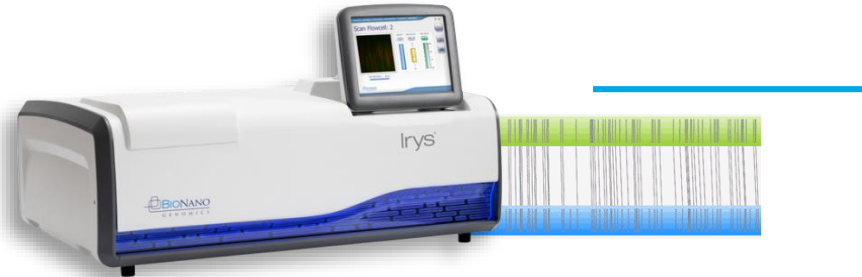
Project aiming at understanding plant genome evolution (Poster 40)



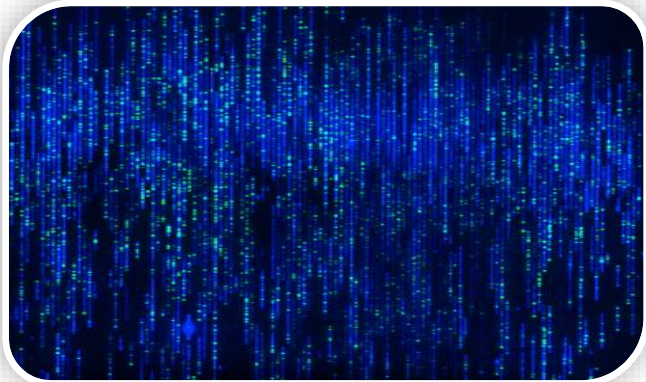
Structural and Evolutive Analysis of an Ancestral Chromosomes Fusion Point within the Hexaploid Wheat Genome

# A way to improve the quality for reference genome sequence : the optical maps

## The BioNanolrys system



- Direct visualization of long DNA molecules (>100 kb) working on non-amplified native genomic DNA
- Real physical distance information
- Labelling of specific sites (nickases)
- Molecular barcodes assembled



## Applications

### Whole Genome Finishing map = scaffolds

- Sequencing contigs converted *in silico* into molecular barcodes (highlighting the same sequence motifs)
- Sequencing based barcodes aligned to the BioNano maps

Targeting a specific genomic region / Comparison to a reference genome, looking for changes in the patterns:

- reveal insertion, deletion, inversion, translocation of genome segments

# Improvement of the Sunflower Genome sequence

- Species: *Helianthus annuus*
- 3.6 Gb
- 2n=34 chromosomes
- Genome sequence >100X PacBio (XRQ genotype)



N. Langlade

# contigs	LEN Max	N50 BP	#>N50	MEDIAN	BP
12 318	3,35 Mb	524 kb	1 684	120 kb	2,93

**=> 80% of the genome inside contigs**

Gouzy et al., 2016

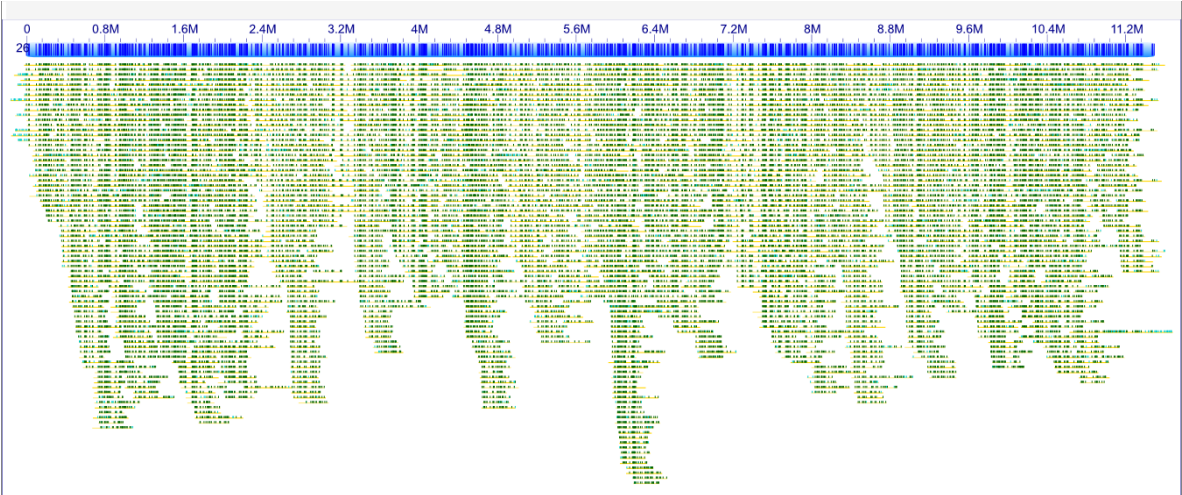
**Two major repeats in the sunflower genome: 8 kb and 11.5 kb**

# Optical maps to improve the genome assembly

	BspQ1	Bsss1
	5'...GCTCTTCN <sup>v</sup> ...3' 3'...CGAGAAAGN...5'	5'...CACGAG...3' 3'...GTGCTC...5'
Theoretical nb labels / 100kb	7,2	17,2
Real nb labels / 100kb	6,4	12,8
Raw data (Gb)	846 (235X)	845 (235X)
Filtered data >100kb (Gb)	635 (176X)	600 (167X)
Molecules N50 (kb)	206	187

**2 nicking enzymes (BspQ1 & BssS1)**

BioNano map



➤ **176X coverage, molecules from 150kb to 2,3Mb**

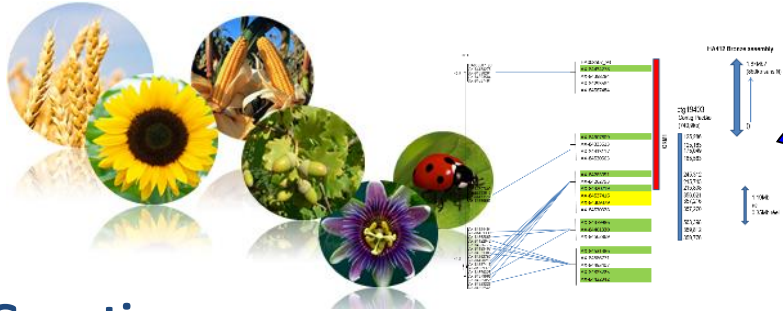
# The 2-steps hybrid scaffolding strategy improves significantly the resulting N50

	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				24.670
Total length (Mb)	2930	3191	2922	3112	2960
% genome	81%	88%	81%	86%	82%

More than 7 fold increase



# Dedicated genomic tools to better understand the role of regions of interest



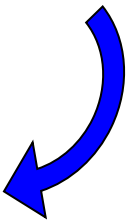
1. Optical map
2. BAC library from various genotypes
3. Sequence Capture

- Genetic map
- Physical map established on other genotypes
- Specific markers available in the region of interest



Sequencing (NGS)  
Comparison

- Physical characterisation of regions of interest (MTP)
- Isolation of the region of interest
- Identification of the region
- Comparison with reference map



# Focusing on a genomic region of interest in Sunflower



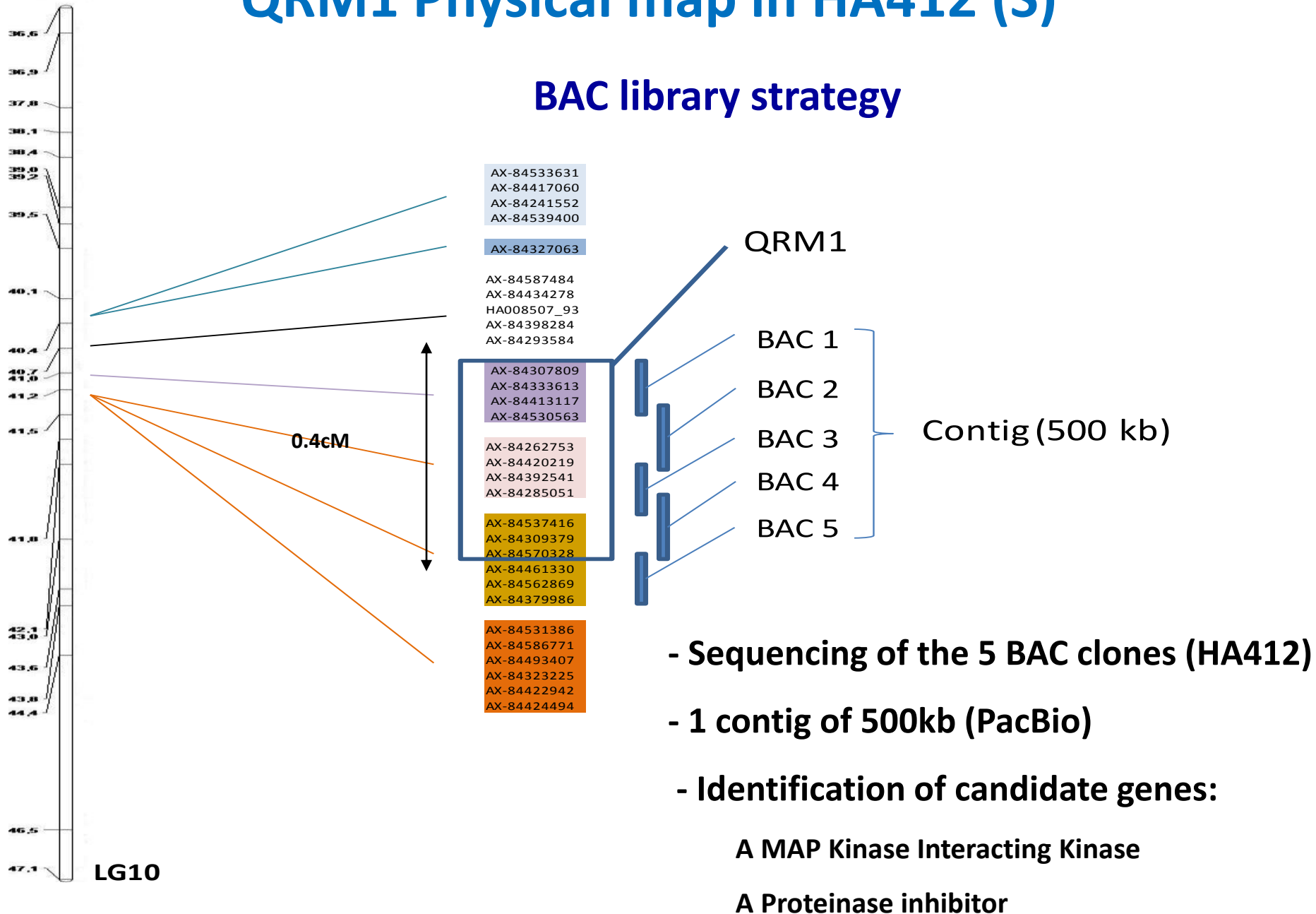
S. Munos



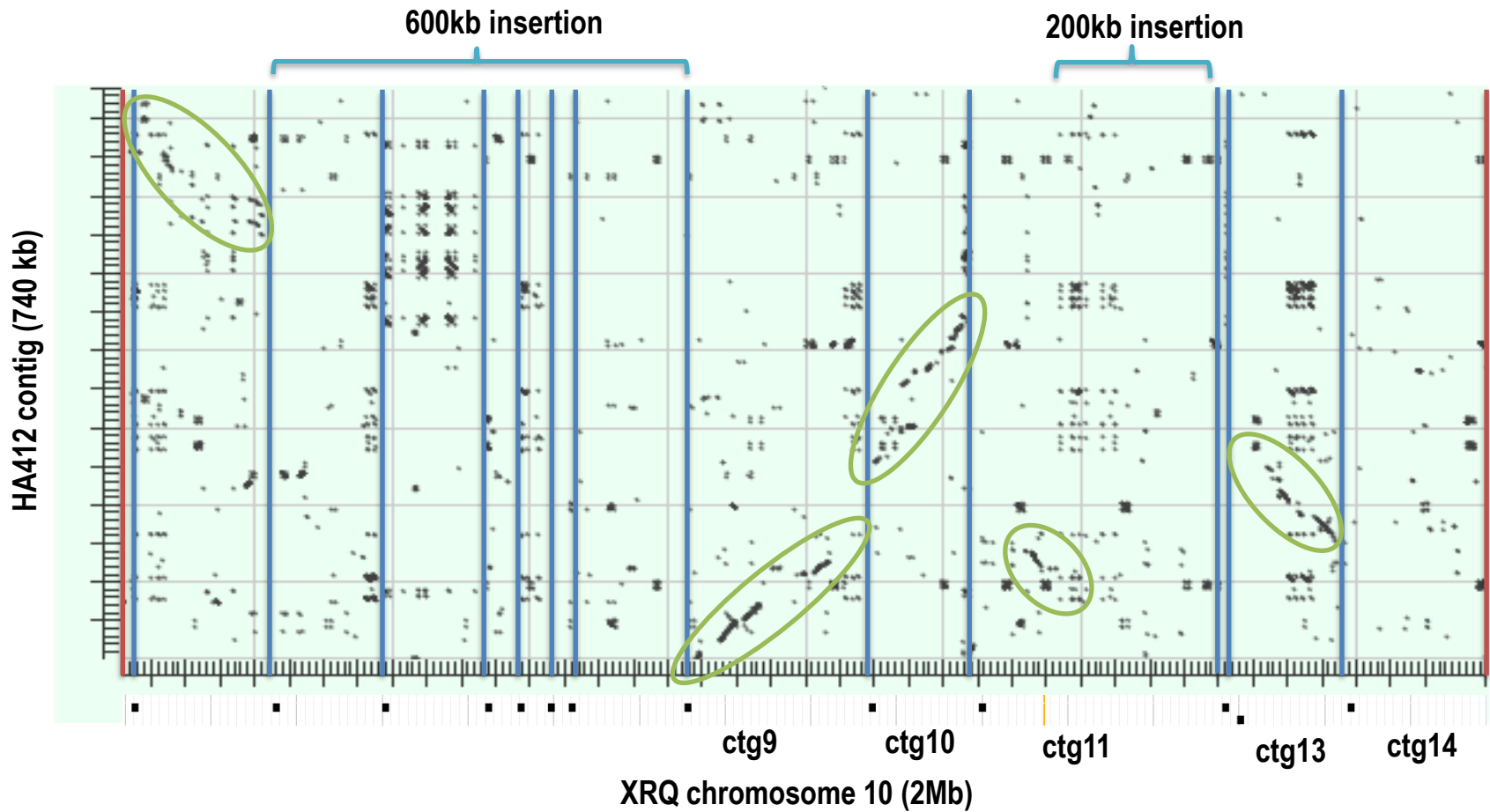
S. Vautrin

- **QRM1 controls quantitative resistance to downy mildew**  
**Susceptible (HA412) /Resistant (XRQ)**
- **Establishment of a genetic map (0.4 cM window on LG10)**
- **Markers definition on the QMR1 locus**
- **XRQ : *in silico* analysis of the 2Mb sequence on chromosome 10 (based on 20 markers alignment) composed of 14 scaffolded Pacbio contigs separating by N gaps (10k missing nucleotides)**

# QRM1 Physical map in HA412 (S)



# Comparison of the XRQ genome vs HA412 BAC clones



**Low collinearity**

**Fragmented alignment / orientation inconsistencies :**

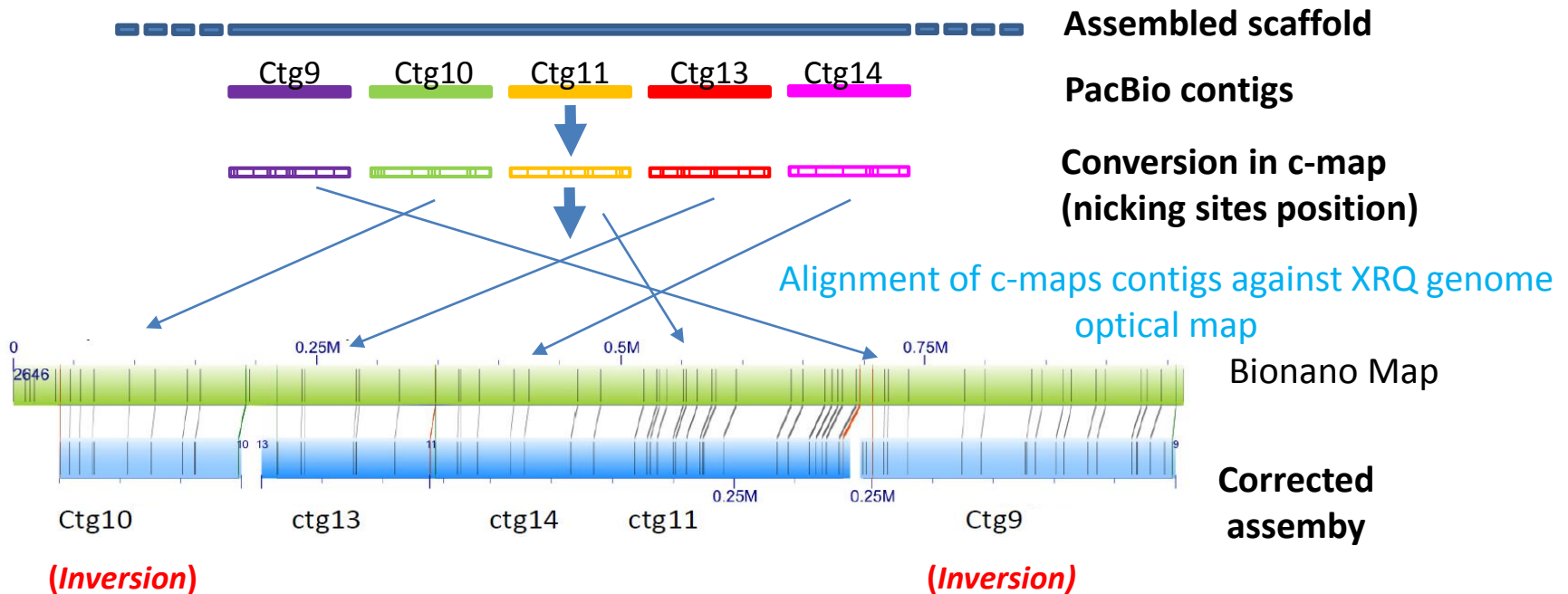
**Scaffolding errors OR true variability?**

# Optical maps to solve conflicts in the assembly



S. Cauet

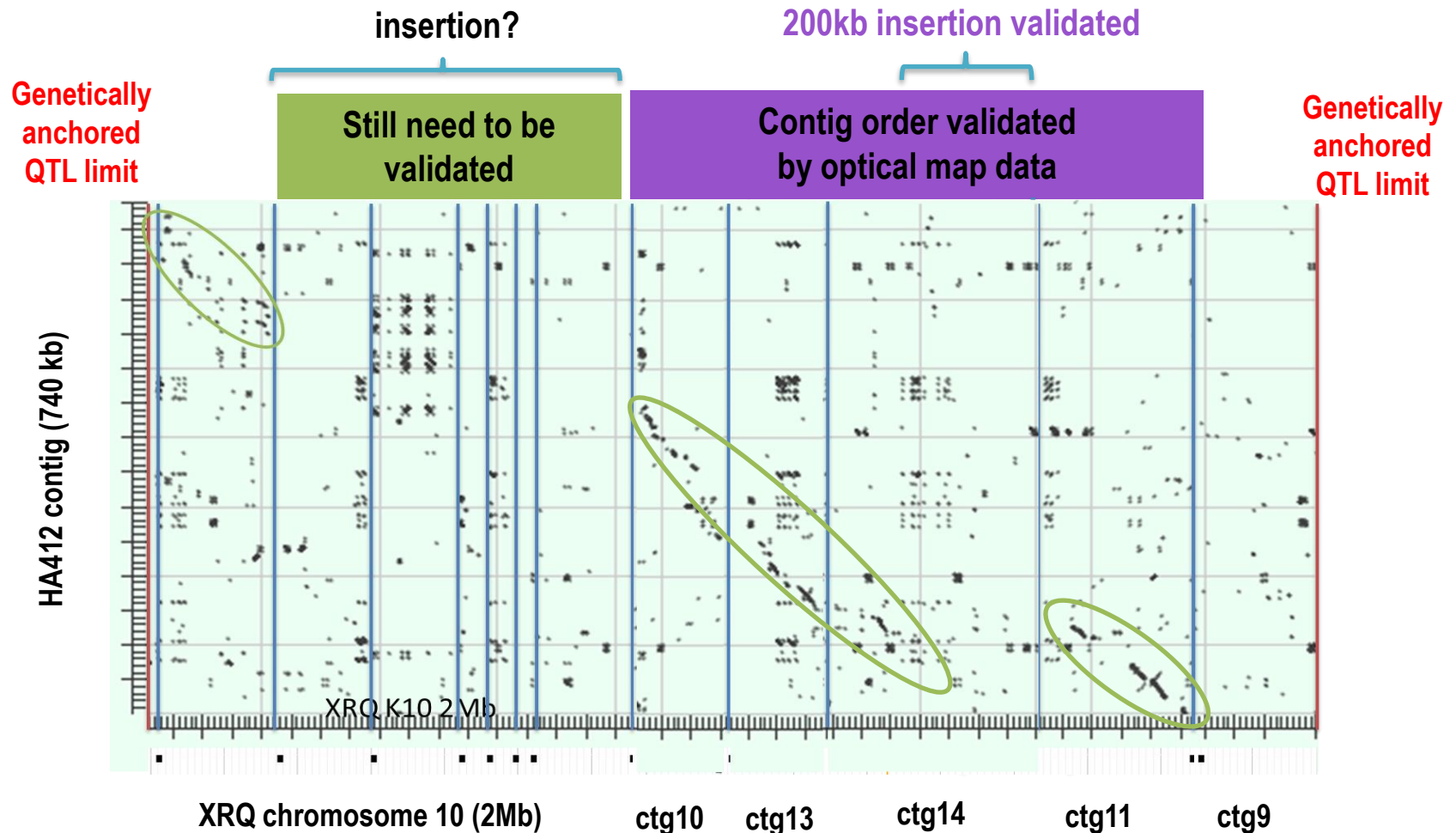
## Alignment of the contig against the BioNano assembly of XRQ genome



On this targeted region, Optical Bionano map allowed:

- to orientate some contigs
- to correct scaffolding of the PacBio contigs

# Comparison of the XRQ genome vs HA412 BAC clones



- Validation of the collinearity between XRQ and Ha412 sequences on QRM1 locus
- High variability observed: 2 major insertions of several hundreds of kb in XRQ
- Annotation of the 2 sequences and comparative analysis are under progress (9 candidates genes have been identified)

# Genome assembly improvement will help linking genotype / phenotype



N. RODDE Poster41

Sunflower proves again to be a highly complex genome, showing very high diversity between genotypes

**One reference genome is not enough!**

Despite long reads sequencing, assembly (scaffolding) has to be checked when working on reference genomes

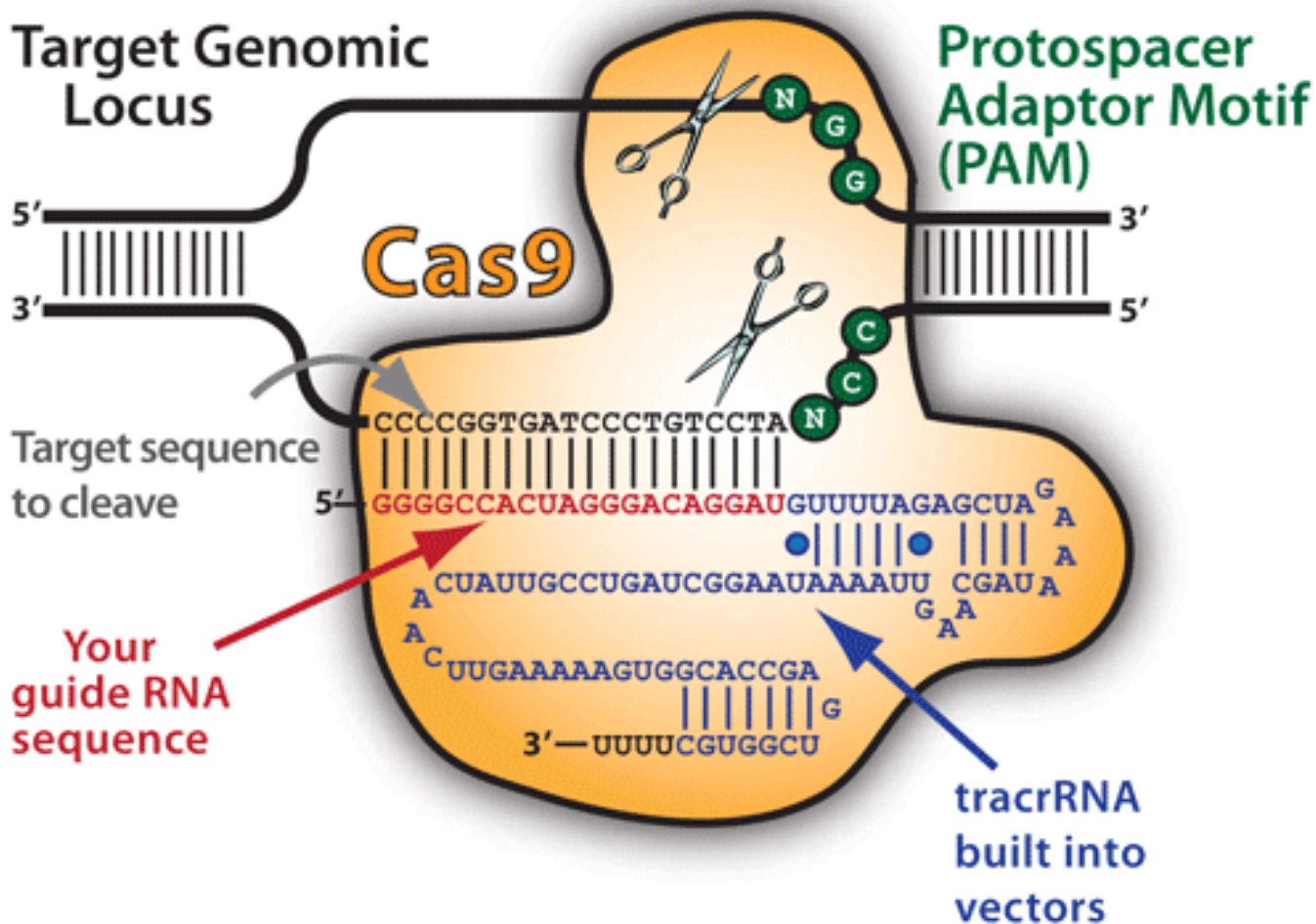
The optical map allowed to validate major rearrangements between the 2 genotypes

Proven interest of complementary approaches (NGS – optical map – BAC)

**=> How to be more efficient in focusing on genomic regions?**

# Another way to target a genome segment

## Based on the CRISPR-Cas9 technique



E. Charpentier /  
J. Doudna



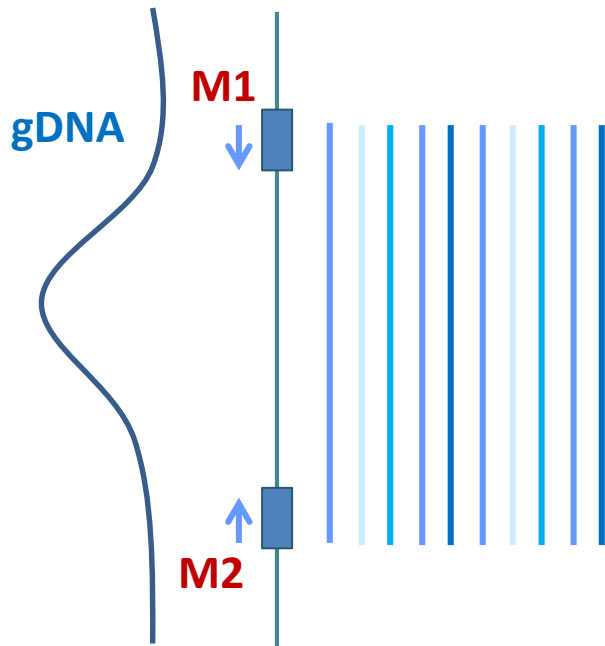
# CRISPR-CATCH: Cas9-Assisted Targeting of Chromosome segments



C. CHANTRY-DARMON

## QTL identification

- Guide RNA to target a specific region in the genome (**flanking markers of the QTL**)
- Cas 9 : nuclease DNA to unzip and cut the DNA at the chosen locus



### ARTICLE

Received 11 Feb 2015 | Accepted 16 Jul 2015 | Published 1 Sep 2015

DOI: 10.1038/ncomms9101

OPEN

Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters

Wenjun Jiang<sup>1</sup>, Xuejin Zhao<sup>2</sup>, Tslil Gabrieli<sup>3</sup>, Chunbo Lou<sup>2</sup>, Yuval Ebenstein<sup>3</sup> & Ting F. Zhu<sup>1</sup>

### PROTOCOL

Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments

Wenjun Jiang & Ting F. Zhu

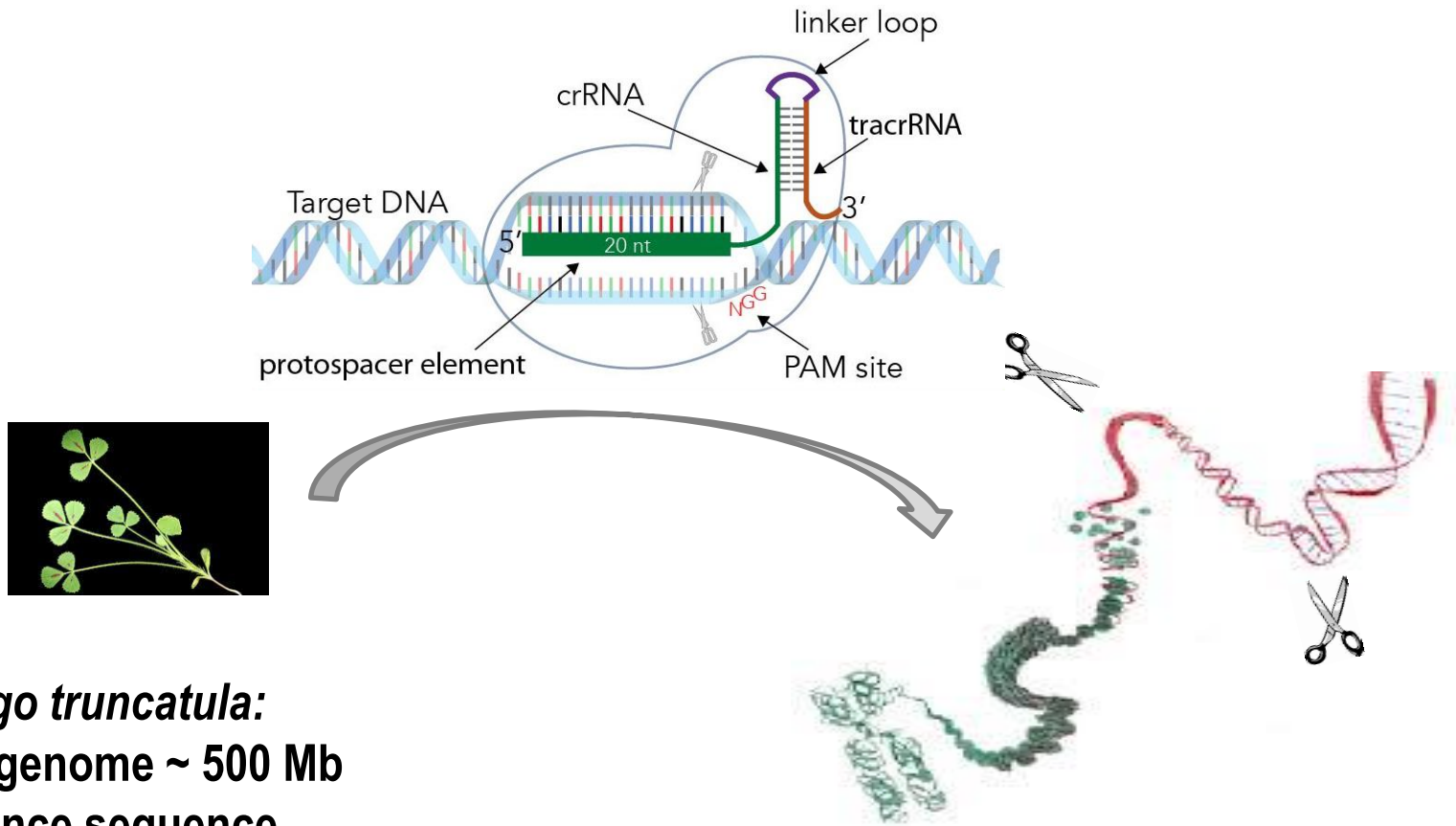
School of Life Sciences, Center for Synthetic and Systems Biology, Ministry of Education Key Laboratory of Bioinformatics, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Tsinghua University, Beijing, China. Correspondence should be addressed to W.J. (jiangwj12@mails.tsinghua.edu.cn) or T.F.Z. (tzhu@biomed.tsinghua.edu.cn).

Published online 21 April 2016; doi:10.1038/sprot.2016.055

➤ From a region of several kb to Mb

# Targeting genomic regions with CRISPR-CATCH

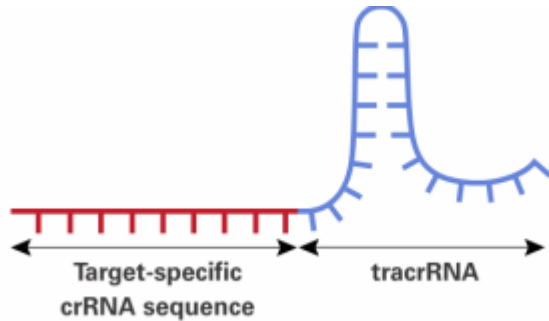
## Pilot test on *Medicago truncatula*



### *Medicago truncatula*:

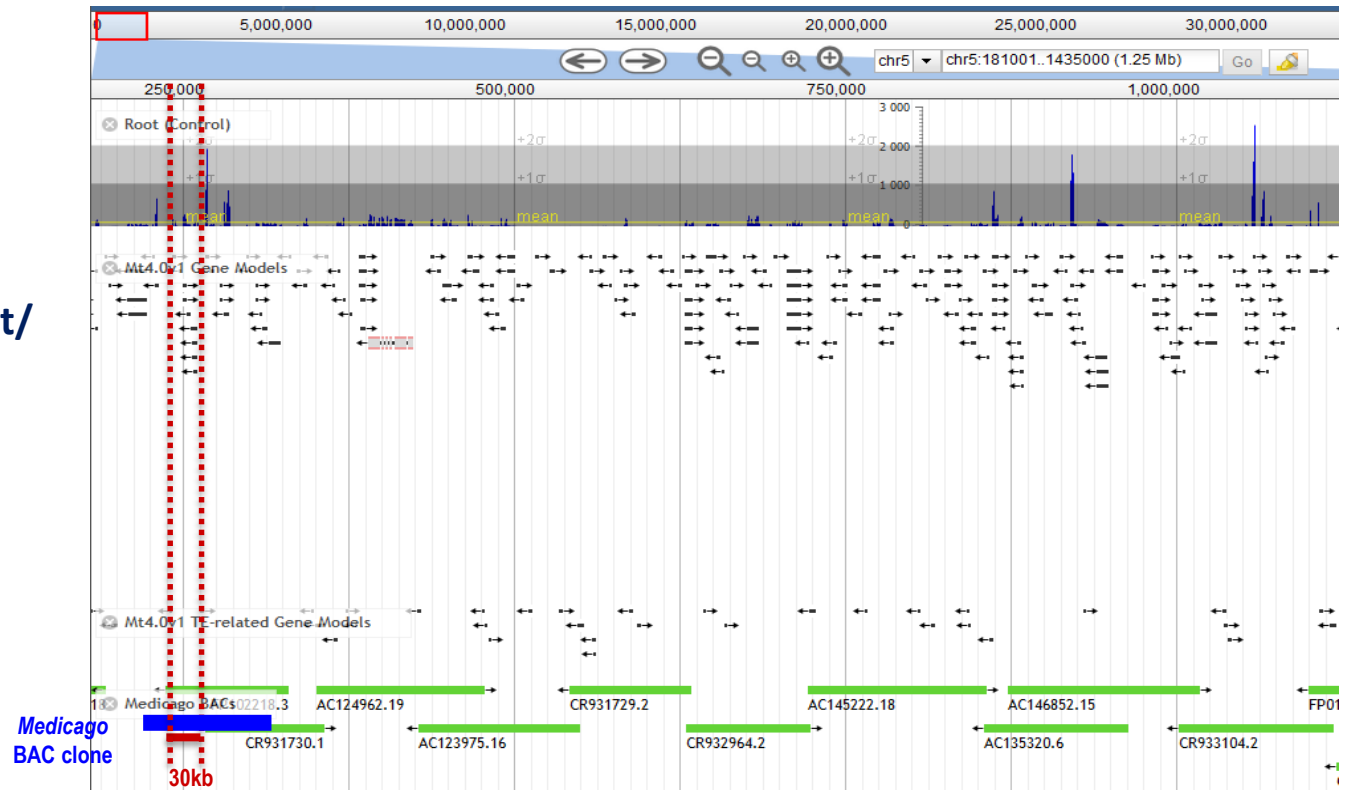
- Small genome ~ 500 Mb
- Reference sequence
- BAC libraries

# Key step : single guide RNA design

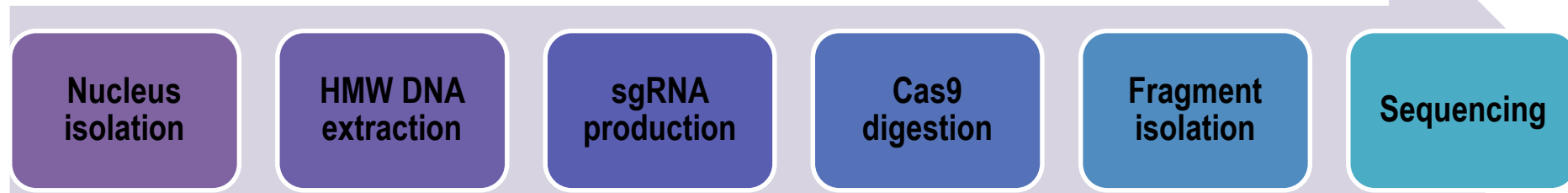
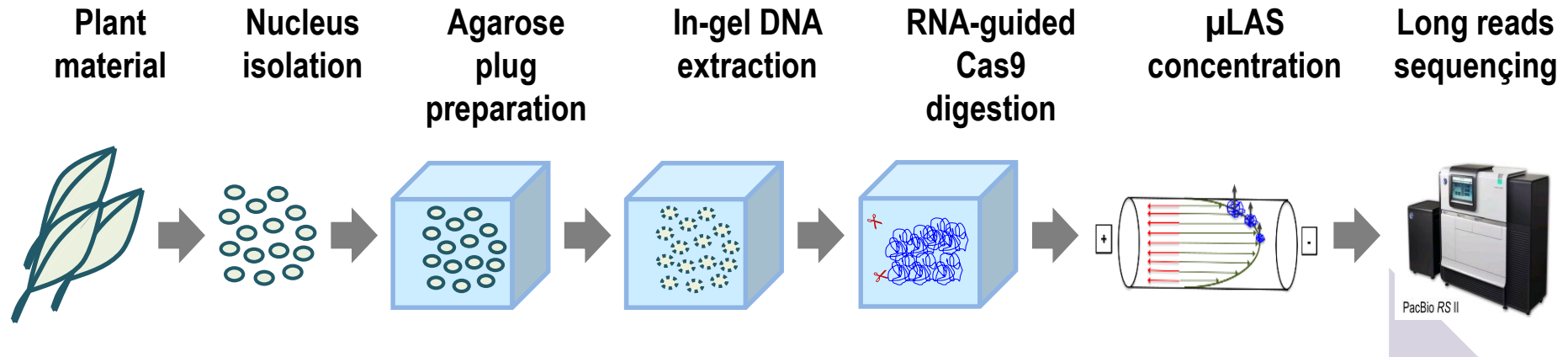


sgRNA contains a **targeting sequence (crRNA sequence)** and a **Cas9 nuclease-recruiting sequence (tracrRNA)**.

<http://crispor.tefor.net/>



# CRISPR-CATCH Workflow



How to recover the fragment of interest?

# Development of a new technology to enhance limit of DNA detection

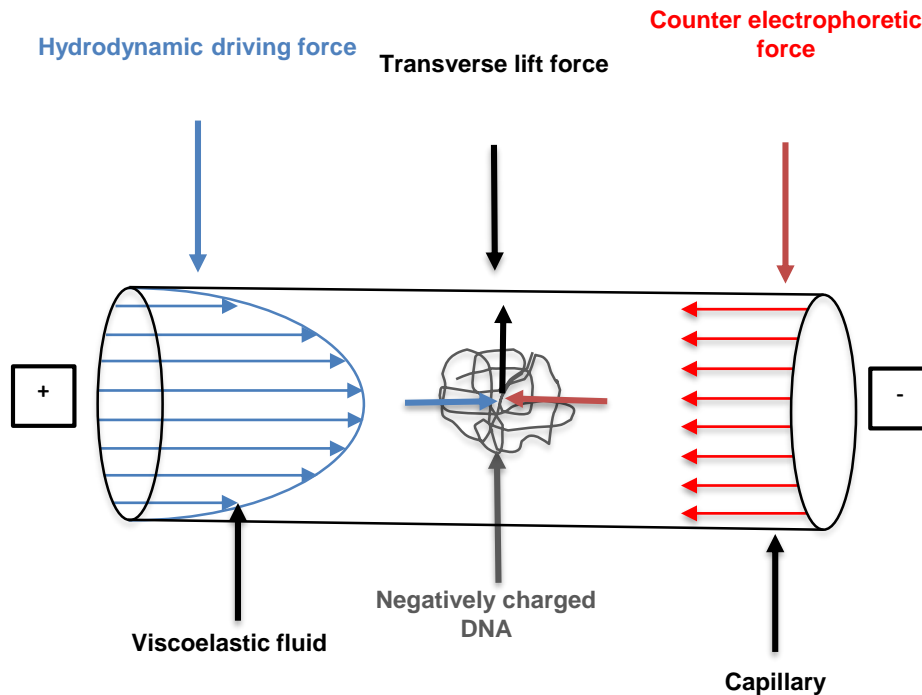
## $\mu$ LAS: separation of large DNA fragments

Microfluidic device with controlled force to avoid DNA fragmentation



A. BANCAUD

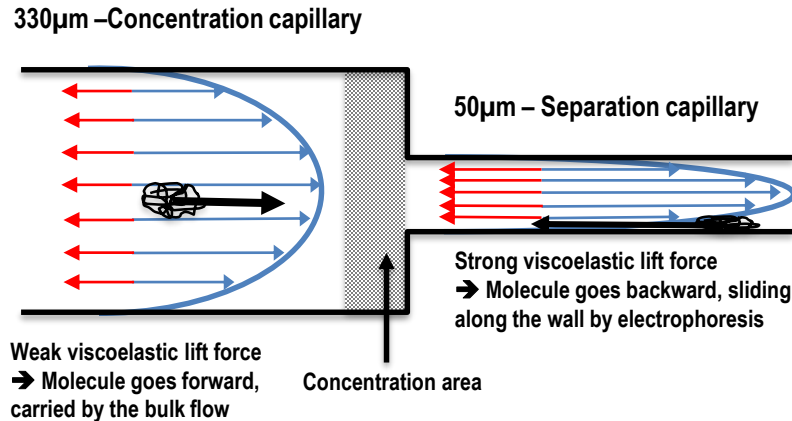
N. MILON



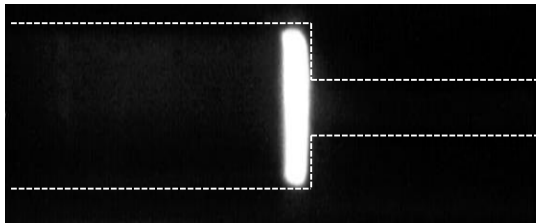
Counter-electrophoresis of DNA deforms fluid flow  
Force intensity depends on flow speed, shearing, electric field and DNA size  
Larger molecules are sent closer to the capillary wall where the flow speed is lower leading to a size separation mechanism

Enhance the performance of DNA analysis based on a polymer with specific viscosity

# $\mu$ LAS: Selective enrichment of large DNA fragments



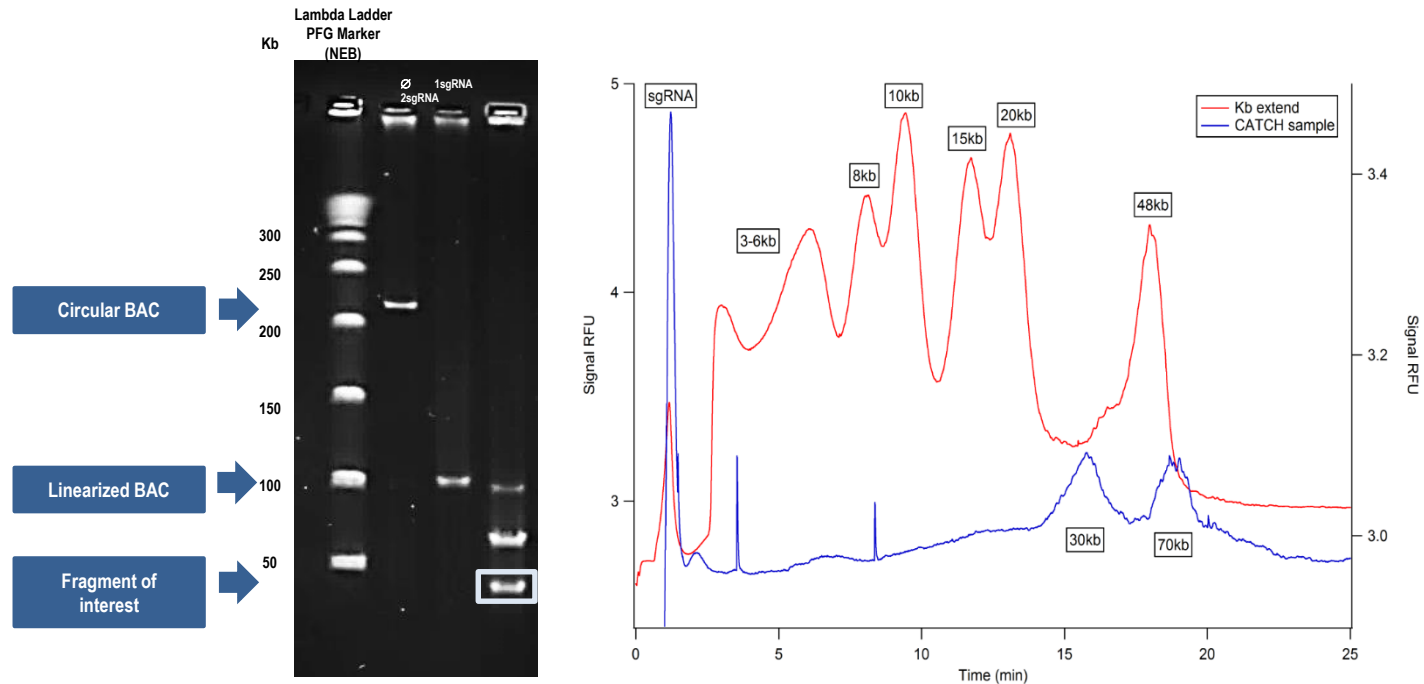
- Selection and concentration of DNA fragments > 8kb
- Removal of any RNA traces, small fragments, uncharged molecules
- Sensitivity of few pg/mL



DNA concentration at capillary junction



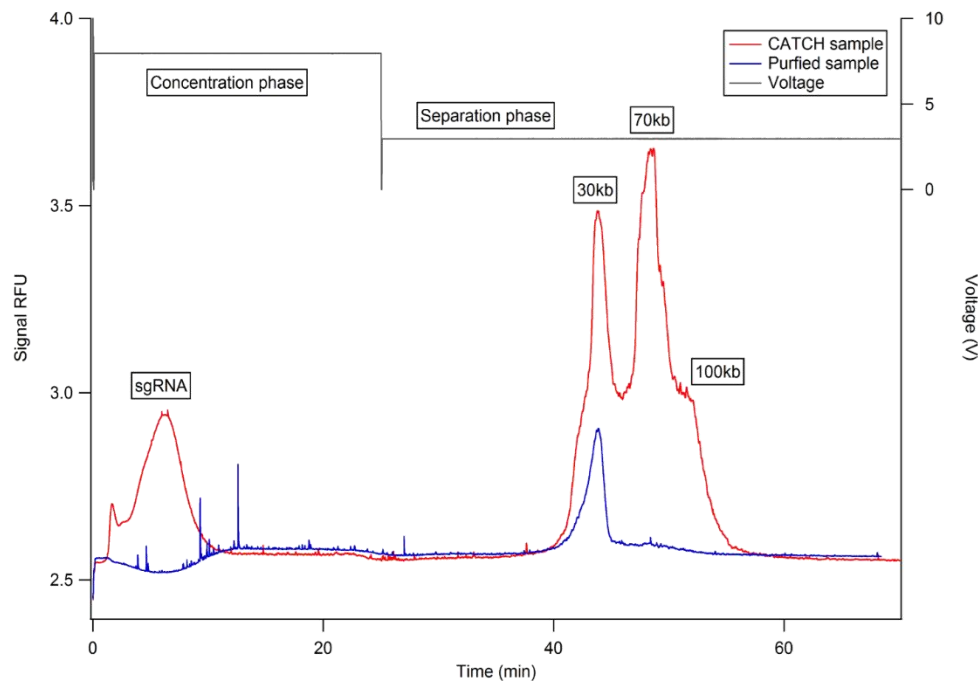
# Proof of concept with a *Medicago* BAC containing the region of interest



**Pulsed Field Gel  
Electrophoresis  
μLAS electrophoresis  
capillary system**

# Proof of concept with a *Medicago* BAC containing the region of interest

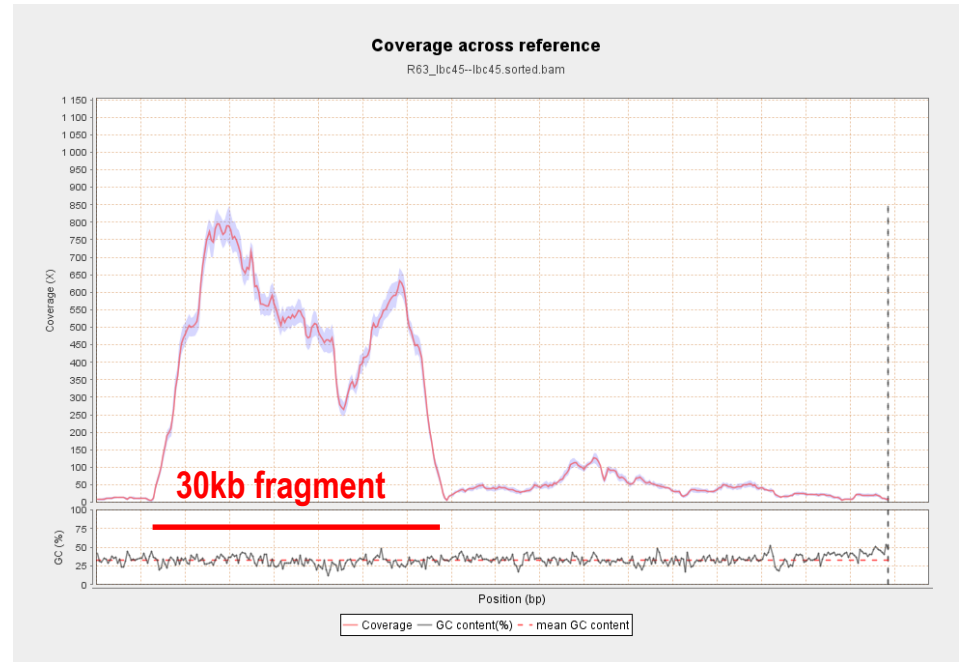
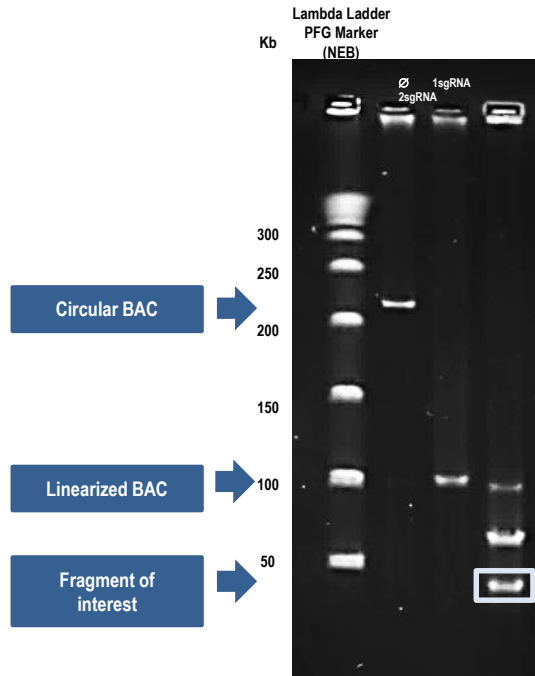
## Isolation and enrichment of the fragment of interest



- sgRNA, 70kb and 100kb removal
- Collection of 25 $\mu$ L at 150pg/ $\mu$ L
- Collection yield of 65% for the 30kb fragment



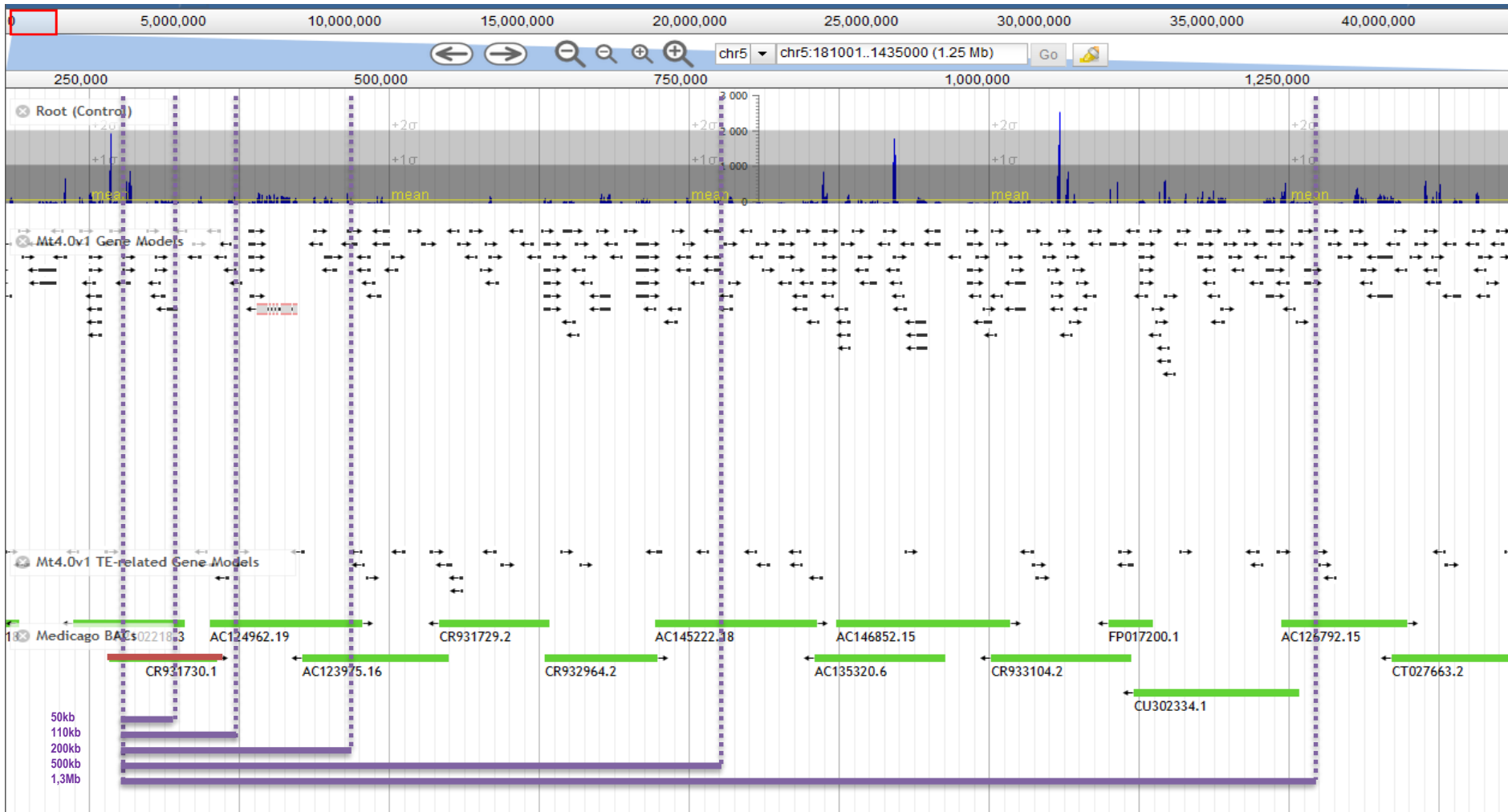
# Proof of concept with a *Medicago* BAC containing the region of interest



Phi29 DNA  
amplification

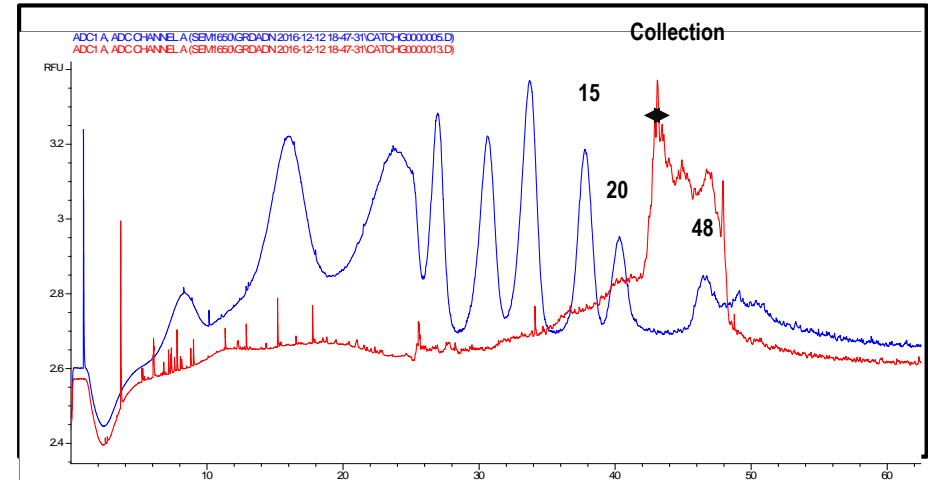
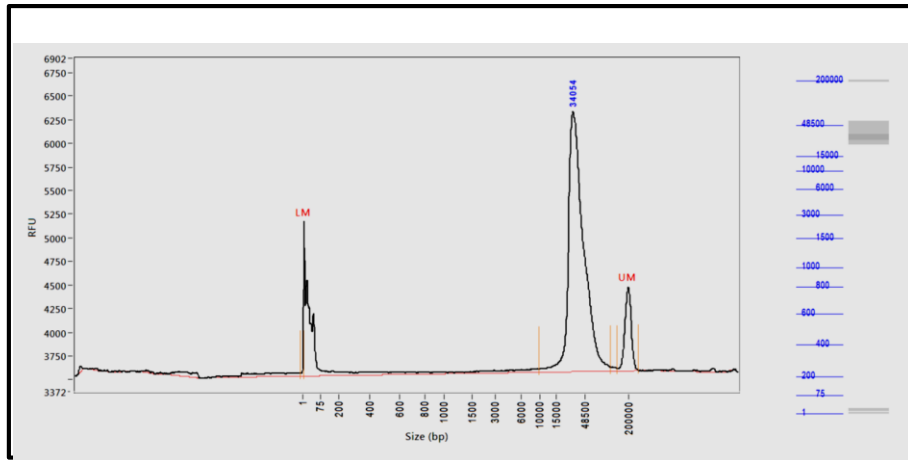
DNA sequencing

# Transfer of the method on gDNA



	sgRNAB1	sgRNAB2	sgRNAB3	sgRNAB4	sgRNAB5	sgRNAB6
Position	270 000	320 000	370 000	470 000	770 000	1 270 000
Theo fragment size	0	50 000	100 000	200 000	500 000	1 000 000
Position	270 000	320 000	380 000	470 000	770 000	1 300 000
Real fragment size	0	50 000	110 000	200 000	500 000	1 030 000

# Transfer of the method on gDNA



- Analysis of the sequences under progress
- Improvement of the HMW DNA extraction
- Improvement of the Cas9 digestion
- Improvement of the fragment isolation
- Tests on larger fragments under progress

**2-years postdoc position open!**

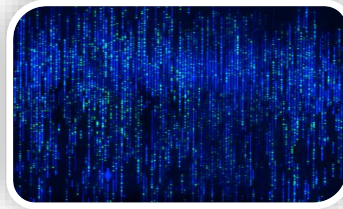
**<http://cnrgv.toulouse.inra.fr/News/Come-working-with-us-on-a-very-exciting-project>**

# Genomics to help agriculture facing challenges

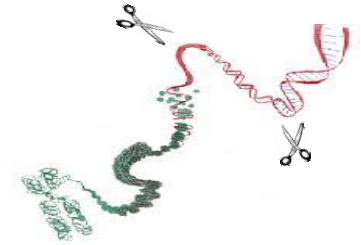
Integrated approaches to combine complementary technologies and tools



NGS  
Ref sequence  
genomes



Optical maps



BAC library  
Sequence Capture



**Make the world get a grain!**





# Acknowledgements



PLANT GENOMIC CENTER



@CNRGV  
@SUNRISE\_France

<http://cnrgv.toulouse.inra.fr/>

**Hélène BERGES**

Arnaud BELLEC

Sonia VAUTRIN

Céline CHANTRY-DARMON

Nathalie RODDE

Céline JEZIORSKI

William MARANDE

Stéphane CAUET

Nadège ARNAL

Caroline CALLOT

Joëlle FOURMENT

Nadine GAUTIER

Elisa PRAT

David PUJOL

Roseana RODRIGUES

Sandrine ARRIBAT

Laetitia HOARAU



A. BANCAUD

N. Milon

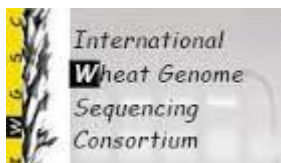
F. Ginot



John BAETEN

Kees-Jan FRANCOIJS

Jérôme GOUZY  
Nicolas LANGLADE  
Stéphane MUNOS



Kellye Eversole

