PLANT GENOMIC CENTER

Low coverage BAC libraries and BAC NGS sequencing : a powerful tool for comparative genomics of complex plant genomes

Genséric Beydon[1], Nathalie Rodde[1], Audrey Courtial[1], Elisa Prat[1], Joelle Fourment[1], Nadine Gautier[1], Nadege Arnal[1], Justine Mas[1], William Marande[1], Stephane Cauet[1], Arnaud Bellec[1], Sonia Vautrin[1], Clémentine Vitte[2], Johann Joets[2] and Hélène Bergès[1]

(1)INRA - CNRGV, Castanet Tolosan, France, (2)INRA/CNRS/Univ Paris-Sud/AgroParisTech, Gif-sur-Yvette, France

The French Plant Genomic Resource Center (CNRGV) is unique in both France and Europe. It is responsible for preserving and maintaining plant genomic resources generated by research. It already supplies laboratories throughout the world with genomic resources, and related tools. The CNRGV interacts with laboratories around the world as services provider or through collaboration. The CNRGV is a reliable partner to assist your genomic projects.

Plant Genome Complexity and Diversity

Arabidopsis : 130 Mb

Tomato:1Gb
75 % of TE

Maize : 2.5 Gb
80% of TE

Very Large genome Size

Barley : 5 Gb

Drososophila : 120 Mb

Oak : 1 Gb

Pepper : 2.7 Gb

E. Coli 4.5 Mb

Grappewine : 475 Mb

Passiflora : 1.35 Gb

Rye : 9.1 Gb

Wheat: 17 Gb
90 % of TE - 6X

Eucalyptus : 600 Mb

Oil palm : 1.8 Gb

Sunflower : 3.5 Gb

TE : Transposable Elements

Level of Polyploïdy

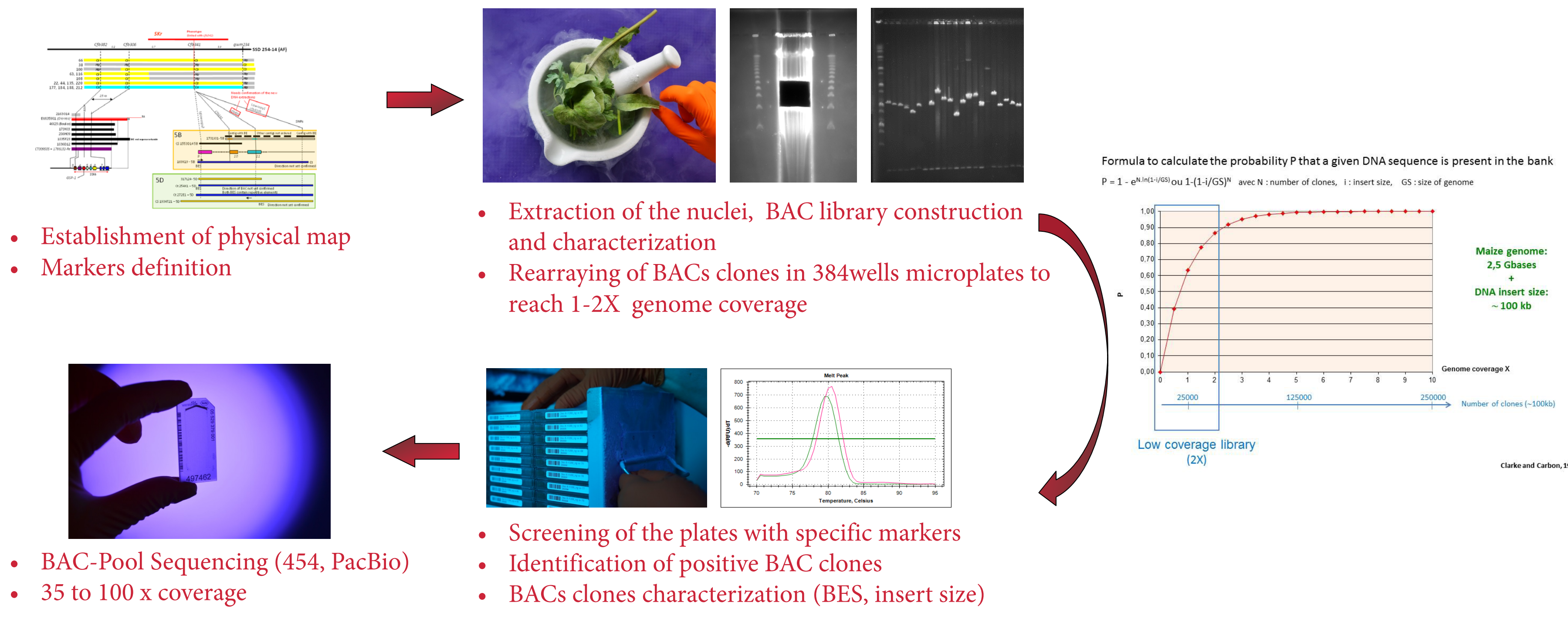Rape : 550 Mb

Strawberry : 1.9 Gb - 8 X

Human : 3 Gb

Sugarcane : 12 Gb -12 X

CNRGV develops, provides and maintains BAC libraries as essential tools to decipher plant genome complexity and explore diversity.

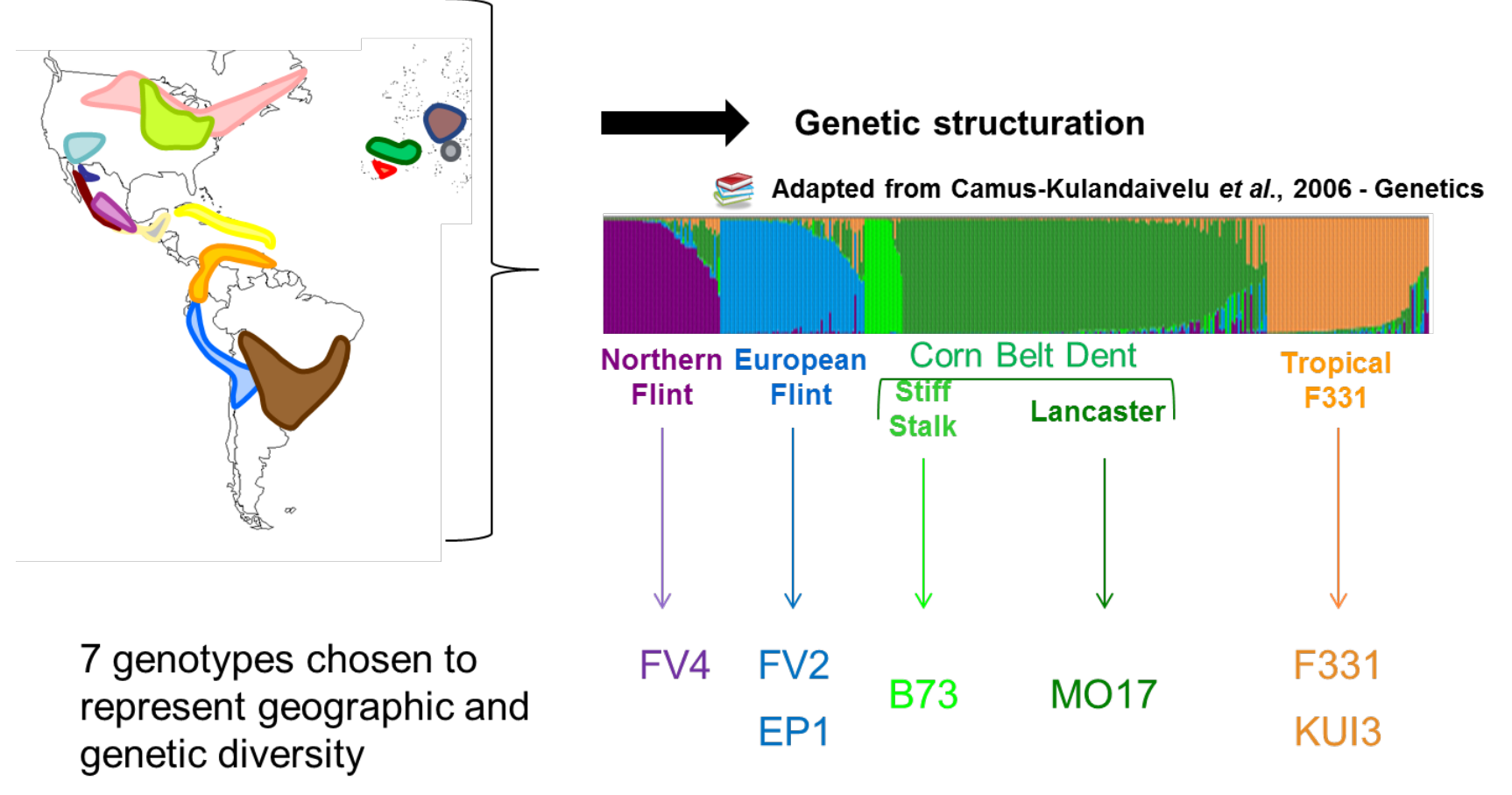## Low coverage BAC library strategy

Objective
Access to a BAC Library screening of various genotypes at the lowest cost / delay ratio by decreasing production and storage costs

Target projects
1 - Establishment of physical maps:
On one or few limited zones
In genotypes with no BAC libraries available (ex : resistant plant or genotype with a specific interest phenotype)
Thanks to sequences or marker available on a model genotype or model plant
2 - Filling of residual gaps on physical map obtained with classical BAC libraries
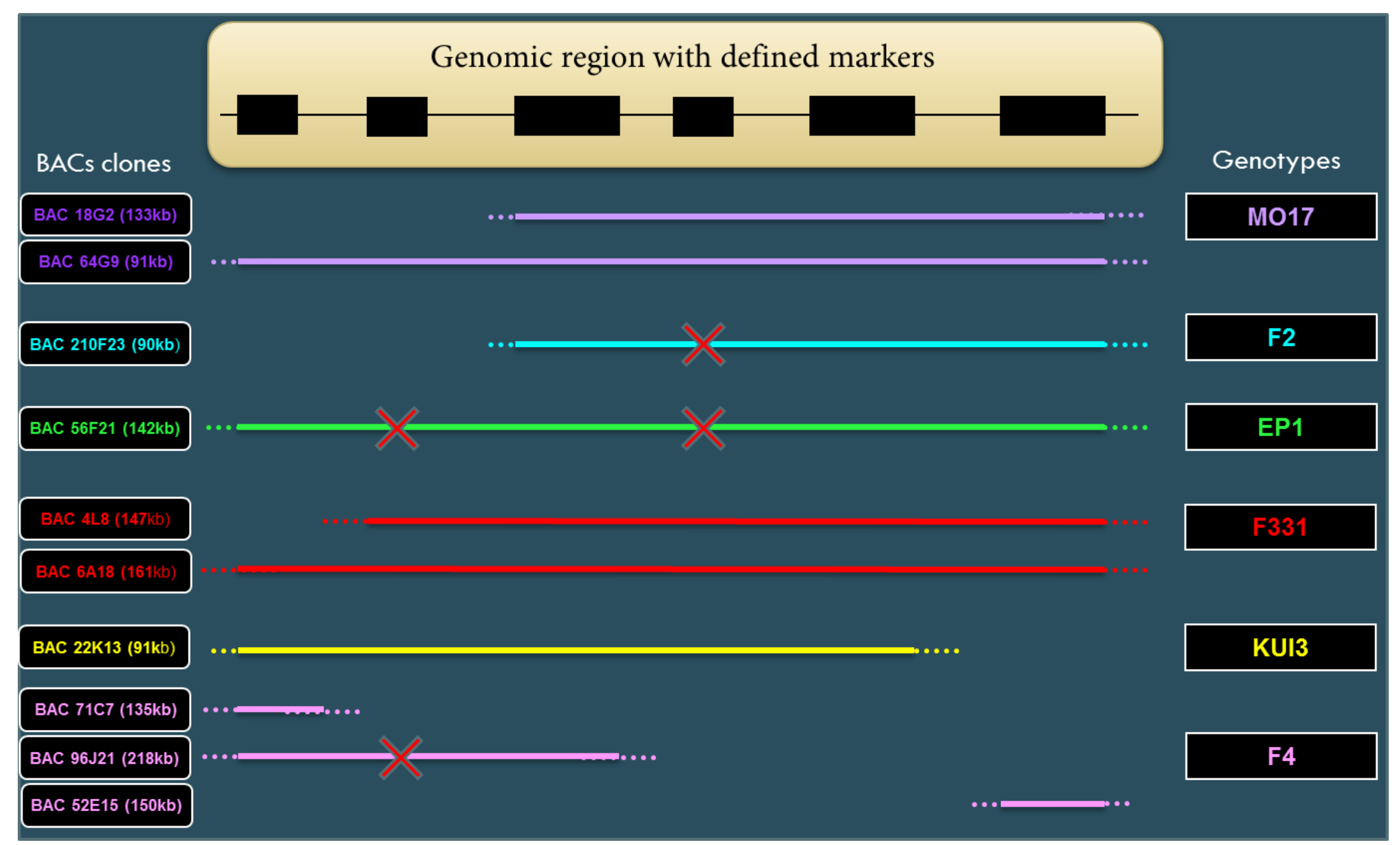3 – Sequencing of a target zone or genes in several different genotypes / species for syntenic analysis

- Establishment of physical map
- Markers definition

- Extraction of the nuclei, BAC library construction and characterization
- Rearraying of BACs clones in 384wells microplates to reach 1-2X genome coverage

Formula to calculate the probability P that a given DNA sequence is present in the bank

$$P = 1 - e^{N \ln(1 - I/GS)} \text{ ou } 1 - (1 - I/GS)^N$$ avec N : number of clones, I : insert size, GS : size of genome

Maize genome:
2,5 Gbases
+
DNA insert size:
~ 100 kb

Genome coverage X

Low coverage library (2X)

Number of clones (1-100Kb)

Clarke and Carbon, 1976

- BAC-Pool Sequencing (454, PacBio)
- 35 to 100 x coverage

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

Example : comparison of different genomic regions between 7 maize inbred lines

Selection of maize inbred lines

Genetic structuration

Adapted from *Camus-Kulandaivelu et al.,* 2006 - Genetics

Northern Flint | European Flint | Corn Belt Dent (Stiff Stalk / Lancaster) | Tropical F331

FV4 | FV2 EP1 | B73 | MO17 | F331 KUI3

7 genotypes chosen to represent geographic and genetic diversity

BACs libraries specifications

Z. mays line	FV2	MO17	EP1	F331	KUI3	F4
Enzyme	*Hind* III	*Hind* III	*Hind* III	*Hind* III	*Hind* III	*Hind* III
Vector	pBELOBAC11	pIndigoBAC5	pIndigoBAC5	pIndigoBAC5	pIndigoBAC5	pIndigoBAC5
E. coli cells strain	DH10B	DH10B-T1R	DH10B-T1R	DH10B-T1R	DH10B-T1R	DH10B-T1R
		phage resistant	phage resistant	phage resistant	phage resistant	phage resistant
Total BAC clones	84864	28800	28416	26112	31104	38400
Plates number	221	51	74	68	81	100
Average insert size (kb)	90	135	120	136	115	112
Genome coverage (X)	3	1	1	1	1	1

Genomic region with defined markers

BACs clones | Genotypes

BAC 18G2 (133kb) | MO17
BAC 64G9 (91kb) |

BAC 210F23 (90kb) | F2

BAC 56F21 (142kb) | EP1

BAC 4L8 (147kb) | F331
BAC 6A10 (161kb) |

BAC 22K13 (91kb) | KUI3

BAC 71C7 (135kb) | F4
BAC 96J21 (218kb) |
BAC 52E15 (150kb) |

The low coverage BAC library strategy is a powerful tool to investigate variability in specific genomic regions of interest with a set of well defined markers.

INRA SCIENCE & IMPACT

CNRGV
24 Chemin de Borde Rouge
Auzeville - CS 52627
31326 Castanet tolosan cedex
infocnrgv@toulouse.inra.fr

http://cnrgv.toulouse.inra.fr

@cnrgv