



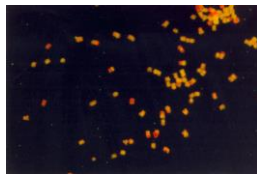
## A Reference Sequence of the Monoploid Genome of Sugarcane

Olivier GARSMEUR  
CIRAD, Montpellier, France

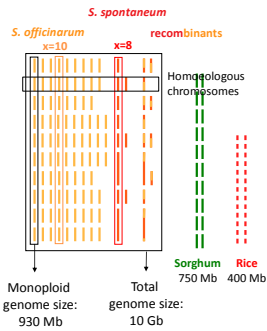


## Global organization of the complex genome of sugarcane

Genomic in situ hybridization



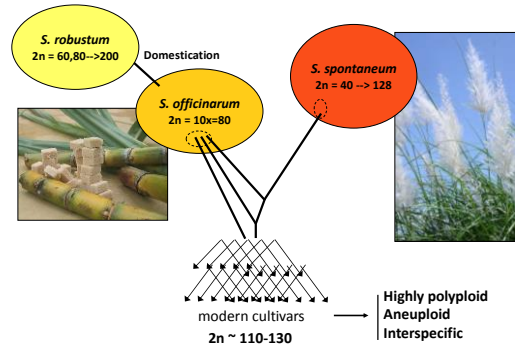
R570, 2n=ca 115  
80 % *S. officinarum*  
10% *S. spontaneum*  
10% recombinants



(D'Hont et al, 1996; D'Hont et al, 1998; D'Hont 2005, Piperidis et al 2010)

CIRAD, Garsmeur 2017

## Modern sugarcane cultivars



CIRAD, Garsmeur 2017

## Fine structure and evolution of the sugarcane genome

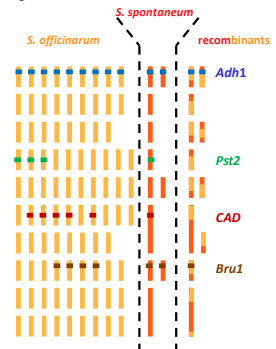
→ Comparison of sugarcane homoeologous regions

Thirteen hom(oe)ologous BAC clones corresponding to the *Adh1* region

Four hom(oe)ologous BAC clones corresponding to the *Pst2* region

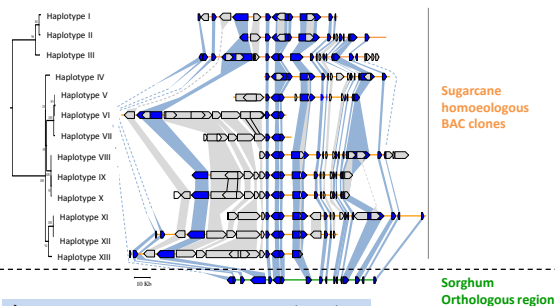
Six hom(oe)ologous BAC clones corresponding to the *CAD* region

Seven hom(oe)ologous BAC clones corresponding to the *Bru1* region



CIRAD, Garsmeur 2017

## Analysis of sugarcane homeologs and comparison with sorghum



→ Gene structure conservation among sugarcane homeologs  
→ All hom(oe)-alleles predicted functional  
→ High colinearity with sorghum  
→ Transposable elements variations

Garsmeur et al. New Phytologist 2010  
Charron et al. submitted

## Sugarcane sequencing strategy

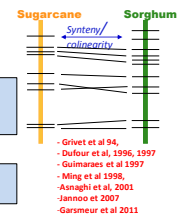
- Genes are conserved among sugarcane hom(oe)ologous chromosomes

→ if we could sequence a set of BACs representing one monoploid genome, it would represent a very useful reference sequence

- High colinearity between sugarcane and sorghum

→ We should be able to use sorghum to identify a core set of sugarcane BACs representing one monoploid genome

→ Mainly the genes are conserved → we focus on the gene-rich part of the genome

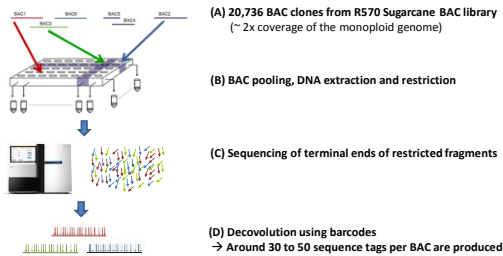


- Griet et al 94,  
- Dufour et al, 1996, 1997  
- Guimaraes et al 1997  
- Ming et al 1998,  
- Anaghi et al, 2001  
- Jannou et 2007  
- Garsmeur et al 2011

DHONT PAG 2015

## BAC selection through Whole Genome Profiling (WGP) Technology

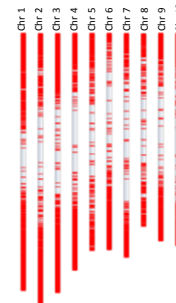
WGP technology generates short sequence tags from the terminal ends of restriction fragments from pooled BACs



→ Anchoring of the produced WGP-tags onto the sorghum genome

cirod, Garameur 2017

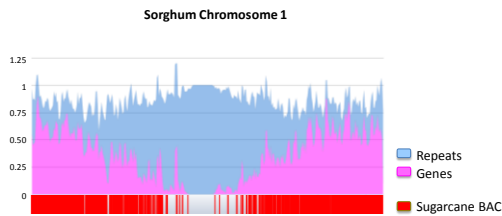
## Distribution of the sugarcane BAC in the sorghum genome



→ 11,732 R570 sugarcane BACs anchored onto the sorghum

cirod, Garameur 2017

## Sugarcane BACs anchor in sorghum gene-rich regions

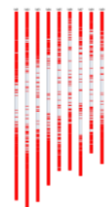


→ BAC mostly distributed in gene-rich regions

DHONT PAG 2015

## Minimum tiling path (MTP) of sugarcane BACs

MTP = minimum set of BACs to be sequenced to obtain the best coverage of chromosomes.



11,732 R570 sugarcane BACs anchored onto the sorghum

Sorghum	Number of genes on sorghum	Genes covered by BACs	MTP of BACs
Sb01	5447	4510 83%	822
Sb02	4317	3314 77%	602
Sb03	4461	3631 81%	654
Sb04	3599	2829 79%	522
Sb05	2322	1558 67%	298
Sb06	2822	2195 78%	414
Sb07	2277	1719 75%	313
Sb08	1933	1325 69%	274
Sb09	2610	2043 78%	399
Sb10	2801	1966 70%	390
	32589	25090 77%	4688

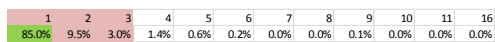
→ MTP ~4,700 BACs to cover the basic sugarcane genome

cirod, Garameur 2017

## Sequencing the MTP of R570 BAC

BAC sequenced through international collaboration

BAC sequenced using PacBio RSII technology and 100X depth coverage



97.5% of BAC assembled in less than 3 contigs

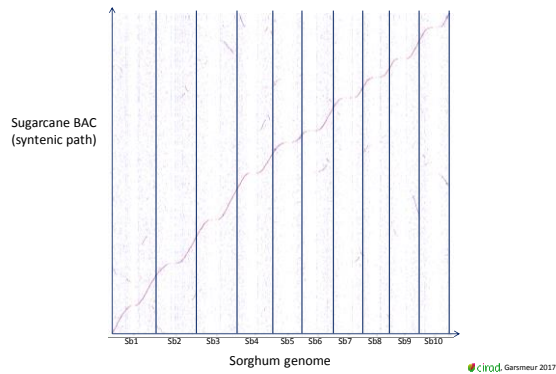


cirod, Garameur 2017

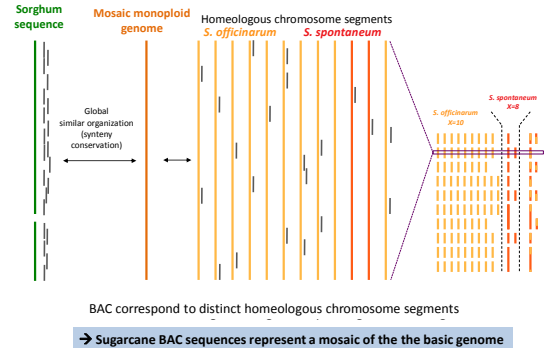
## The monoploid sugarcane reference sequence

- ☑ 4688 sugarcane BAC sequenced covering the gene-rich part of the basic genome
- ☑ high quality sequence
- ☑ 85 % assembled in one contig
- ☑ Cover ~80 % of the sorghum genes

## Coverage of the sorghum genome by the sugarcane sequence



## A mosaic sequence of the monoploid genome

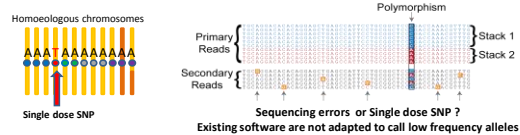


## Anchoring BAC sequences onto sugarcane chromosomes

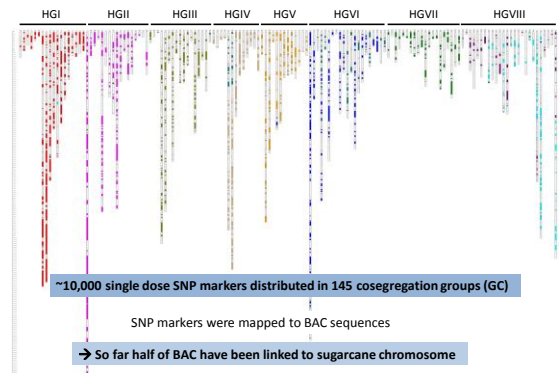
Sugarcane BAC were selected based on synteny conservation with sorghum. It's now essential to anchor them on sugarcane chromosomes

### → Development of a R570 high density genetic map

- Genotyping By Sequencing (Reduced-complexity method)
  - Avoid having to genotype the entire genome
  - Target and enrich specific loci to ensure sufficient sequence coverage
  - Produce thousands of SNP markers
- GBS of a mapping population (94 individuals)
- Development of bioinformatics tools for identifying single dose SNP that are useful for genetic mapping



## R570 genetic map



## BAC sequences annotation

Annotation of the BAC is in progress using a pipeline dedicated to sugarcane BAC annotation

- ☑ Predict location, structure and putative function of genes
- ☑ Has been successfully used to annotate sugarcane BAC sequence



### Exploiting R570 transcriptomic resources to improve annotations

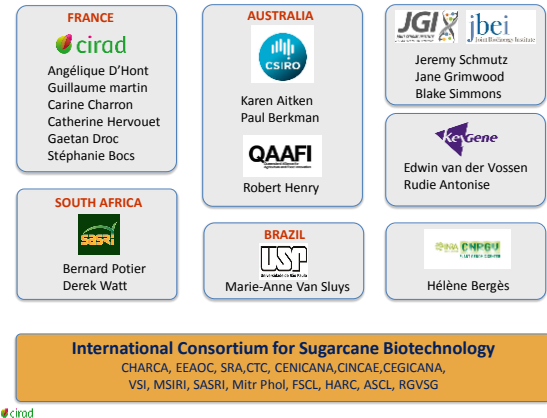
Sample	No of raw reads	No of filtered reads	% of reads after filtration
Leaf	64 146 407	62 951 164	98.14
Root	55 065 789	53 976 785	98.02
Stem	57 822 126	56 758 545	98.16
TOTAL	177 034 322	173 686 474	98.11

## Sugarcane Genome Hub



### This sugarcane sequence will represent an essential reference

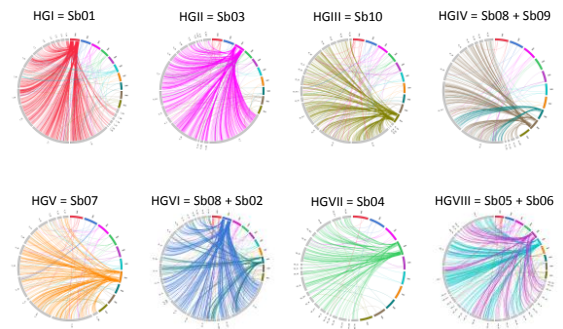
- To serve as template to align Genotyping By Sequencing data from any variety
  - for association and QTL studies, genomic selection using GBS markers
- To serve as template to align RNAseq data from any variety
  - to study gene expression in specific conditions
- To finely map or clone genes of interest
  - for marker-assisted selection
- To serve as high quality framework to help assemble the whole genome sequence



cirad

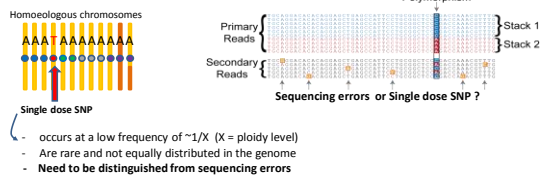
### Colinearity of the R570 Genetic map with sorghum

SUPP



### Genetic mapping in polyploids is challenging ...

The first difficulty resides in the identification of single dose SNP markers that are the one used for genetic mapping.

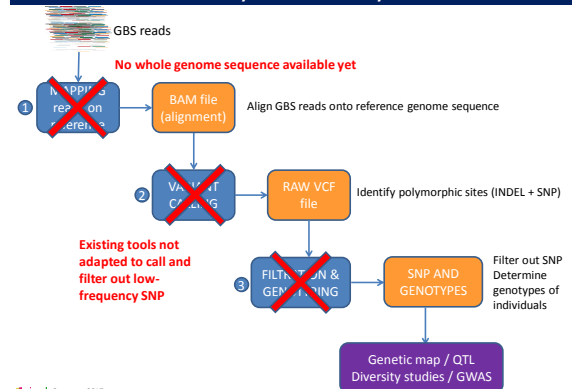


The 2<sup>nd</sup> difficulty is that existing softwares/methods such as TASSEL/GATK+SAMtools or denovo SNP pipelines such as /UNEAK/STACKS do not well manage with genotyping in polyploids, particularly with highly polyploid and heterozygous genomes

Need to develop and test analytic method adapted to complex genomes to identify high quality single dose SNP markers

cirad, Gansmeur 2017

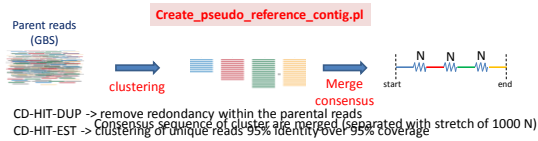
### SNP discovery : What is usually done ...



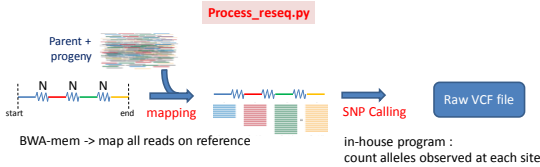
cirad, Gansmeur 2017

## Development of bioinformatics tools for GBS analyses in polyploids

### STEP1 = Build a reference pseudo contig from parental GBS reads

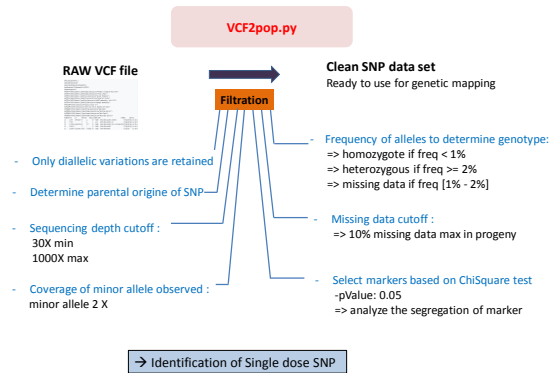


### STEP2 = Mapping of all reads onto the reference and SNP calling



cirod, Garsmeur 2016

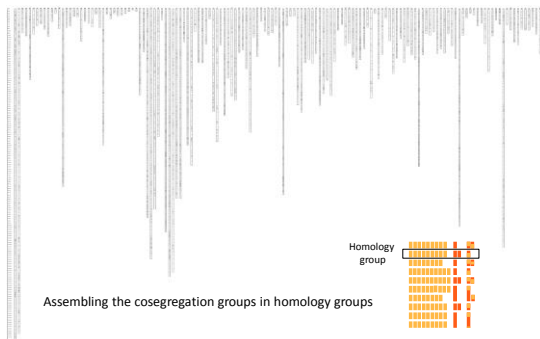
## Development of bioinformatic tools to identify simplex markers



cirod, Garsmeur 2017

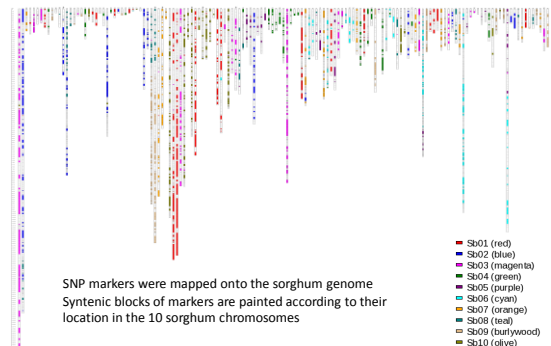
## R570 Sugarcane Genetic Map

~10,000 single dose SNP markers distributed in 145 cosegregation groups (GC)



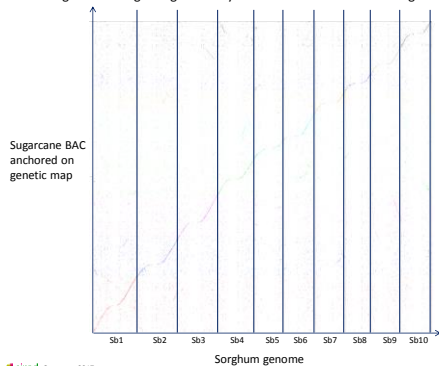
## Syntenic blocks between R570 genetic map and Sorghum

Homologous sugarcane chromosomes would correspond to same sorghum regions (global collinearity)



## Anchoring BAC sequences onto the sugarcane chromosomes

Coverage of the sorghum genome by the BAC anchored onto the sugarcane chromosomes



cirod, Garsmeur 2017

## Conclusions

- ☐ We developed a sequencing strategy to produce a reference sequence corresponding to the gene-rich regions of the basic (monoploid) sugarcane genome
- ☐ We have identified and sequenced a set of 5000 BAC covering the gene-rich part of the 10 basic sugarcane chromosomes
- ☐ We tested GBS and developed bioinformatics tools able to discover single dose SNP markers in complex polyploid genomes
- ☐ We built a genetic map comprising ~10,000 SNP markers and use it to anchor around half of the BAC

→ Need to increase the number of marker : GBS on selfed progeny from R570 ?  
 → Identifying SNP on BAC : Targeted sequence capture on BAC sequence ?

This reference sequence will be a very useful resource for genetics (GWAS, GS) and genomics studies in sugarcane ( WGD, expression, ..)  
 It will also represents an essential high quality frame to help building a whole genome sequence