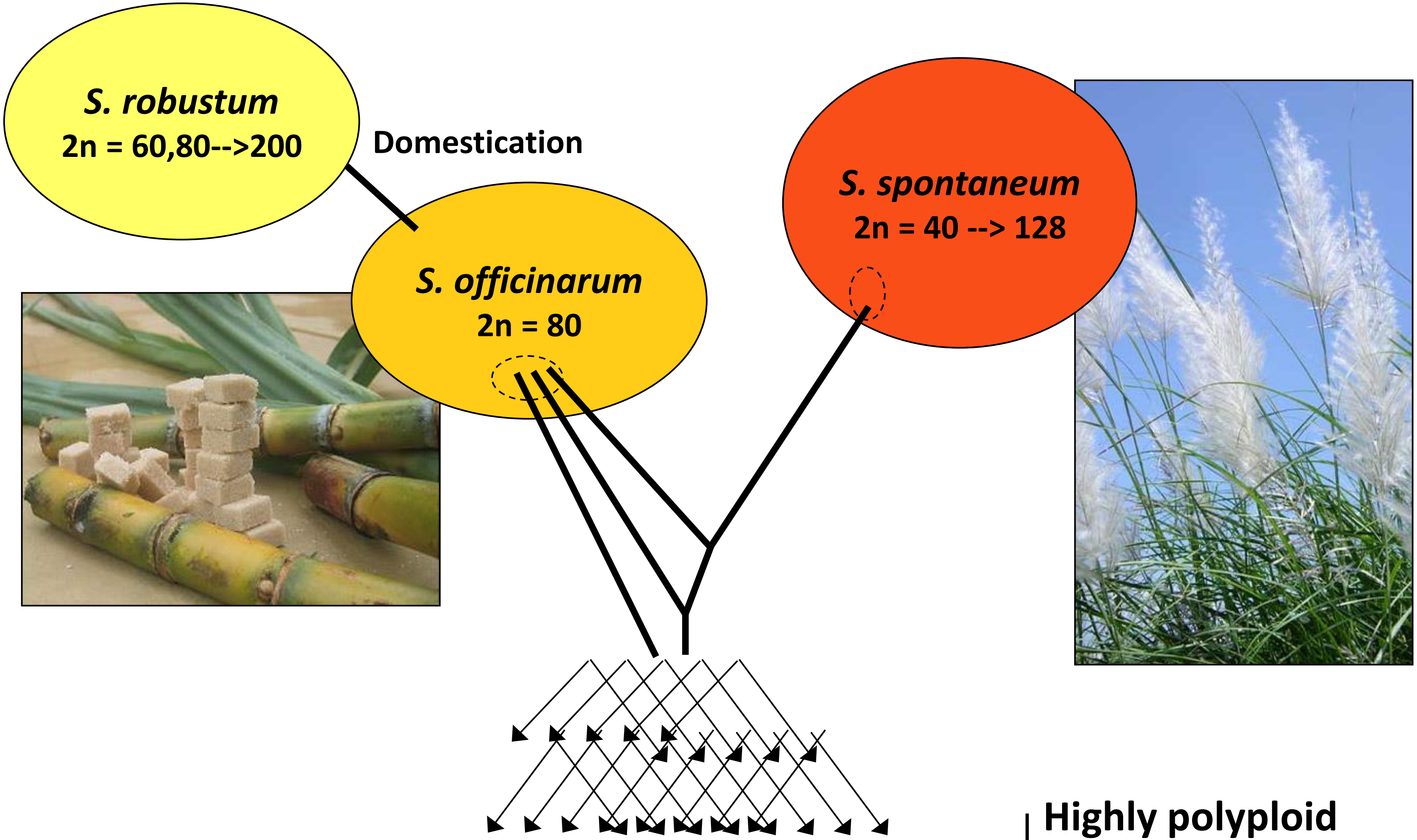
A mosaic monoploid reference sequence for the highly complex genome of sugarcane

Olivier GARSMEUR
CIRAD, équipe Structure et Evolution des Génomes

cirad CSB sasri CSIRO KeyGene QAAFI jbei JGI

Modern sugarcane cultivars



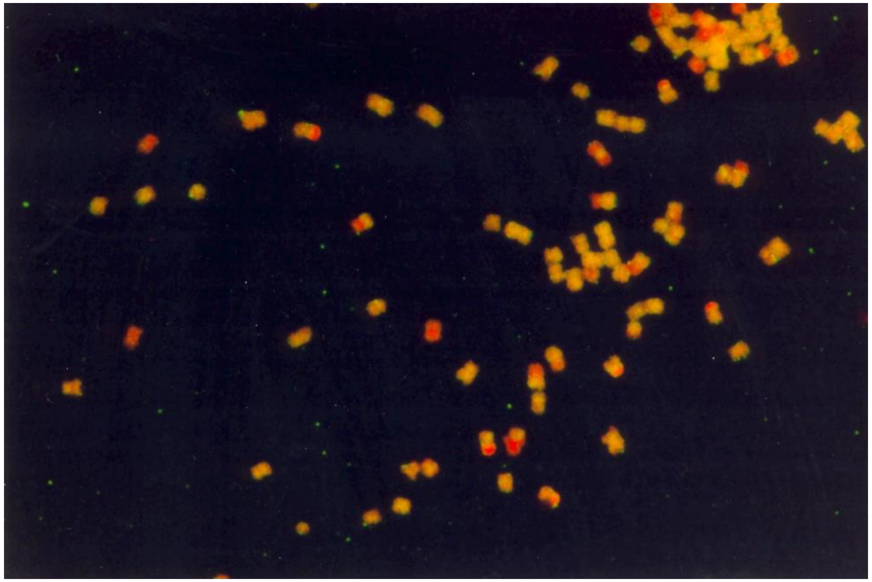
modern cultivars

2n ~ 110-130

Highly polyploid
Aneuploid
Interspecific

Global organization of the complex genome of sugarcane

Genomic in situ hybridization

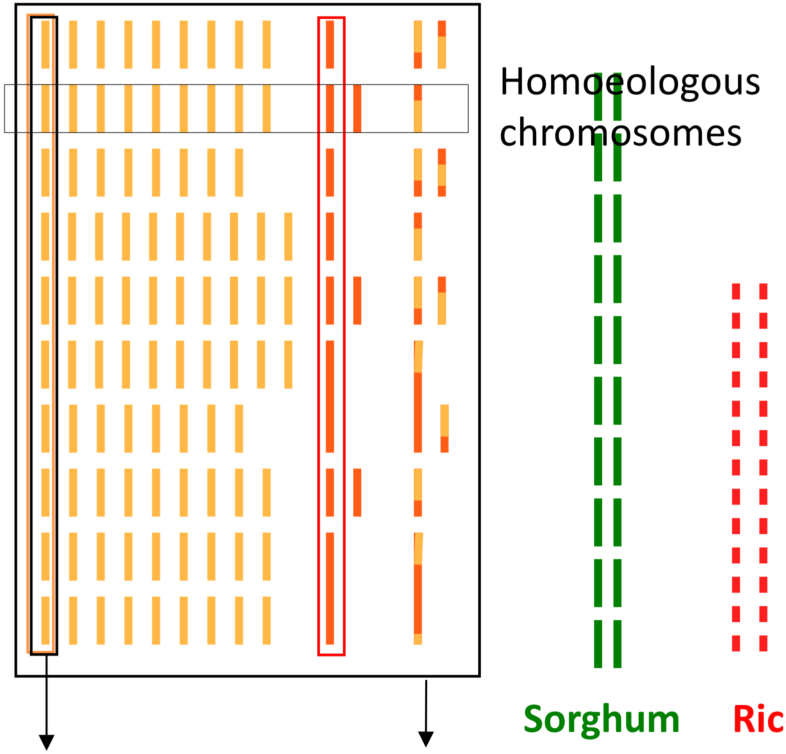


R570, 2n=ca 115

80 % S. officinarum

10% S. spontaneum

10% recombinants



S. spontaneum

S. officinarum

x=10

x=8

recombinants

Homoeologous chromosomes

Sorghum

750 Mb

Rice

400 Mb

Monoploid genome size: 930 Mb

Total genome size: 10 Gb

(photo and figure adapted from D'Hont et al, 1996; D'Hont et al, 1998; D'Hont 2005, Piperidis et al 2010)

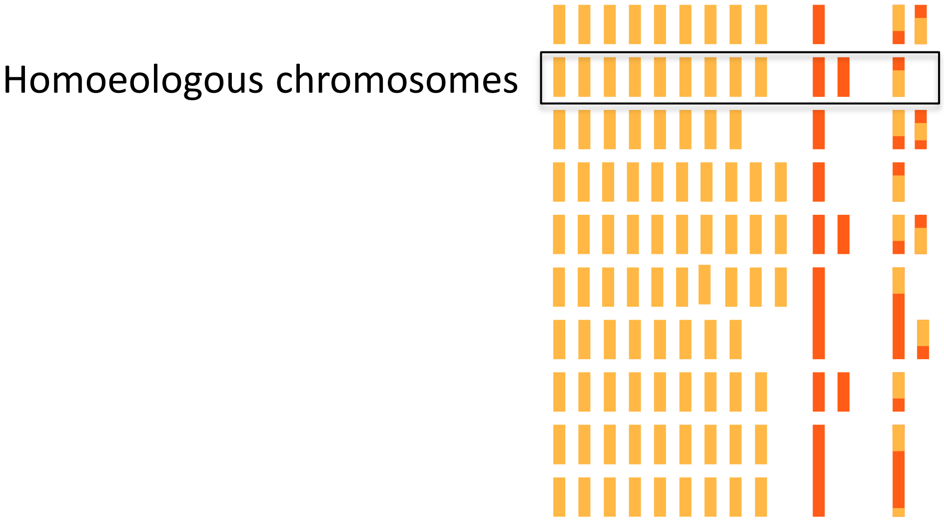
Sequencing sugarcane

The sugarcane genome poses challenges that have not been addressed in any prior sequencing project.

The main difficulties reside in

- its **high polyploidy**, with a complete set of homoeologous chromosomes predicted to range from **10 to 12 copies**

- its **high level of heterozygosity** which makes an assembly of the whole genome very challenging through a classical shotgun sequencing approach.



Sugarcane

Comparison of sugarcane homeologs and comparison with sorghum

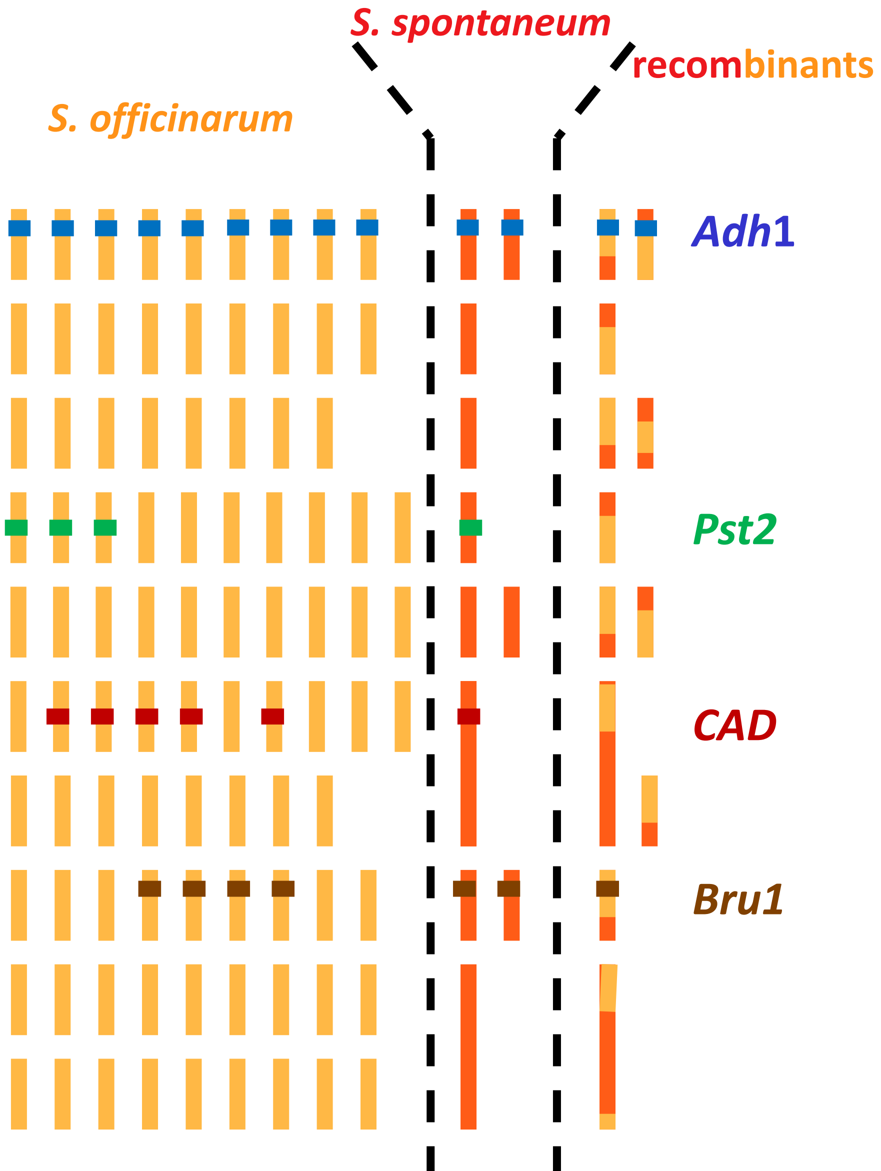
→ Comparison of sugarcane homoeologous regions

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

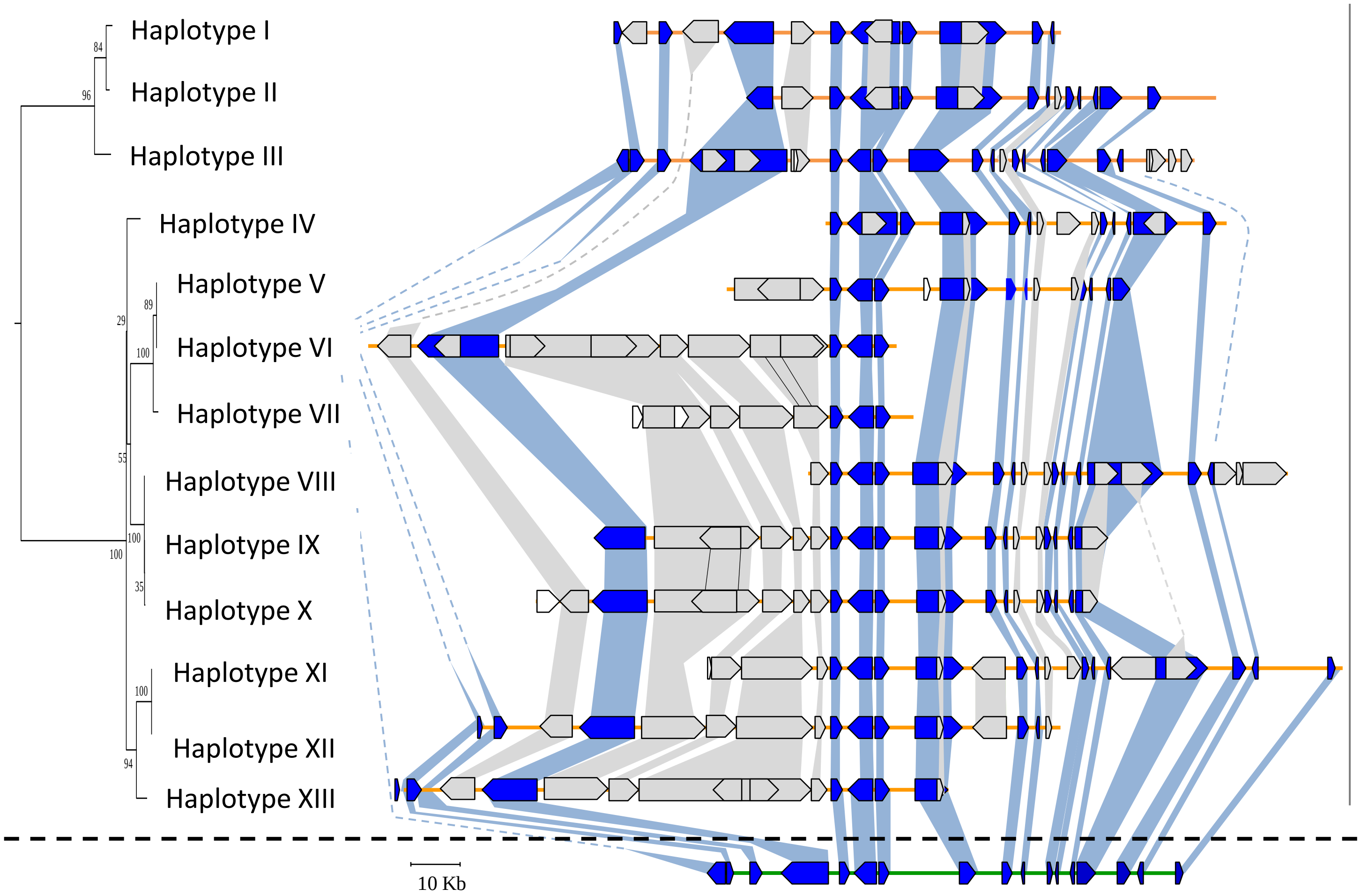
Four hom(oe)ologous BAC clones corresponding to the ***Pst2*** region

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

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



Comparison of sugarcane homeologs and comparison with sorghum



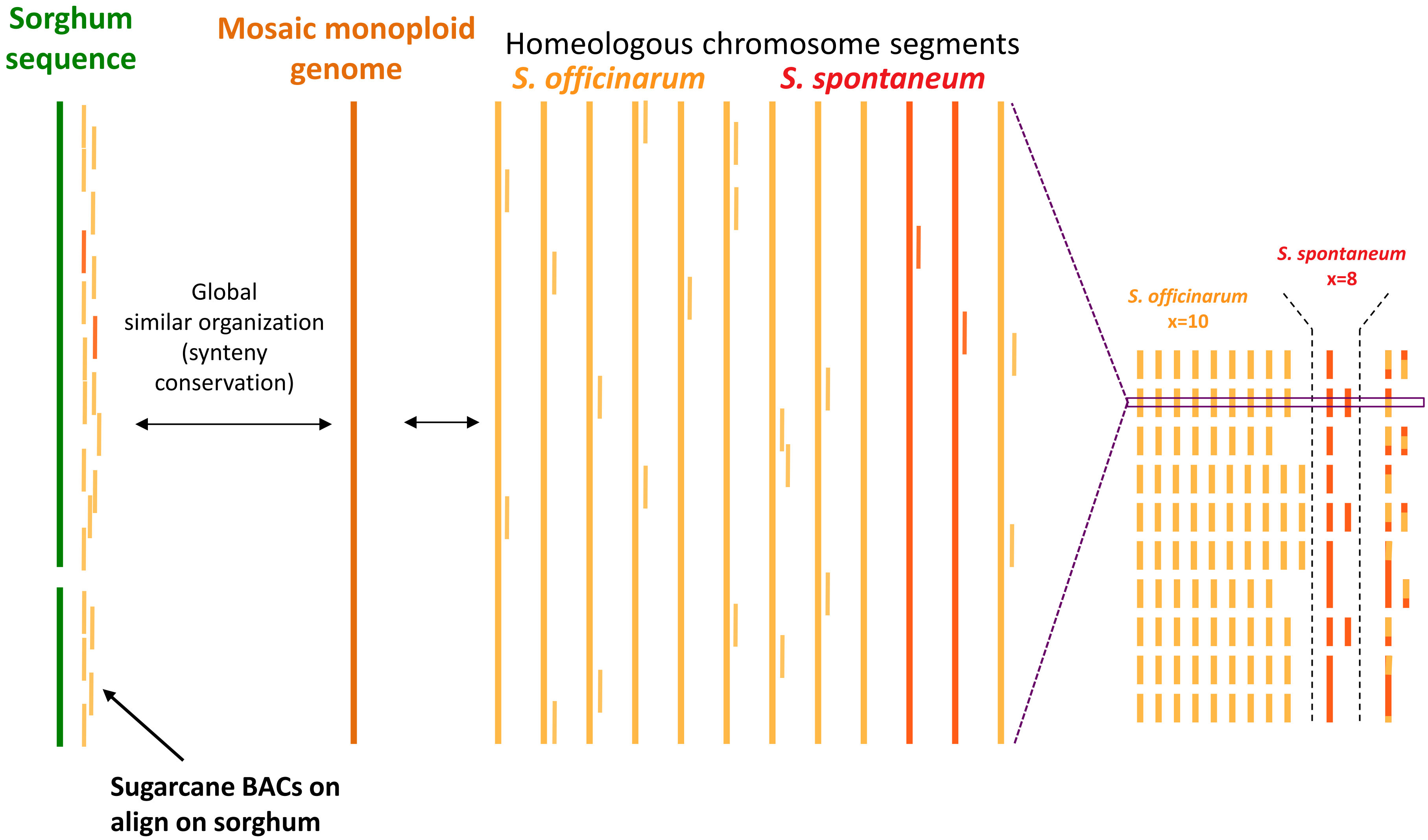
Sugarcane homoeologous BAC clones

Sorghum Orthologous region

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

Garsmeur *et al*. New Phytologist 2010
unpublished data

The sugarcane sequencing strategy



Sorghum sequence

Mosaic monoploid genome

Homeologous chromosome segments

S. officinarum ***S. spontaneum***

Global similar organization (synteny conservation)

Sugarcane BACs on align on sorghum

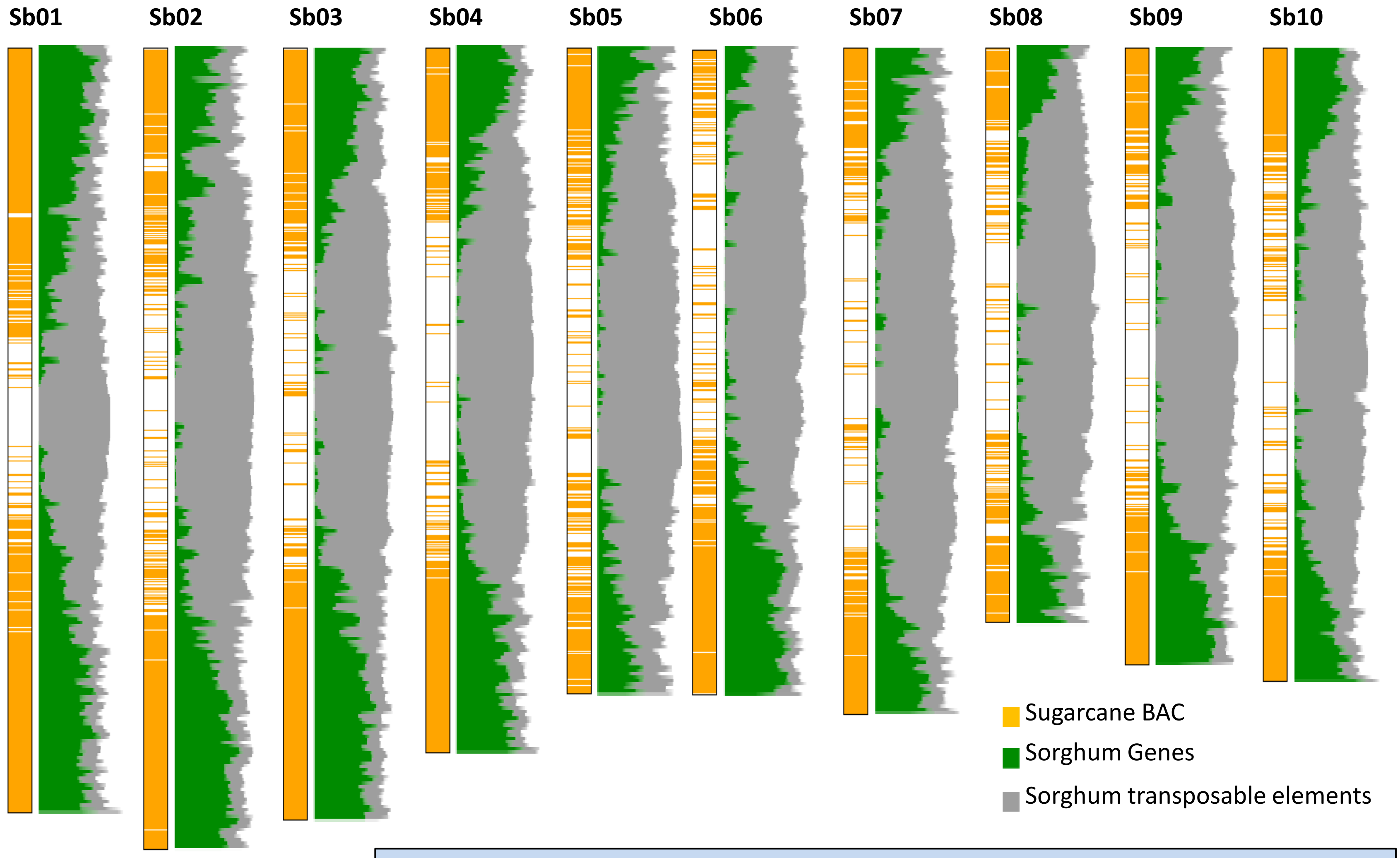
→ Sugarcane BAC sequences represent a mosaic of the monoploid genome

Sugarcane BAC selection using the sorghum genome sequence

→ **Whole Genome Profiling (WGP) technology**

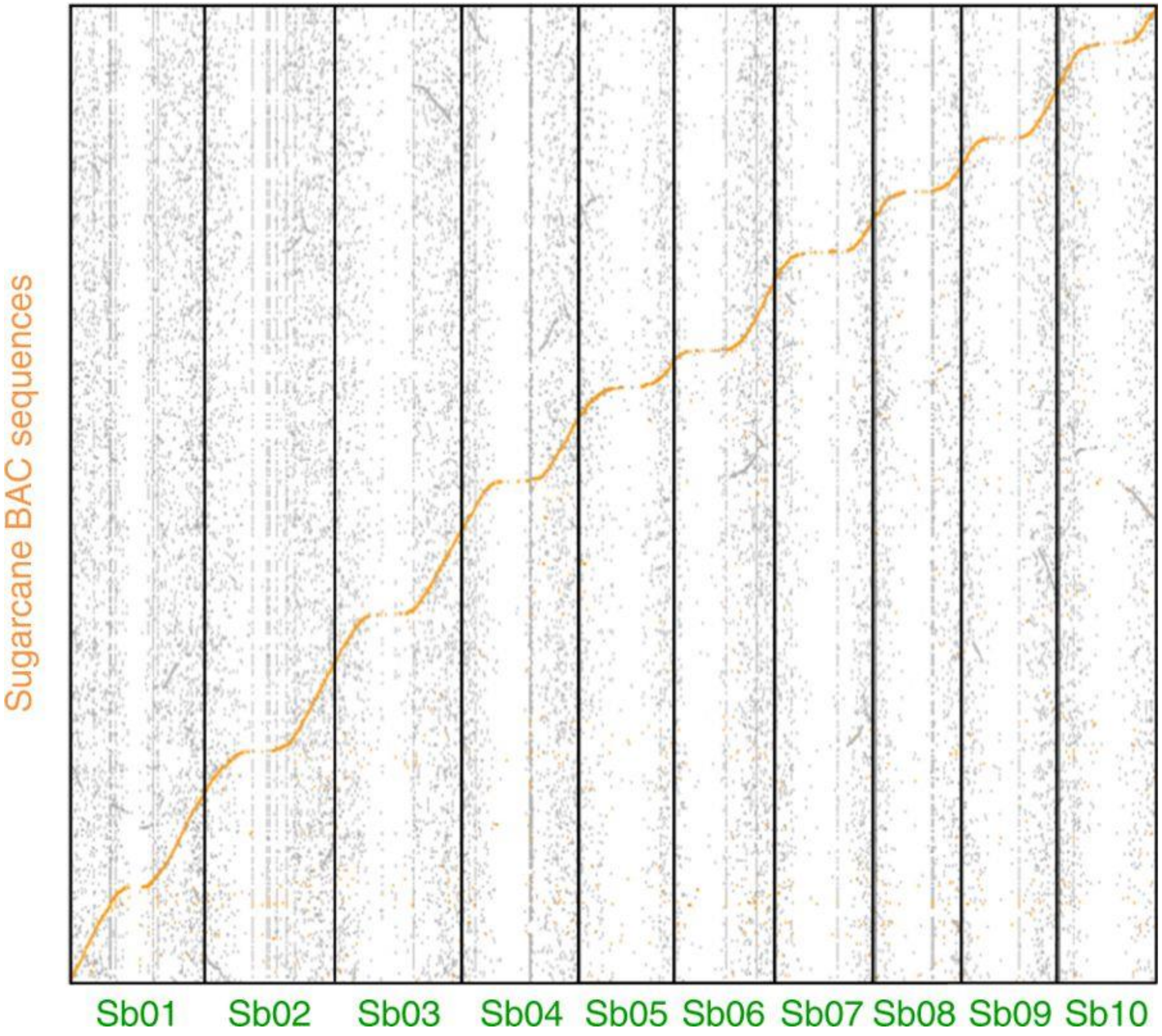
produce short illumina sequence tags from terminal ends of restricted BAC fragments

20,736 sugarcane BAC → 455,656 uniq WGP tags



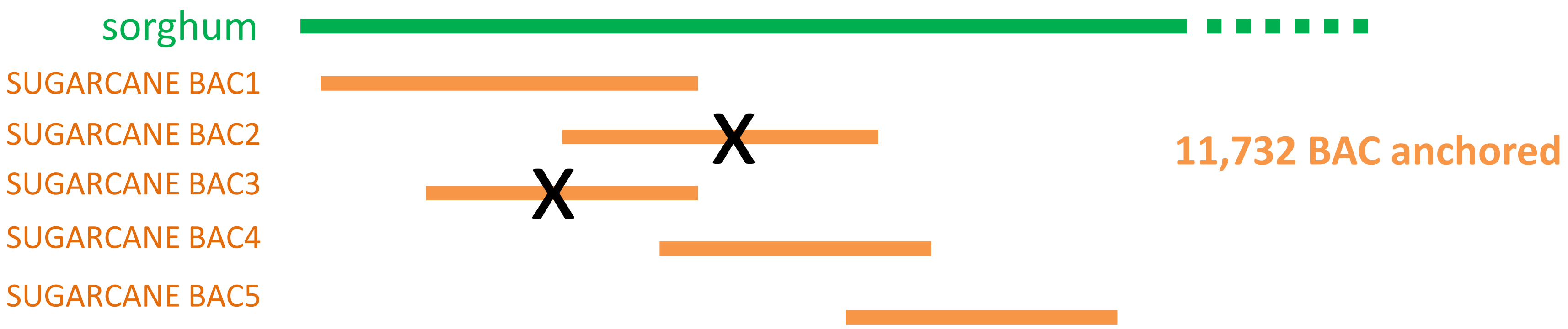
→ **11,732 R570 sugarcane BACs anchored onto the sorghum**

Coverage of the sorghum genome by sugarcane BAC sequences



Selection of a Minimum Tiling Path (MTP) of sugarcane BACs

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



11,732 BAC anchored => Minimun Tiling Path 4,660 sugarcane BACs to be sequenced

Represent the number of BAC to be sequenced to cover the gene-rich part of a basic/monoploid sugarcane genome

Sequencing a Minimum Tiling Path (MTP) of R570 BAC

Aim = High quality sequence and BAC assembled in 1 contig

BAC sequenced using PacBio RSII technology and 100X depth coverage

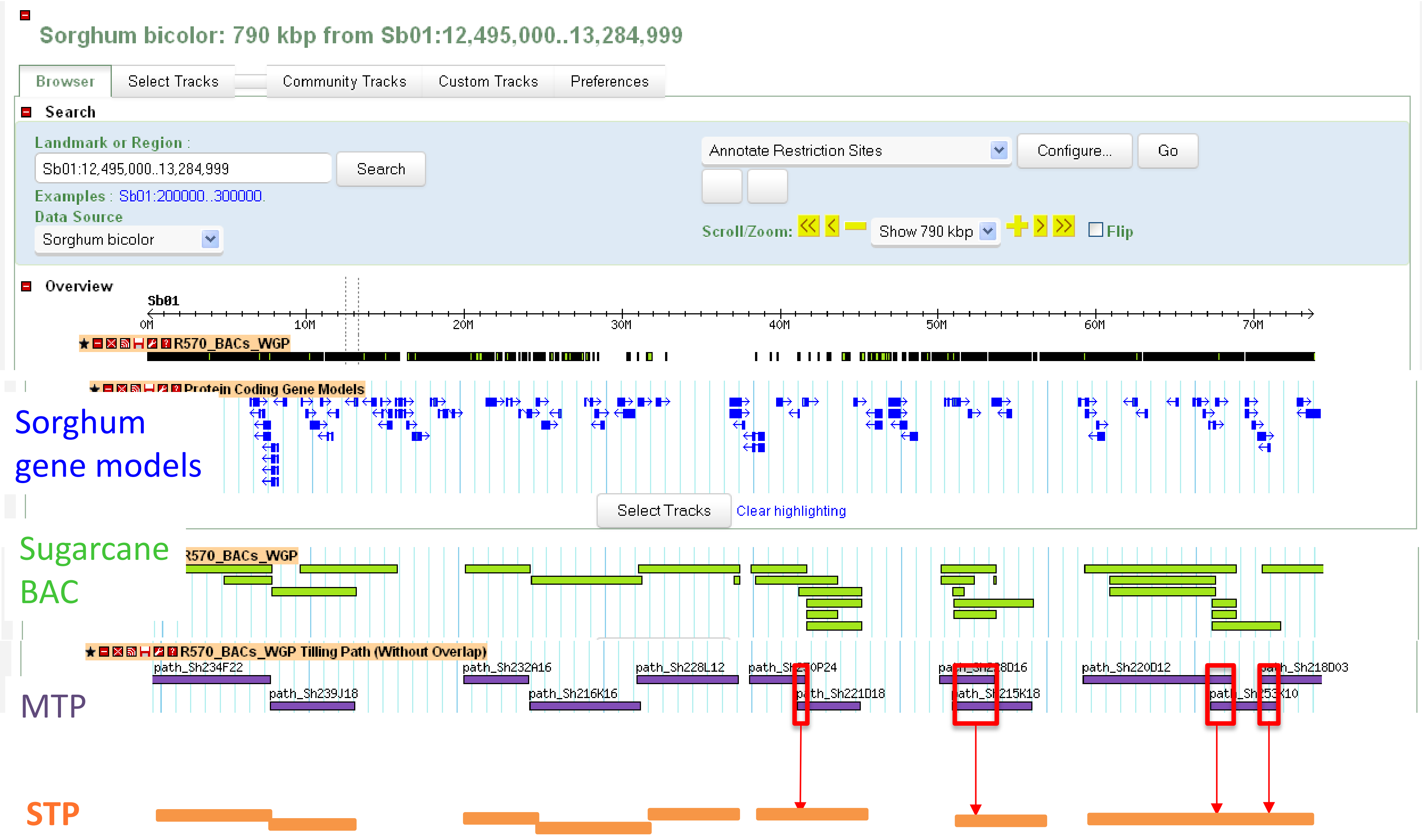
Single Molecule Real Time sequencing (SMRT) → produce long reads (several kb)
high 100X sequencing → High quality sequence

BAC sequenced through international collaboration

→ 4535 BAC sequenced
→ 97.5% of BAC assembled in <= 3 contigs

→ 531 Mb of high quality sequences
(0.02% of N)

Assembly of Single tiling path (STP) of sugarcane BACs



Sorghum bicolor: 790 kbp from Sb01:12,495,000..13,284,999

Landmark or Region : Sb01:12,495,000..13,284,999

Examples : Sb01:200000..300000.

Data Source: Sorghum bicolor

Annotate Restriction Sites

Scroll/Zoom: Show 790 kbp ☐ Flip

From MTP = 531 Mb → Assembly STP = 382 Mb

Global genome comparison with sorghum

→ Construction of a R570 SNP genetic map

Map constructed with 186 individuals from the self progeny of R570 genotyped with **DArTseq**

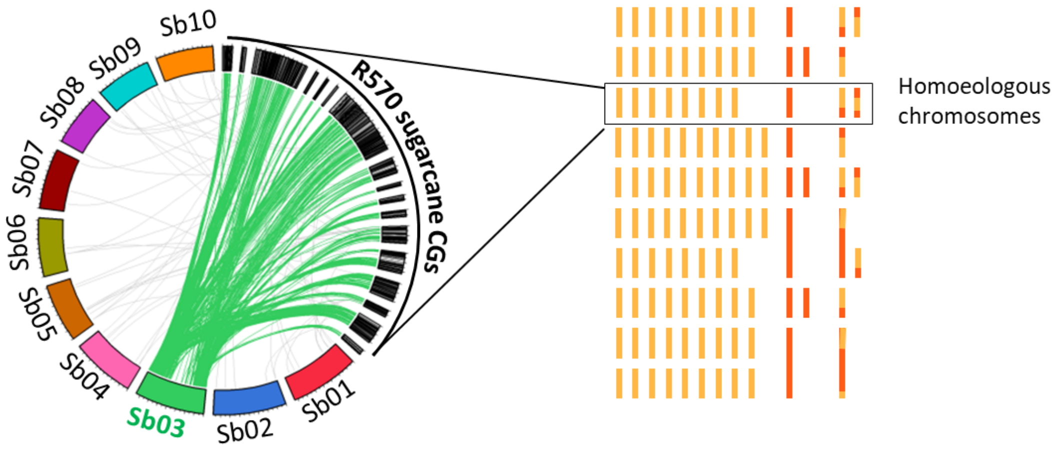
Data analyzed with in-house pipeline : **Process_reseq and vcf2pop**
Available at https://github.com/SouthGreenPlatform/VcfHunter/

→ 12,468 Single dose SNPs assembled in 132 Cosegregation Groups (CG)

Markers aligned onto the sorghum genome

→ 95 % of the CGs with most markers aligning on one sorghum chromosome

→ Allow assembling CGs in sugarcane Homoeology Groups (HG)

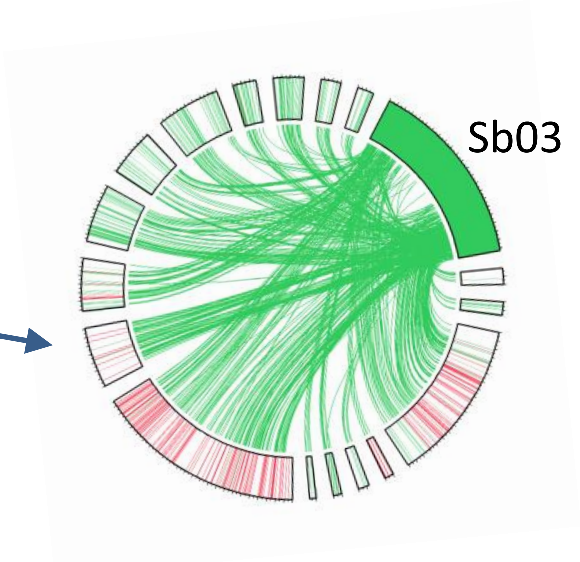


origin of R570 cultivar CGs

DArTseq Genotyping of a panel of 34 Saccharum accessions

-12 *S. officinarum*
- 9 S. robustum
- 13 S. spontaneum

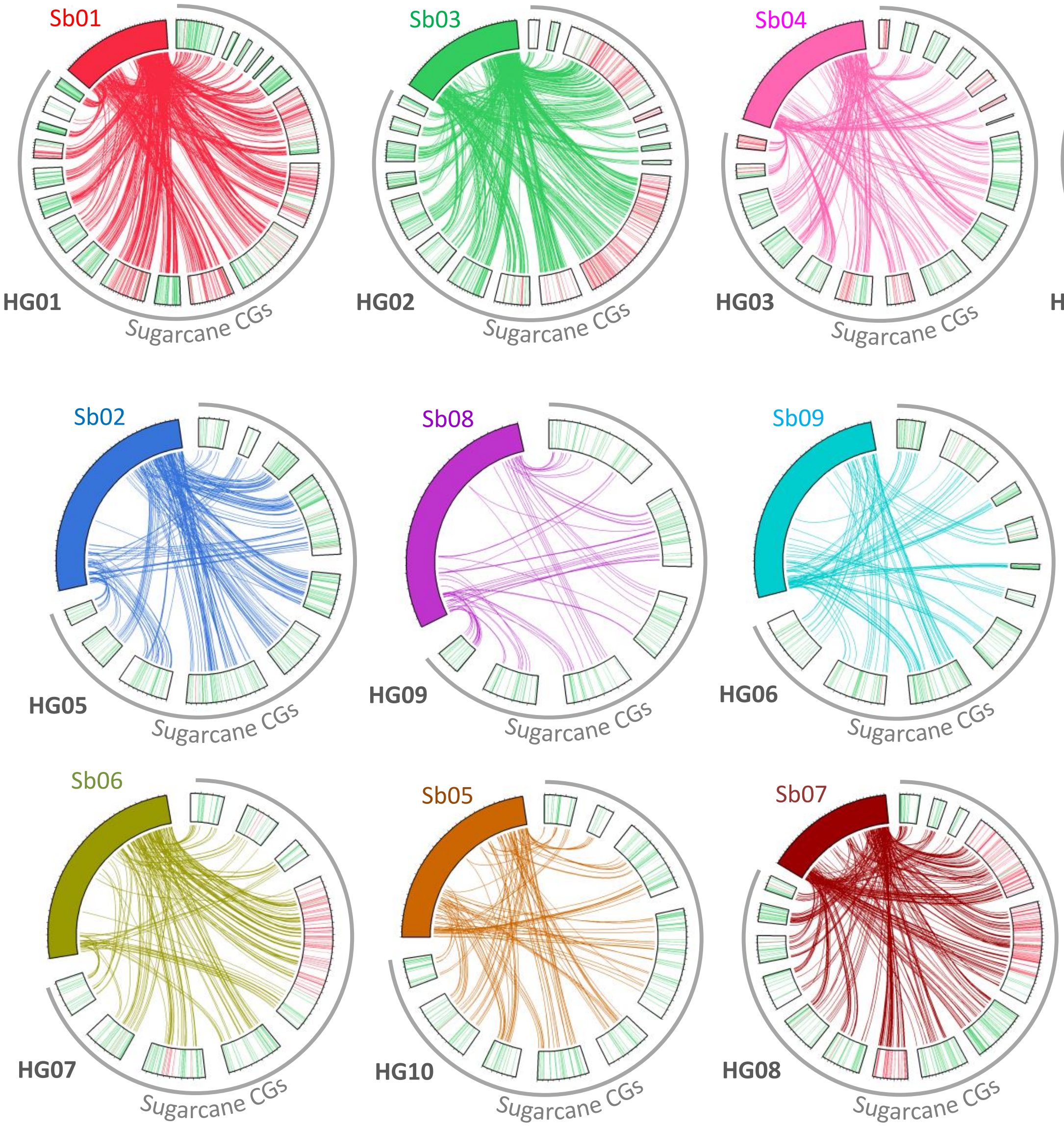
→ Identification diagnostic SNP



S. officinarum/robustum VS S. spontaneum

Global genome comparison with sorghum

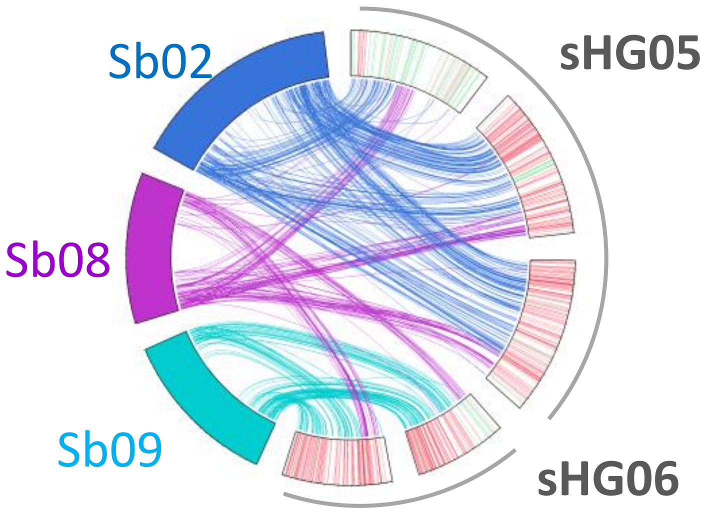
→ **125 CGs assembled in 10 homoeology groups aligning on one sorghum chromosome**

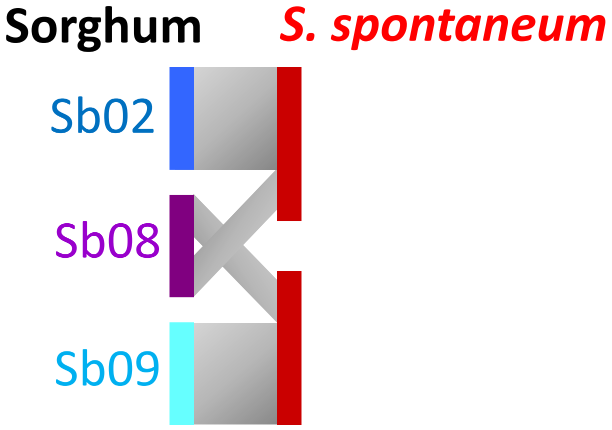


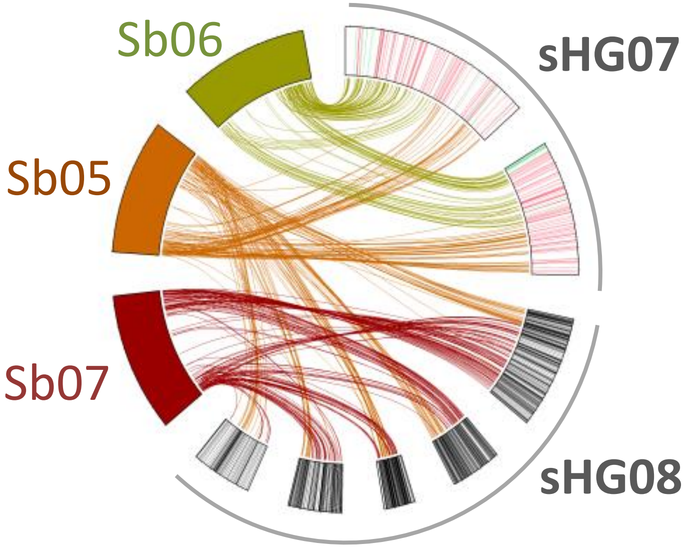
S. spontaneum
S. officinarum

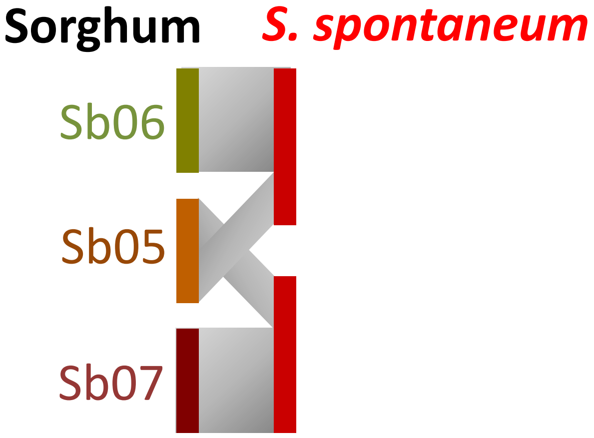
Global genome comparison with sorghum

→ **For a few CGs most markers aligning on two sorghum chromosomes**

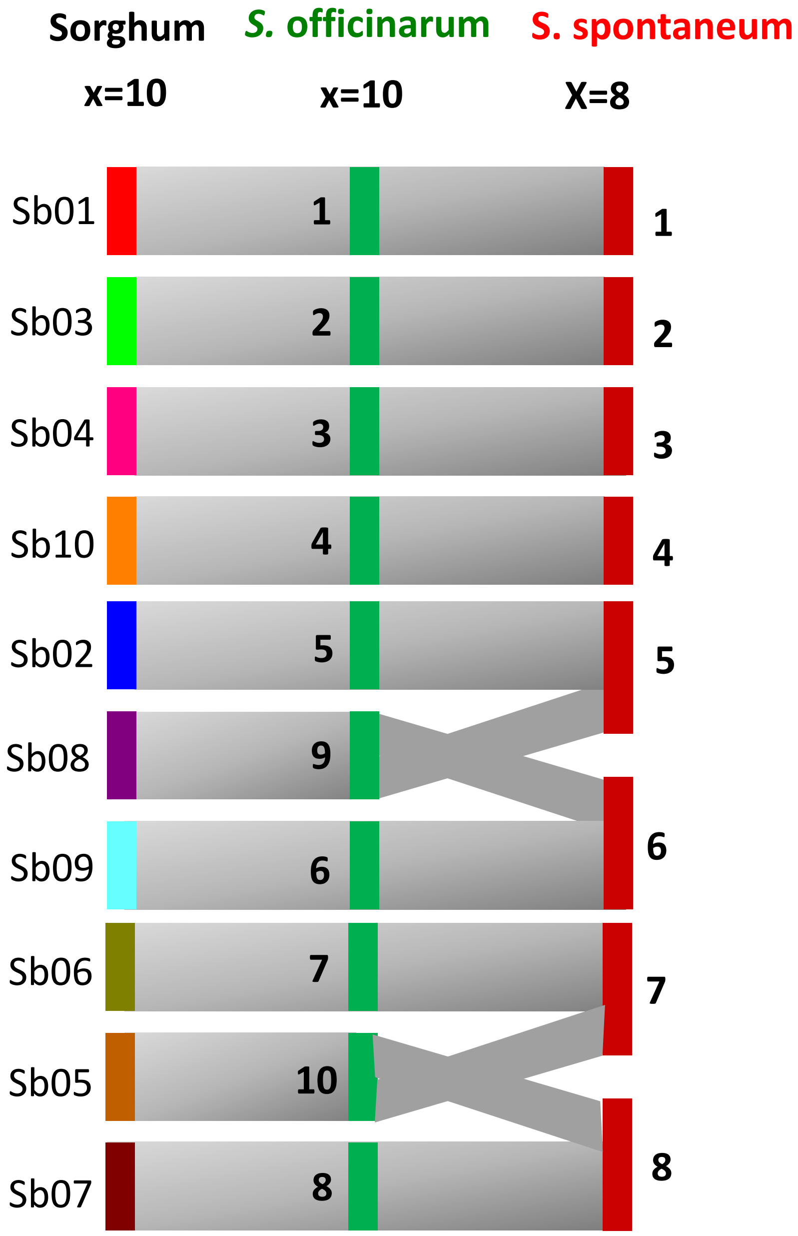








Global Sorghum and Sugarcane genomes comparison



Conclusions

☑ A high quality sequences of 382 Mb representing the gene rich fraction of the sugarcane monoploid genome

☑ 25,316 protein-coding gene models within their genomic sequence context providing access to regulatory elements for functional studies

Serve as high quality framework to help assemble the whole genome of sugarcane

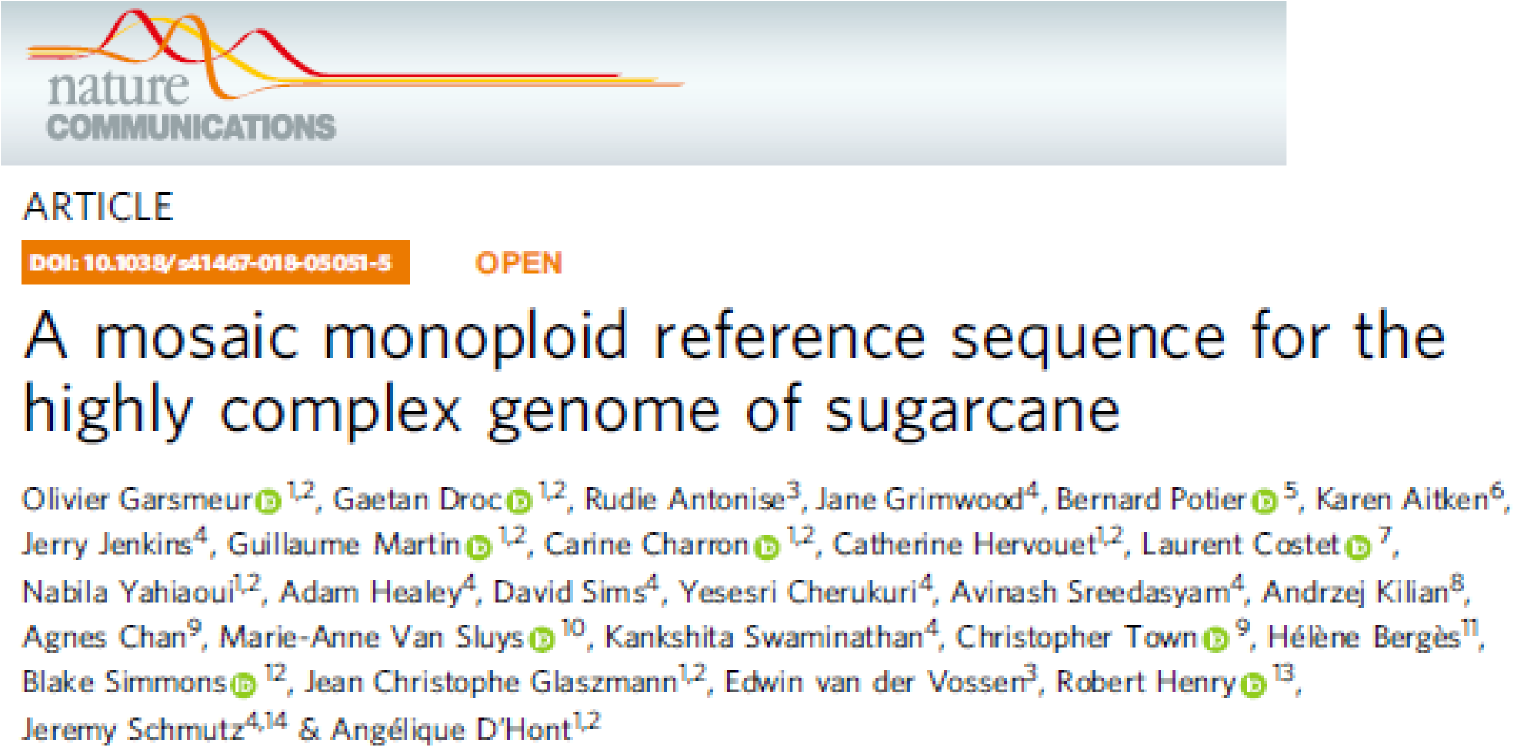
→ Project in progress in progress with JGI

☑ Confirm global colinarity between sugarcane and sorghum

☑ A few chromosome rearrangements between *Sorghum/S. officinarum and S. spontaneum*

→ These rearrangements corroborate and explain a variation of basic chromosome numbers in *S. officinarum* vs *S. spontaneum* with x= 10 and x=8, respectively

→ x=10 is recognized to be ancestral in the Saccharinae-Sorghinae, these rearrangements thus must have arisen in the *S. spontaneum* lineage after its divergence from the lineage of *S. officinarum*


nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-05051-5 OPEN

A mosaic monoploid reference sequence for the highly complex genome of sugarcane

Olivier Garsmeur[1,2], Gaetan Droc[1,2], Rudie Antonise[3], Jane Grimwood[4], Bernard Potier[5], Karen Aitken[6], Jerry Jenkins[4], Guillaume Martin[1,2], Carine Charron[1,2], Catherine Hervouet[1,2], Laurent Costet[7], Nabila Yahiaoui[1,2], Adam Healey[4], David Sims[4], Yesesri Cherukuri[4], Avinash Sreedasyam[4], Andrzej Kilian[8], Agnes Chan[9], Marie-Anne Van Sluys[10], Kankshita Swaminathan[4], Christopher Town[9], Hélène Bergès[11], Blake Simmons[12], Jean Christophe Glaszmann[1,2], Edwin van der Vossen[3], Robert Henry[13], Jeremy Schmutz[4,14] & Angélique D'Hont[1,2]

http://sugarcane-genome.cirad.fr

Welcome to the Sugarcane Genome Hub

Genomic resources for sugarcane

An integrated web-based database providing centralized access to the sugarcane reference genome sequences and genomic resources.

FRANCE

cirad

Olivier Garsmeur
Gaetan Droc
Guillaume martin
Carine Charron
Catherine Hervouet
Angélique D'Hont

AUSTRALIA

CSIRO

Karen Aitken

QAAFI
Queensland Alliance for Agriculture and Food Innovation

Robert Henry

USA

JGI JOINT GENOME INSTITUTE DEPARTMENT OF ENERGY

jbei Joint BioEnergy Institute

Jeremy Schmutz
Jane Grimwood
Jerry Jenkins
Blake Simmons

NETHERLANDS

KeyGene

Edwin van der Vossen
Rudie Antonise

SOUTH AFRICA

sasri

Bernard Potier

BRAZIL

USP Universidade de São Paulo

Marie-Anne Van Sluys

FRANCE

INRA CNRGV PLANT GENOMIC CENTER

Hélène Bergès

International Consortium for Sugarcane Biotechnology

ICSB

CHARCA, EEAOC, SRA,CTC, CENICANA,CINCAE,CEGICANA,
VSI, MSIRI, SASRI, Mitr Phol, FSCL, HARC, ASCL, RGVSG