SNP Identification to Breed for Biomass Production in Sorghum: an alternative to non-renewable resources in tropical and temperate regions

Vilmus Ingrid[1], Berger Angélique[1], Trouche Gilles[2], Maleyrat Julien[1], Vaksmann Michel[2], Bastianelli Denis[3], Clerget Benoit[4], Chantereau Jacques[1], Rivallan Ronan[1], Bonnal Laurent[3], Mournet Pierre[1], Couvin David[5], Dufayard Jean-François[5], Barrière Yves[6], Höfte Herman[7], Rami Jean-François[1], Pot David[1]
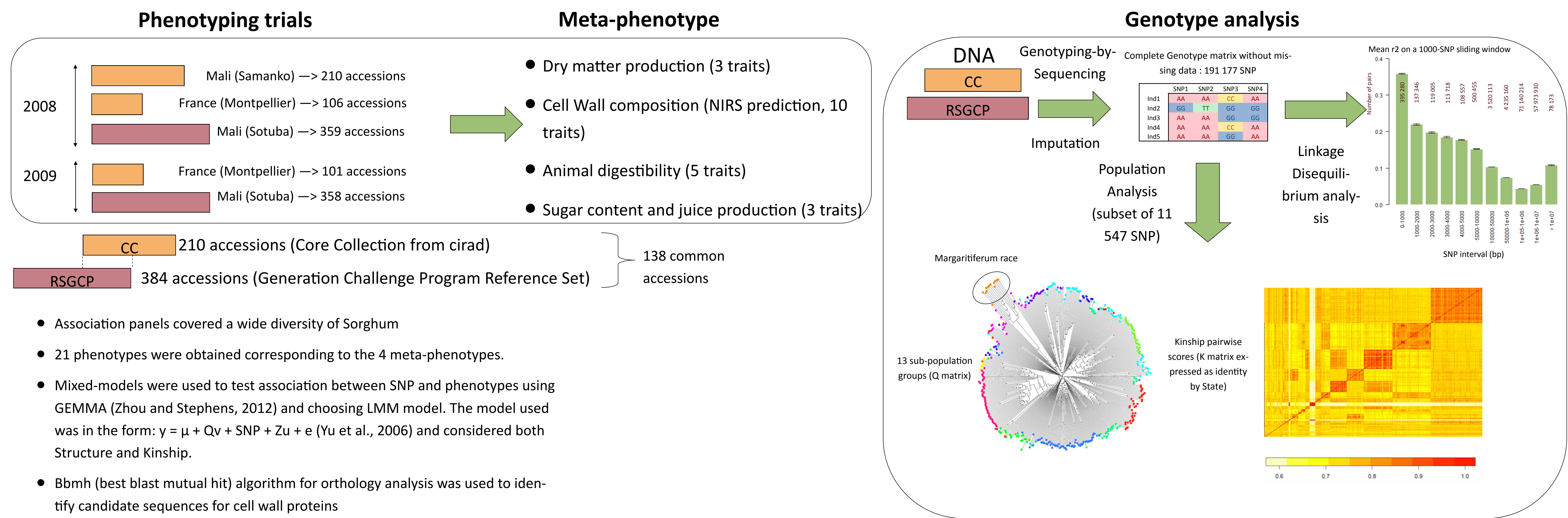
1 CIRAD, UMR AGAP, DDSE team, TA 108/03, Av Agropolis, 34398 Montpellier CEDEX 5, France
2 CIRAD, UMR AGAP, GIV team, TA 108/03, Av Agropolis, 34398 Montpellier CEDEX 5, France
3 CIRAD, UMR SELMET, TA 108/03, Av Agropolis, 34398 Montpellier CEDEX 5, France
4 CIRAD, UMR AGAP, PAM team, TA 108/03, Av Agropolis, 34398 Montpellier CEDEX 5, France
5 CIRAD, UMR AGAP, ID team, Av Agropolis, 34398 Montpellier CEDEX 5, France
6 INRA, Unité de Génétique et d'Amélioration des Plantes Fourragères (UGAPF), UR889, Route de Saintes, 86600 Lusignan, France
7 INRA, Institute Jean-Pierre Bourgin, Laboratoire de Biologie Cellulaire, 78026, Versailles, France
Email : ingrid.vilmus@cirad.fr

1. Introduction : Sorghum biomass as a Carbon Source

- To substitute to non-renewable fossil carbon sources through biorefinery approach

  ⟹ To produce energy : bioethanol, methane, electricity

  ⟹ To make biocomposites (matrix + natural fiber) and biomaterials for building

  ⟹ To support the green-chemistry value chain

- To feed animal

These goals can be achieved focusing on various traits relative to biomass composition and production:

- Dry matter production (plant height, dry matter content, principal stem dry matter weight)

- Cell Wall composition (lignin, cellulose, hemicellulose content)

- Animal nutrition (organic fiber digestibility, dry matter digestibility, solubility, nitrogen and mineral content)

- Sugar content and juice production (Brix, Brix at 30% moisture and juice weight in principal stem)

Analysis of genetic determinism is of key importance to:
i. Identify genome areas involved in trait variation
ii. Support and optimise breeding strategies adapted to the different industrial processes using knowledge from i.

Main objective : Highlight SNP to be used in a breeding process for biomass production

2. GWAS through Mixed-Model to detect associations between biomass characteristics and genotypes

Phenotyping trials

2008
- Mali (Samanko) —> 210 accessions
- France (Montpellier) —> 106 accessions
- Mali (Sotuba) —> 359 accessions

2009
- France (Montpellier) —> 101 accessions
- Mali (Sotuba) —> 358 accessions

CC 210 accessions (Core Collection from cirad)
RSGCP 384 accessions (Generation Challenge Program Reference Set)
138 common accessions

Meta-phenotype
- Dry matter production (3 traits)
- Cell Wall composition (NIRS prediction, 10 traits)
- Animal digestibility (5 traits)
- Sugar content and juice production (3 traits)

Genotype analysis

- Association panels covered a wide diversity of Sorghum
- 21 phenotypes were obtained corresponding to the 4 meta-phenotypes.
- Mixed-models were used to test association between SNP and phenotypes using GEMMA (Zhou and Stephens, 2012) and choosing LMM model. The model used was in the form: y = μ + Qv + SNP + Zu + e (Yu et al., 2006) and considered both Structure and Kinship.
- Bbmh (best blast mutual hit) algorithm for orthology analysis was used to identify candidate sequences for cell wall proteins
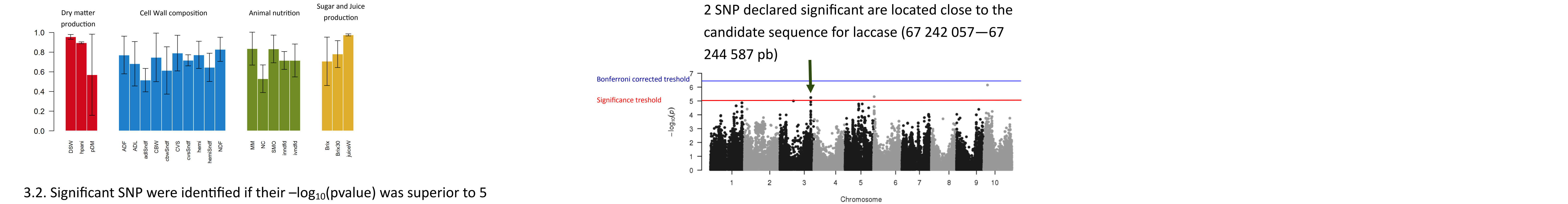
3. Association results match with some candidate sequences obtained from orthology analysis with other species

3.1. Heritabilitiy values were calculated using REML variances estimated from a mixed-model without SNP in Asreml-R (Butler, et al., 2009) for each variables in each trial.

3.2. Significant SNP were identified if their $-\log_{10}$(pvalue) was superior to 5 (average threshold value for 5% FDR correction on every trials-phenotype combination) and allowed to:

- Define genome areas containing at least 1 significant SNP
- Check for matching with candidate sequences from bbmh analysis

Meta-phenotype	genome areas associated	areas matching with candidate sequences (distance < 100 kb)	SNP inside or close to (< 5 kb) candidate sequences
Dry matter production	18	8	1
Cell Wall composition	85	43	26
Animal digestibility	50	19	6
Sugar and juice production	18	8	1

Number of genome areas associated with phenotypes variation and SNP directly linked to candidate sequences for cell wall proteins

3.3. A focus on the variable adlSndf (lignin/total fiber) observed in France in 2008 shows interesting candidate ortholog sequence for a laccase previously identified on Arabidopsis as implied in lignin polymerization (Zhao et al., 2013) and in maize as impacting biomass quality (Dhugga et al., 2011, patent US 2011/0239329 A1).

2 SNP declared significant are located close to the candidate sequence for laccase (67 242 057—67 244 587 pb)

Conclusion :

- Population analysis highlights 13 sub-population groups showing a genetic continuum except for one isolated group composed of acccesions from the margaritiferum race. Linkage disequilibrium decreased rapidly for distances superior to 1 kb and confirms the high number of SNP necessary to cover the entire Sorghum genome in association panels (Bouchet et al., 2012).

- GWAS analysis allowed to identify promising SNP to breed for biomass production, some being close or inside candidate sequences obtained from an orthology analysis with maize, rice and Arabidopsis. Because of the wide genetic diversity covered by association panels chosen, those results could serve in a breeding program using genitors from the cultivated pool of Sorghum.

Bibliography :
Bouchet, Sophie, David Pot, Monique Deu, Jean-François Rami, Claire Billot, Xavier Perrier, Ronan Rivallan, et al. 2012. "Genetic Structure, Linkage Disequilibrium and Signature of Selection in Sorghum: Lessons from Physically Anchored DArT Markers." PLoS ONE 7 (3): e33470. doi:10.1371/journal.pone.0033470.
Butler, DG, BR Cullis, AR Gilmour, and BJ Gogel. 2009. [ASReml]-R Reference Manual.
Kanwarpal S. Dhugga, Robert B. Meeley, Carl R. Simmons. Secondary Wall Forming Genes from Maize and Uses Thereof. US20110239329 A1. 2011-09-29.
Yu, Jianming, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." Nature Genetics 38 (2): 203–8. doi:10.1038/ng1702.
Zhao, Qiao, Jin Nakashima, Fang Chen, Yanbin Yin, Chunxiang Fu, Jianfei Yun, Hui Shao, Xiaoqiang Wang, Zeng-Yu Wang, and Richard A. Dixon. 2013. "LACCASE Is Necessary and Nonredundant with PEROXIDASE for Lignin Polymerization during Vascular Development in Arabidopsis." The Plant Cell 25 (10): 3976–87. doi:10.1105/tpc.113.117770.
Zhou, Xiang, and Matthew Stephens. 2012. "Genome-Wide Efficient Mixed Model Analysis for Association Studies." Nature Genetics 44 (7): 821–24. doi:10.1038/ng.2310.

http://www.biomassforthefuture.org/

cirad