

A whole genome scan of cultivated sorghum (*Sorghum bicolor* L. Moench): Evolving diversity in the light of different marker systems



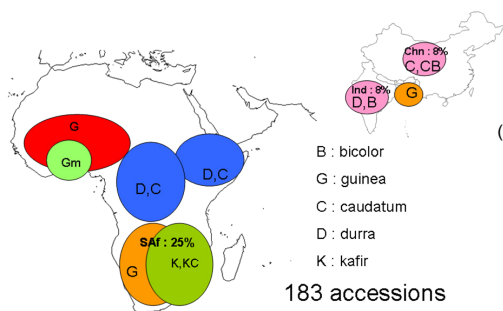
Jean-Christophe Glaszmann¹, Sophie Bouchet¹, Monique Deu¹, Jean-Francois Rami¹, Claire Billot¹, Ronan Rivallan¹, Laetia Gardes¹, Ling Xia², Andrej Kilian²

1 - CIRAD, Avenue Agropolis, 34398 Montpellier Cedex 5, FRANCE
2 - DArT PL, Cambia, AUSTRALIA



Sorghum diversity has been characterized over time with various generations of molecular markers. The first assessments of the extent of linkage disequilibrium place it between that of maize and that of *Arabidopsis thaliana*, making whole-genome scans realistic. We developed and characterized DArT markers for this purpose. A well documented Core Collection was genotyped and its structure compared according to various available data sets: 713 DArTs, 60 RFLPs and 40 SSRs. The three marker systems revealed similar patterns of diversity. The information retrieved by DArT appeared much closer to saturation than the other two systems. Compared to DArT and RFLP, SSR yielded lower stability across simulations and lower differentiation between the groups with higher residual diversity within the groups. The proportion of 'unclassified' accessions appeared consistently higher with DArT markers, suggesting that higher genome coverage reveals higher admixture, and pointing out recombination as a major source of sorghum germplasm diversity. Taking into consideration the structure level of the collection, linkage disequilibrium was weak, with a decrease within one Mb. Our DArT markers represent thus a significant contribution in search for adaptive variation in the sorghum genome.

A worldwide core collection with 3 marker systems



3 marker systems
SSR (GCP SP1)
RFLP (Deu et al. 2008)
DArTs (Figure 1)

A worldwide core collection of *S. Bicolor*
(Figure 2) representing all geographic origins as well as racial characteristics

The 3 marker systems agree for 6 groups

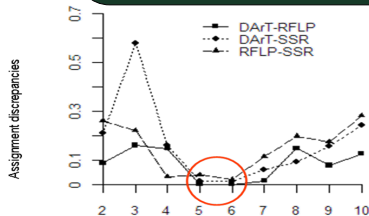


Figure 2. Repartition of the Core Collection samples according to race and geographic origins

The sorghum DArT genetic map

- 1 RIL population from CIRAD-INERA-IER
- 4 RILs from QDPI (Australia)
- 1 RIL from TAMU (USA)
- 1 consensus genetic map (Mace et al 2009)

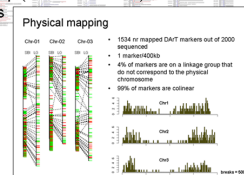


Figure 3. Proportion of assignment discrepancies between the three marker systems. Values are represented for a number of hypothetical ancestral genetic groups K on the X-axis, varying from 2 to 10 according to STRUCTURE software (Pritchard et al, 2000) assignments

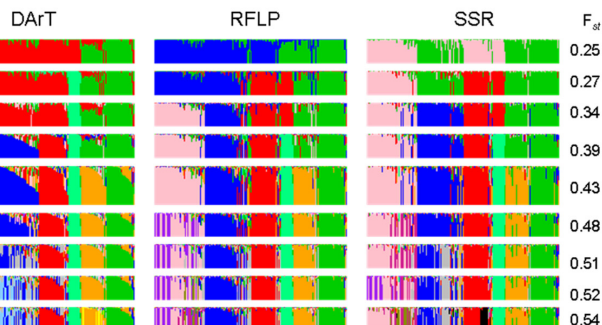


Figure 4. Genome representation

STRUCTURE software represents each sample in K dimensions, with K the number of hypothetical genetic groups that compose the collection. In this study, K varied from 2 to 10 from the top to the bottom. Three different data sets were tested from the left to the right: (A) 713 DArTs, (B) 60 RFLPs, (C) 40 SSRs. Each color corresponds to one genetic group. Each accession on the X-axis is represented by K colors ordered according to a decreasing genome fraction on the Y-axis. For each data set, 171 accessions were ordered according to DArT assignments and a decreasing proportion of genome assigned to the main group according to STRUCTURE at K=10.

The between-marker system group (color) correspondence was defined at K=10. The iterative first two groups having the most individuals in common received the same color. For each K value, colors were attributed according to groups that had the most accessions in common at K+1.

At K=6,
Group I in pink: D and B from India, C and CB from China.
Group II in blue: C and D from Africa.
Group III in red: G from western Africa
Group IV in spring green: Gm from western Africa
Group V in orange: G from southern Africa and Asia
Group VI in green: K and KC from southern Africa

Groups are congruent according to marker systems, allowing the construction of a balanced core collection

Linkage disequilibrium is weak in Sorghum with a decrease within one Mb

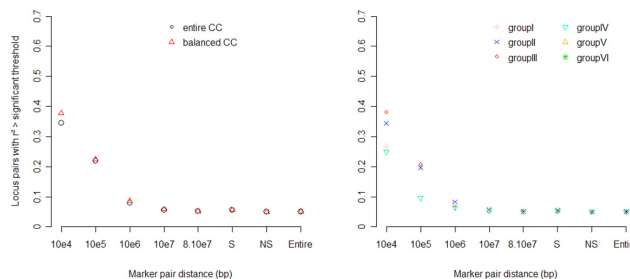
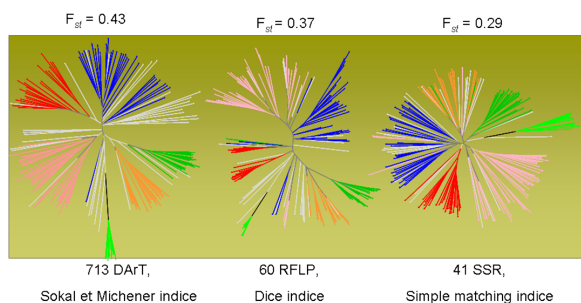


Figure 6. Linkage Disequilibrium over the genome

Proportion of significant long-range LD (r^2) between DArT loci.

Pair-wise DArT markers have been classified based on the intermarker physical distance d (bp) as follows : $d \leq 10^4$, $10^4 < d \leq 10^5$, $10^5 < d \leq 10^6$, $10^6 < d \leq 10^7$, $10^7 < d \leq 5 \cdot 10^7$, $5 \cdot 10^7 < d \leq 8 \cdot 10^7$.

"Entire" stands for total marker pairs.
"S" stands for marker pairs on the same chromosome.
"NS" stands for independent marker pairs.

LD analysis was performed on the total set of 177 accessions and on a minimum LD core collection comprising 100 accessions (balanced collection), and on each group (I to VI).

The significance threshold isolates the 5% highest r^2 among independent loci (distinct chromosomes). The respective significance thresholds were 0.11, 0.09, 0.15, 0.14, 0.25, 0.72, 0.28, 0.30.

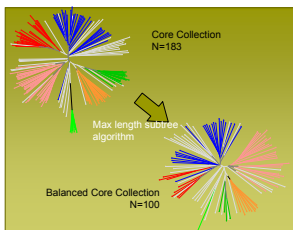


Figure 5. Selection of a balanced core collection (ie less structured)

100 accessions were chosen to represent the largest genetic diversity of the initial core-collection, and to reduce the structural linkage disequilibrium (DARwin software, Perrier et al.)