Increased Power to Dissect Adaptive Traits in Global Sorghum Diversity using a Nested Association Mapping Population

Relevant information for the study

Sophie Bouchet[*,1], Marcus O. Olatoye[*], Sandeep R. Marla[*], Ramasamy Perumal[*,†], Tesfaye Tesso[*], Jianming Yu[‡], Mitch Tuinstra[§], and Geoffrey P. Morris[*,2]

[*] Department of Agronomy, Kansas State University, Manhattan, Kansas 66506
[†] Kansas State University, Agricultural Research Center, Hays, Kansas 67601
[‡] Department of Agronomy, Iowa State University, Ames, Iowa 50011
[§] Department of Agronomy, Purdue University, West Lafayette, Indiana 47907
[1] Current address: GDEC, INRA, UBP, 63039, Clermont-Ferrand, France.
[2] Corresponding author: gpmorris@k-state.edu

Plant material

NAM seeds will be submitted to GRIN. Please contact corresponding author for availability.

Files for the following sections are available at the Dryad Digital Repository (doi:10.5061/dryad.gm073)

Raw Phenotype files

We phenotyped plant height (HT) for F_6 NAM RILs at two locations, in western Kansas (Hays KS, 38.8541°N 99.3385°W, semi-arid climate) and eastern Kansas (Manhattan, KS, 39.2125°N 96.5983°W, continental-humid climate) in 2014. Plant height was estimated as the mean of two representative plants per row measured using a barcoded ruler after physiological maturity. We also phenotyped flowering time (FT) in the Manhattan, KS experiment. Flowering time was defined as number of days until 50% of plants were in anthesis.

height_manhattan_2014.csv
height_hays_2014.csv
Flowering_Time_MH2014.csv

Filtered and normalized Phenotype files

FT_HT_NAM.csv
FT_SAP_weslaco.csv
FT_SAP_lubbock.csv
FT_GRIN.csv
HT_SAP.csv

Raw Genotype files

fastq files were submitted to NCBI: SRA NCBI Project accession SRP095629.
Raw genotypes were obtained from fastq files. Reads were trimmed, and genotypes were called and filtered using TASSEL 5 pipeline GBS v2 Pipeline (https://bitbucket.org/tasseladmin/tassel-5-source/wiki/Tassel5GBSv2Pipeline) (Glaubitz *et al.* 2014).
NAM population: nam.hmp.txt.zip
diverse association panels: div.hmp.txt.zip

Filtered and imputed genotype files
For the NAM RILs, raw genotypes were filtered for tag coverage (> 5% of taxa), minor allele frequency (MAF) (> 0.03), and single marker missing data (< 0.8), and 90,441 SNPs were retained for further analysis. Missing genotypes were imputed using the FSFHap Plugin (Swarts *et al.* 2014) implemented in TASSEL 5, which corrects genotyping errors for inbred individuals in full-sib families. We removed 96 RILs with more than 10% residual heterozygosity and retained 2214 RILs for further analyses (NAM2214).
For the diverse association panels, taxa with call rate > 95%, markers with MAF > 0.01 and missing data < 0.8 were retained. Missing genotypes were imputed using Beagle 4 (Browning and Browning 2013), which is more accurate than FSFHap for diverse germplasm (Swarts *et al.* 2014).

For the NAM QTL mapping, 90441 markers with MAF > 0.03 were used:
Tassel format: nam_composite_FT.hmp.txt.zip
Plink format: nam_composite.ped

For the GRIN700 panel, 204557 markers with MAF > 0.01 were used:
geno_grin.csv

For the SAP340 diversity panel, 190555 markers with MAF > 0.01 were used:
geno_wes.csv
geno_lub.csv

For simulations, 60864 that were polymorphic in the NAM and diversity panels were used:
geno_grin.map
geno_grin.ped
geno_nam.map
geno_nam.ped
geno_sap.map
geno_sap.ped

Marker information files
map.composite.csv

Results files - Dryad Digital Repository (doi:10.5061/dryad.gm073)
genome scan
For each marker, information about genic region (CDS, 3'_UTR, 5'_UTR) and distance from closest gene: info_marker_gene.csv
Heterozygosity rate in the NAM population and within each family: heterozygosity.csv
Proportion of RILs with Tx430 allele in the NAM population and within each family: prop_Tx430_distorsion.csv
Recombination rate in the NAM population and within each family: recombination.csv
Monomorphic status of each marker in each population: monomorphism.csv

To map QTL in the NAM population (NAM2214), we used joint linkage (JL) (Würschum *et al.* 2012) or multi-locus linear regression model (MLLM) (Giraud *et al.* 2014).
HA stands for Hays. MN for Manhattan. For the diverse association panels (SAP340 and

GRIN700), GWAS were performed using a forward–backward stepwise multi-locus mixed model (MLMM) (Segura *et al.* 2012).

joint linkage analyses
HT_HA2014_JL.csv
HT_MN2014_JL.csv
FT_MN2014_JL.csv

multi-locus linear regression model
HT_SAP_MLLM.csv
FT_SAP_LUBBOCK_MLLM.csv
FT_SAP_WESLACO_MLLM.csv
FT_GRIN_MLLM.csv
HT_NAM_MN2014_MLLM.csv
HT_NAM_HA2014_MLLM.csv
FT_NAM_MN2014_MLLM.csv

Source code for simulations
To compare the power of QTL detection using NAM versus GWAS, we studied simulated QTL. First, 50 random samples of 50 SNPs (from the 60K SNPs) were assigned as QTL with additive effects following a geometric series (Lande and Thompson 1990; Yu *et al.* 2008). The genotypic value of each RIL was defined as the sum of genotypic values across all loci. The entry-mean heritability (h^2) was set to either 0.4 or 0.7. Phenotypic values of RILs were obtained by adding normally-distributed error to the genotypic values such that the residual variation was 60% (h^2 = 0.4) or 30% (h^2 = 0.7) of phenotypic variation. Power was compared between a diverse association panel (using MLMM) and the NAM population (using JL or MLLM). The power was calculated for 50 independent runs and then averaged for each simulation scheme.
script_simulation.r