

BACKGROUND

Genomic structural mutations are an important source of variation in many species and can play key roles in phenotypic diversification and evolution. Previous work in many plant species, including some crops, has identified multiple instances of structural variations (SVs) occurring in or near genes related to stress response and disease resistance, suggesting a possible role for SVs in local adaptation. Sorghum (Sorghum bicolor (L.) Moench) is the fifth most important crop in the world, and over the course of its history it has been adapted to a wide array of different climates as well as bred for multiple purposes, resulting in a striking phenotypic diversity within the existing germplasm, but currently little is known about what role SVs may have played in sorghum's diversification.

OBJECTIVES

- Examine whether or not SVs in sorghum are under positive or negative selection.
- Examine the extent to which particular SVs are associated with the geographic origins of each genotype
- Determine if any segionally-specific SVs occurred in genes with functions potentially involved in local adaptation

Genomic Patterns of Structural Variation in the Sorghum Biomass Association Panel

Kittikun Songsomboon and Elizabeth Cooper

Department of Bioinformatics and Genomics, College of Computing and Informatics, University of North Carolina at Charlotte

METHODS

Illumina resequencing data of 347 diverse genotypes with ~30X coverage on average was processed as follows:

- Raw data were aligned with speedseq
- SVs were called by sytools as described in

ttps://github.com/hall-lab/sytools/blob/master/Tutorial.md

 SV calls were filtered and analyzed as described in https://github.com/skittikun/SVs_in_sorghum_BAP

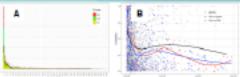


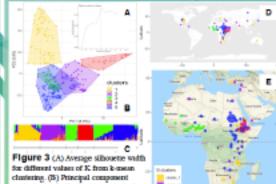
Figure 2 (A) Global site frequency spectrum of SVs. (B) Genome-wide decay of linkage disequilibrium (LD).

RESULTS

The total number of SVs before any filters was 1,468,572 SVs. After applying filters, we retained 24,648 SVs. There were 8,170 SVs with a minor allele frequency > 5%, which we used to plot Figure 1. The outer ring of Figure 1 shows the overall distribution of SVs where each dash represents a single SV. The blue line in the outermost ring shows the average density of SVs in 500 kb sliding windows. The middle ring shows the relative abundance of SVs in 500 kb sliding windows within each of the five groups identified by k-means clustering. The innermost ring is the significance (log(p)) value of each 500 kb window, indicating whether or not the relative abundance in the window is significantly different from the expected distribution of SVs among the five groups. Most SVs are rare and appear to be under neutral selection based on the global site frequency spectrum, but the SVs within CDS regions show increased linkage disequilibrium (LD) compared to the genomewide rate (Figures 2A and 2B). Both K-means clustering and ADMIXTURE analyses separated the genotypes into five groups (Figures 3A, 3B and 3C). These groups strongly corresponded with the known geographic distributions and historical races of genotypes (Figures 3D and 3E). For example, cluster 3 is exclusive to the guinea race in west Africa. There were 1,570 cluster-specific SVs and 120 of them occurred within genes, several of which have annotated functions related to local adaptation (Table 1).

HIGHLIGHTS

- Among 347 sorghum genotypes, we found 8,170 SVs with MAF > 5% distributed across all 10 chromosomes, and primarily located around the gene-rich high recombination regions.
- K-means clustering and ADMIXTURE based on SVs grouped these genotypes into five clusters which corresponded to both geography and race.
- Among 1,570 cluster-specific SVs, 120 of them potentially affected gene functions. Some of these affected genes had functions known to be related to local adaptation.



for different values of K from k-mean chastering. (B) Principal component analysis plot for the first two principal components. (C) Cluster membership for each individual in the 5 clusters identified

by both ADMIXTURE and k-means clustering. (D) The geographic distributions of populations semipled in this study, with colors corresponding to the cluster membership of each population. (P) The distribution and cluster membership of African populations.

Table 1 Cluster-specific SVs affecting gene functions with local adaptation potentia

	17 mars	Terms.	Ten market	Transaction between the contrader to
į		The Real Property lies	The second second second second second	-
				Personal Print Services
Ĭ			COLUMN TO SERVICE AND ADDRESS.	ALC: N
	THE RESERVE OF THE PARTY OF THE PARTY.		INCOLDED SPACE SPACE SPACE	THE RESERVE
	the state of the s	_	CONTRACTOR OF THE PARTY.	
-	CONTRACTOR AND ADDRESS OF THE PARTY OF THE P			THE RESERVE
	mark to the second			THE PARTY NAMED IN
-	THE RESERVE OF THE PARTY OF THE			THE REAL PROPERTY.
	Mark Street Street			
	BOARD OF BUILDING		ACCORDING A CASE AND ADDRESS.	THE PERSON
	The April 2 of the State of the		Mission or departments where all while	
	OF REAL PROPERTY.			City and Control of the Control of t
			CO was the box beat making real	DOM: NO
_	BOTH THE RESERVE			CONTRACTOR OF THE PARTY OF THE
	BALL OF THE PARTY			
			and the contract of the latter posts.	Conglications.
_			THE RESERVE AND ADDRESS OF THE PARTY OF THE	