

**SUPPLEMENTARY FIGURES AND TABLES**

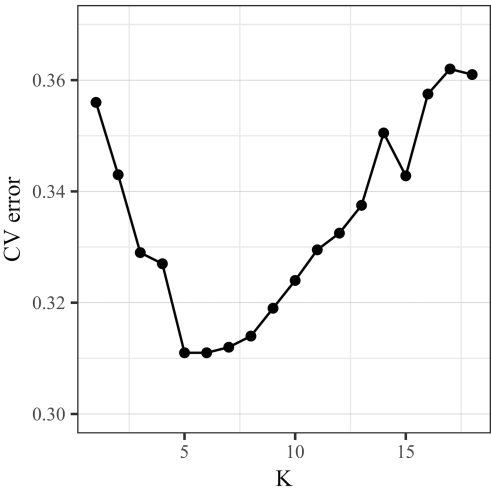
**Impact of Sorghum Racial Structure and Diversity on Genomic Prediction of Grain Yield  
Components**

Sirjan Sapkota,\* Richard Boyles, Elizabeth Cooper, Zachary Brenton, Matthew Myers, and  
Stephen Kresovich

Affiliations: S. Sapkota, Z. Brenton, M. Myers and S. Kresovich, Advanced Plant Technology  
Program, Clemson University, Clemson, SC 29634; S. Sapkota, R. Boyles, and S. Kresovich,  
Department of Plant and Environmental Sciences, Clemson University, Clemson SC 29634; R.  
Boyles, Pee Dee Research and Education Center, Clemson University, Florence, SC 29506; E.  
Cooper, Department of Bioinformatics and Genomics, University of North Carolina at Charlotte,  
Charlotte, NC 28223.

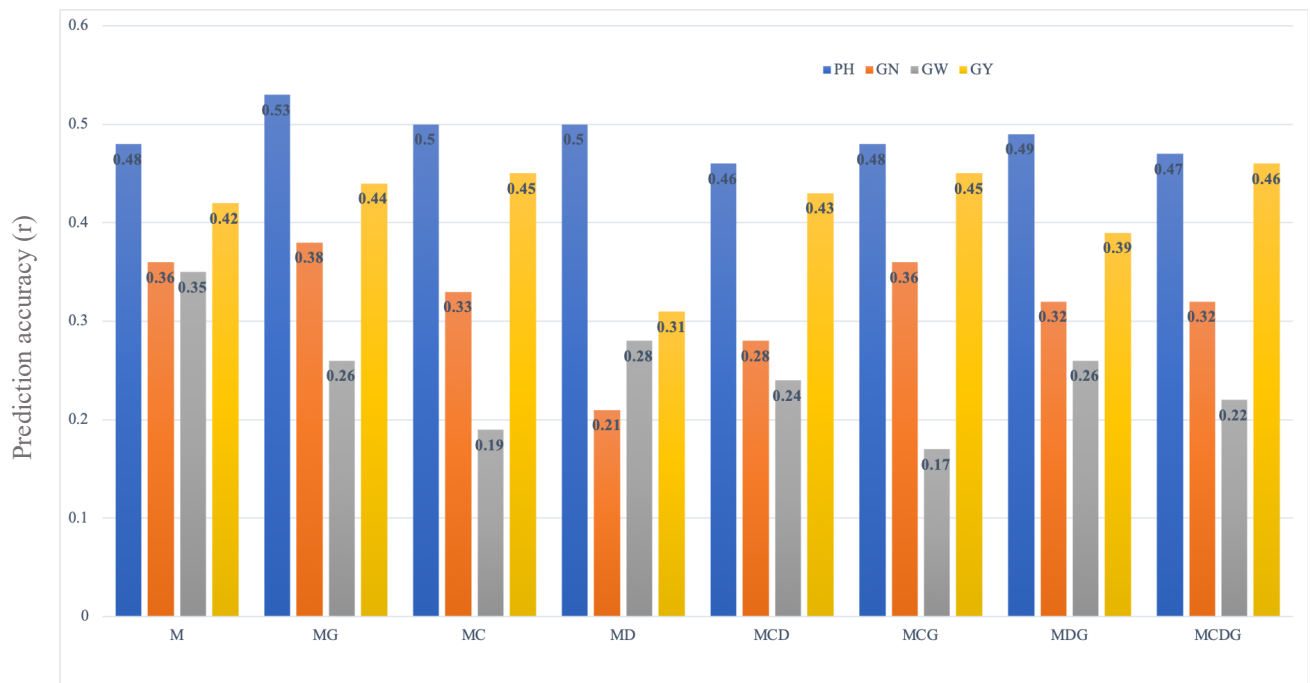
15     **Supplementary Figure S1.** Cross validation (CV) of error for number of clusters ( $K = 1$  to 18)

16     in admixture.

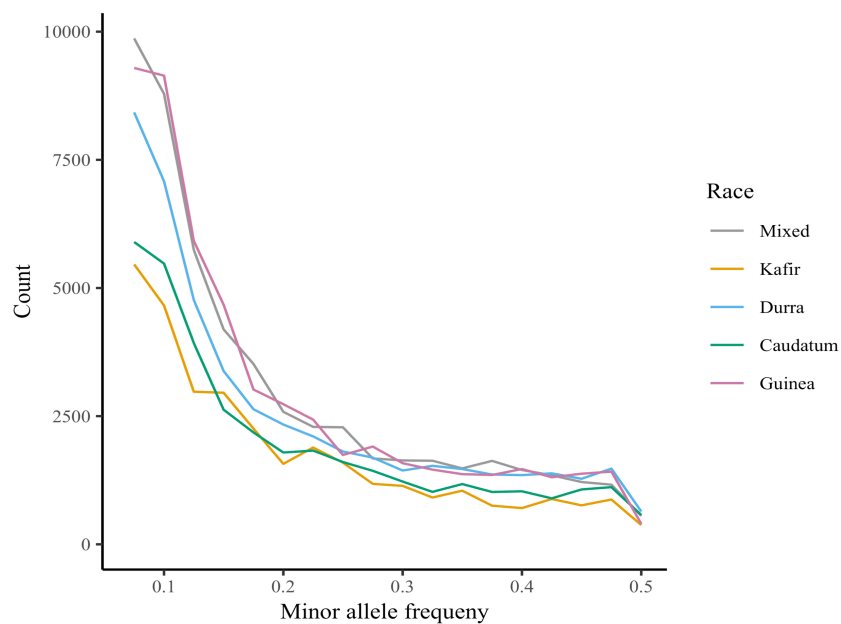


17

**Supplementary Figure S2.** Cross-validation accuracy for 36 kafir accessions using combination of one or many other races as training population. Method: We randomly sampled 36 kafir accessions and used it as validation population for the all cross-validation runs. The training population size used was always 36 individuals. A total of 36, 18, 12, and nine individuals from each training race was used when one, two, three, and four training race were used, respectively. The alphabets in x-axis represent races: M = mixed, C = caudatum, D = durra, and G = guinea. Colors in bars represent traits: GN= grain number, GW = grain weight, GY = grain yield, and PH = plant height. The respect numerical values for prediction accuracies are shown at the top of the bars.



28 **Supplementary Figure S3.** Allele frequency spectra for minor alleles across accessions within  
29 each race. The minor alleles were plotted at *binwidth* of 0.02 with total number (count) of alleles  
30 with that minor allele frequency.



**Supplementary Table S1.** Genetic differentiation between the races based on  $F_{st}$  (bottom and left of the diagonal) and Euclidean distance between centroids (top and right of the diagonal) of race clusters from first five principal components.

	Mixed	Kafir	Durra	Caudatum	Guinea
Mixed		41.12	38.2	30.07	36.03
Kafir	0.08		61.34	60.79	63.12
Durra	0.07	0.16		54.76	55.32
Caudatum	0.05	0.15	0.13		54.71
Guinea	0.07	0.17	0.12	0.14	

**Supplementary Table S2.** Mean and standard deviation of all phenotypic traits by sorghum races. BL, terminal branch length; DTA, days to anthesis; FLH, flag leaf height; GN, grain number; GW, grain weight; GY, grain yield; PH, plant height; PL, panicle length.

Race	DTA	PH	GN	GW	GY	FLH	PL	BL
		cm		g	g	cm	cm	cm
Caudatum	75 ± 8	119.74 ± 43.18	1472 ± 407	24.99 ± 4.87	47.13 ± 13.47	90.15 ± 40.91	19.93 ± 4.04	6.31 ± 1.94
Durra	78 ± 10	124.18 ± 47.59	1118 ± 362	21.53 ± 5.66	33.05 ± 12.55	87.2 ± 38.89	19.84 ± 6.8	7.56 ± 4.40
Guinea	70 ± 8	127.02 ± 49.08	1032 ± 423	25.21 ± 6.69	35.06 ± 17.08	81.97 ± 19.36	28.73 ± 6.44	9.74 ± 3.13
Kafir	71 ± 6	130.74 ± 55.9	1404 ± 409	22.07 ± 4.25	39.6 ± 11.69	95.9 ± 47.08	21.87 ± 4.07	7.37 ± 1.78
Mixed	71 ± 8	131.21 ± 50.38	1228 ± 419	21.69 ± 6.12	34.08 ± 12.34	92.68 ± 36.8	23.55 ± 6.08	8.36 ± 2.99
Total	74 ± 8.5	125.2 ± 48.2	1301 ± 433	23.23 ± 5.6	39.48 ± 14.5	90.11 ± 39.3	21.69 ± 6	7.43 ± 3.07

**Supplementary Table S3.** Mean estimates of variance and covariance for WR and AR prediction in CV2 method. AR, Across race; WR, within race.

Trait	covariance		variance (predicted values)	
	WR	AR	WR	AR
Terminal branch length (BL)	271.53	35.42	216.38	34.67
Days to anthesis (DTA)	5.95	1.31	6.06	1.84
Flag leaf height (FLH)	294.55	63.44	268.16	31.94
Grain number per panicle (GN)	8419.33	7350.35	8115.08	5448
1000-grain weight (GW)	8.5	2.16	6.17	1.22
Grain yield per panicle (GY)	20.31	13.14	14.68	9.14
Plant height (PH)	503.43	236.2	455.17	117.51
Panicle length (PL)	4.9	0.75	4.26	1.14

**Supplementary Table S4.** Mean prediction accuracies by races for WR and AR prediction in CV2 method. Higher mean prediction accuracy for each trait across all races and methods is highlighted in bold. Values represent mean  $\pm$  standard deviation.

Traits	Caudatum		Durra		Guinea		Kafir		Mixed	
	WR	AR	WR	AR	WR	AR	WR	AR	WR	AR
Primary branch length	0.21 $\pm$ 0.04	0.03 $\pm$ 0.20	<b>0.81</b> $\pm$ 0.06	0.25 $\pm$ 0.28	0.23 $\pm$ 0.09	0.23 $\pm$ 0.39	0.26 $\pm$ 0.11	0.10 $\pm$ 0.24	0.39 $\pm$ 0.05	0.38 $\pm$ 0.28
Days to anthesis	<b>0.55</b> $\pm$ 0.02	0.34 $\pm$ 0.21	0.33 $\pm$ 0.06	0.00 $\pm$ 0.25	-0.07 $\pm$ 0.08	0.00 $\pm$ 0.44	0.06 $\pm$ 0.07	0.06 $\pm$ 0.25	0.33 $\pm$ 0.04	0.18 $\pm$ 0.25
Flag leaf height	0.48 $\pm$ 0.03	0.47 $\pm$ 0.19	<b>0.66</b> $\pm$ 0.03	0.50 $\pm$ 0.24	0.00 $\pm$ 0.10	0.01 $\pm$ 0.39	0.46 $\pm$ 0.05	0.42 $\pm$ 0.24	0.44 $\pm$ 0.05	0.32 $\pm$ 0.32
Grain number	0.36 $\pm$ 0.03	0.21 $\pm$ 0.18	0.33 $\pm$ 0.06	0.09 $\pm$ 0.25	0.18 $\pm$ 0.12	0.34 $\pm$ 0.41	0.13 $\pm$ 0.07	0.28 $\pm$ 0.24	0.26 $\pm$ 0.09	<b>0.42</b> $\pm$ 0.25
1000-grain weight	0.54 $\pm$ 0.03	0.29 $\pm$ 0.21	0.62 $\pm$ 0.01	0.45 $\pm$ 0.28	<b>0.77</b> $\pm$ 0.02	0.48 $\pm$ 0.31	0.65 $\pm$ 0.02	0.22 $\pm$ 0.24	0.49 $\pm$ 0.04	0.42 $\pm$ 0.20
Grain yield	0.32 $\pm$ 0.04	0.19 $\pm$ 0.18	0.43 $\pm$ 0.05	0.24 $\pm$ 0.27	<b>0.53</b> $\pm$ 0.05	0.44 $\pm$ 0.41	0.28 $\pm$ 0.06	0.43 $\pm$ 0.19	0.23 $\pm$ 0.09	0.46 $\pm$ 0.27
Plant height	0.55 $\pm$ 0.03	0.55 $\pm$ 0.16	<b>0.62</b> $\pm$ 0.03	0.54 $\pm$ 0.20	0.21 $\pm$ 0.06	0.27 $\pm$ 0.42	0.54 $\pm$ 0.04	0.51 $\pm$ 0.19	0.37 $\pm$ 0.05	0.40 $\pm$ 0.24
Panicle length	0.25 $\pm$ 0.07	0.02 $\pm$ 0.17	<b>0.79</b> $\pm$ 0.04	0.29 $\pm$ 0.27	0.02 $\pm$ 0.13	0.10 $\pm$ 0.46	-0.07 $\pm$ 0.09	-0.03 $\pm$ 0.27	0.25 $\pm$ 0.07	0.21 $\pm$ 0.29
Average	0.41 $\pm$ 0.14	0.26 $\pm$ 0.26	0.57 $\pm$ 0.18	0.30 $\pm$ 0.31	0.23 $\pm$ 0.28	0.23 $\pm$ 0.44	0.29 $\pm$ 0.24	0.25 $\pm$ 0.30	0.35 $\pm$ 0.11	0.35 $\pm$ 0.28