

# **Genome-wide association and gene network analysis for starch and protein in sorghum**

Sirjan Sapkota<sup>1,2,\*</sup>, J. Lucas Boatwright<sup>2</sup>, Kathleen Jordan<sup>2</sup>, Richard Boyles<sup>1,3</sup>, and Stephen Kresovich<sup>1,2</sup>,

1 Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA

2 Advanced Plant Technology Program, Clemson University, Clemson, SC, USA

3 Pee Dee Research and Education Center, Clemson University, Florence, SC, USA

\* correspondence: [ssapkot@g.clemson.edu](mailto:ssapkot@g.clemson.edu)

Supplementary File

**Supplementary Table S1.** Variance components for the linear mixed models fit using genetic and environmental variables.

	Intercept	Geno	Year	Geno × Year	Year × Rep	Residual	Total	H <sup>2</sup>
Starch	68.272	3.364	0	1.07	0.141	2.922	7.497	0.799
Protein	12.044	0.869	0.494	0.279	0.066	1.155	2.863	0.753

**Supplementary Table S2.** Pariwise linkage disequilibrium between association SNPs and their neighboring SNPs. R<sup>2</sup>: correlation coefficient, Chr: Chromosome. SNP: single nucleotide polymorphism.

Chr SNP1	Position SNP1	Chr SNP2	Position SNP2	R <sup>2</sup>
4	60623675	4	60577500	0.012
4	60623675	4	60623655	0.513
4	60623675	4	60623675	1
4	60623675	4	60624201	0.941
4	60623675	4	60624444	0.921
4	60623675	4	60626903	0.001
4	63400335	4	63380566	0.004
4	63400335	4	63400335	1
4	63400335	4	63426489	0.031
4	64019590	4	64018526	0.009
4	64019590	4	64019122	0.745
4	64019590	4	64019577	1
4	64019590	4	64019590	1
4	64019590	4	64019619	1
4	64019590	4	64028856	0.014
8	51715166	8	51695627	0.057
8	51715166	8	51715166	1
8	51715166	8	51719632	0.219
8	51715166	8	51719659	0.225
8	51715166	8	51719688	0.216
8	51715166	8	51719704	0.993
8	51715166	8	51720767	0.938
8	51715166	8	51721062	0.816
8	51715166	8	51721065	0.794
8	51715166	8	51726098	0.812
8	51715166	8	51727032	0.061

**Supplementary Table S3.** KEGG Pathway enrichment results for significantly associated genomic regions. FDR: false discovery rate, Observed and Expected refer to the gene count for the network.

Region	Pathway	Observed	Expected	FDR	Genes
Chr-4	Nitrogen metabolism	3	33	1.6E-06	Sb04g024300,Sb04g034470,Sb07g022750
Chr-4	Biosynthesis of unsaturated fatty acids	2	44	0.0004	Sb04g029900,Sb04g029920
Chr-4	Fatty acid metabolism	2	80	0.00085	Sb04g029900,Sb04g029920
Chr-8	Protein processing in endoplasmic reticulum	27	200	1.58E-30	Sb01g005860,Sb01g010460,Sb01g011310,Sb01g013390,Sb01g039390,Sb01g039450,Sb01g039460,Sb01g039470,Sb01g039480,Sb01g039530,Sb01g039780,Sb02g020380,Sb02g021850,Sb02g029650,Sb03g039360,Sb03g041830,Sb04g001140,Sb04g027330,Sb04g030160,Sb06g015020,Sb07g028940,Sb08g003340,Sb08g009580,Sb08g016560,Sb08g018750,Sb09g022580,Sb10g030240
Chr-8	Spliceosome	15	189	5.42E-14	Sb01g011310,Sb01g039390,Sb01g039450,Sb01g039460,Sb01g039470,Sb01g039480,Sb01g039530,Sb01g040250,Sb02g021850,Sb03g039360,Sb03g044450,Sb08g009580,Sb08g015280,Sb08g018750,Sb09g022580
Chr-8	Endocytosis	12	150	1.76E-11	Sb01g011310,Sb01g039390,Sb01g039450,Sb01g039460,Sb01g039470,Sb01g039480,Sb01g039530,Sb02g021850,Sb03g039360,Sb08g009580,Sb08g018750,Sb09g022580
Chr-8	RNA degradation	7	109	1.84E-06	Sb01g017050,Sb01g020010,Sb01g041170,Sb02g028570,Sb04g000370,Sb09g026970,Sb10g001120
Chr-8	Plant-pathogen interaction	6	162	0.00019	Sb03g006570,Sb03g028430,Sb04g021850,Sb07g028940,Sb08g016560,Sb10g030240
Chr-8	Protein export	4	54	0.00022	Sb01g010460,Sb03g041830,Sb04g001140,Sb08g003340