Final Scientific/Technical Report

Award number: DOE Award DE-SC0006995

Project title: Discovering the desirable alleles contributing to the lignocellulosic biomass traits in *Saccharum* germplasm collections for energy cane improvement

Recipient: University of Florida.

PI: Jianping Wang

3. Provide an executive summary, which includes a discussion of 1) how the research adds to the understanding of the area investigated; 2) the technical effectiveness and economic feasibility of the methods or techniques investigated or demonstrated; or 3) how the project is otherwise of benefit to the public. The discussion should be a minimum of one paragraph and written in terms understandable by an educated layman.

1) The success in crop improvement programs depends largely on the extent of genetic variability available. Germplasm collections assembles all the available genetic resources and are critical for long-term crop improvement. This world sugarcane germplasm collection contains enormous genetic variability for various morphological traits, biomass yield components, adaptation and many quality traits, prospectively imbeds a large number of valuable alleles for biofuel traits such as high biomass yield, quantity and quality of lignocelluloses, stress tolerance, and nutrient use efficiency. The germplasm collection is of little value unless it is characterized and utilized for crop improvement. In this project, we phenotypically and genotypically characterized the sugarcane world germplasm collection (The results were published in two papers already and another two papers are to be published). This data will be made available for public to refer to for germplasm unitization specifically in the sugarcane and energy cane breeding programs. In addition, we are identifying the alleles contributing to the biomass traits in sugarcane germplasm. This part of project is very challenging due to the large genome and highly polyploid level of this crop. We firstly established a high throughput sugarcane genotyping pipeline in the genome and bioinformatics era (a paper is published in 2016). We identified and modified a software for genome-wide association analysis of polyploid species. The results of the alleles associated to the biomass traits will be published soon, which will help the scientific community understand the genetic makeup of the biomass components of sugarcane. Molecular breeders can develop markers for marker assisted selection of biomass traits improvement. Further, the development and release of new energy cane cultivars through this project not only improved genetic diversity but also improved dry biomass yields and resistance to diseases. These new cultivars were tested on marginal soils in Florida and showed very promising yield potential that is important for the successful use of energy cane as a dedicated feedstock for lignocellulosic ethanol production.
2) Multiple techniques at different project progress stages were utilized. For example, for the whole world germplasm accession genotyping, a cheap widely used SSR marker genotyping platform was utilized due to the large number of samples (over thousand). But the throughput of this technique is low in generating data points. However, the purpose the genotyping is to form a core collection for further high throughput genotyping. Thus the results from the SSR genotyping was quite good enough to generated the core collection. To genotype the few hundred core collection accessions, an target enrichment sequencing technology was used, which is not only high throughput in generating large number of genotyping data, but also has the candidate genes targeted to genotyping. The data generated would be sufficient in identifying the

alleles contributing to the traits of interests. All the techniques used in this project are effective though extensive time was invested specifically for establish the pipeline in the experimental design, data analysis, and different approach comparison.
3) the research can benefit to the public in polyploid genotyping and new and cost efficient genotyping platform development

4. Provide a comparison of the actual accomplishments with the goals and objectives of the project.

The actual accomplishments were comparable with the goals and objectives of the project except the timeline was extended due to the super long growth cycle of sugarcane, unexpected CoPI moving, retirement, government shut down, and new data analysis pipeline establishment.

5. Summarize project activities for the entire period of funding, including original hypotheses, approaches used, problems encountered and departure from planned methodology, and an assessment of their impact on the project results. Include, if applicable, facts, figures, analyses, and assumptions used during the life of the project to support the conclusions.

1) Phenotypically and genotypically characterize the sugarcane world gerrmplasm collection.

The World Collection of Sugarcane and Related Grasses (WCSRG) in Miami, FL contains diverse and potentially useful germplasm for this and related genera; however, this collection has been underutilized because little is known about the traits of its accessions. Our objectives were to phenotypically characterize the World Collection and select a representative core collection that could then be studied intensively. In total, eight morphological traits of the WCSRG (Table 1) were evaluated three times in 1 year (Figure 1-8). A core of 300 accessions that included each species in the World Collection was selected by using the Maximization Strategy in MStrat software. The core had a higher diversity rating than random selections of 300 accessions (Table 2). The Shannon–Weaver Diversity Index scores of the core and whole collection were similar indicating that the majority of the diversity was captured by the core collection. The ranges and medians between the core and World Collection were similar; only two of the trait medians were not significant at $P = 0.05$ using the non-parametric Wilcoxon method and the coincidence rate (CR % = 96.2) was high (>80) indicating that extreme values were retained. Thus, the phenotypic diversity of these traits in the World Collection was well represented by the core collection. Agronomic studies on the core should be useful for characterizing the World Collection and genes for useful traits should be available in the core collection.

Table 1 Number of accessions by taxon within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses (WCSRG) in Miami, Florida and the core collection

Genus	Species		WCSRG	Core
Coix	*gigantea* (*lacryma-jobi*)	J. Koenig	1	1
Erianthus	sp.		20	8
Imperata	sp.		1	1
Miscanthus	floridulus	Labill.	3	2
Miscanthus	Hybrid		6	2
Miscanthus	*sinensis*	Anderss.	4	3
Miscanthus	unknown		5	2
Narenga	*porphyrocoma*	(Hance) Bor	1	1
Saccharum	*spontaneum*	L.	316	68
Saccharum	*officinarum*	L.	254	64
Saccharum	Hybrid		140	32
Saccharum	*robustum*	Brandes et Jesw. ex Grassl	51	12
Saccharum	*barberi*	Jeswiet	38	8
Saccharum	Unknown		36	14
Saccharum	*sinense*	Roxb.	33	11
Saccharum	*arundinaceum*	Retz.	8	1
Saccharum	*bengalense*	Retz.	5	3
Saccharum	*edule*	Hassk.	6	1
Saccharum	*ravennae*	(L.) Murr. (*Erianthus ravennae* (L.) Beauv.)	3	2
Saccharum	*kanashiroi*	(Ohwi) Ohwi	2	1
Saccharum	*rufipilum*	Steud.	2	1
Saccharum	*brevibarbe*	(Michx.) Pers.	1	1
Saccharum	*procerum*	Roxb.	1	1
Sorghum	*plumosum*	(R. Br.) P. Beauv.	1	1
Sorghum	*arundinaceum*	(Desv.) Stapf	1	1
Unknown	Unknown		238	58
Total			1,177	300

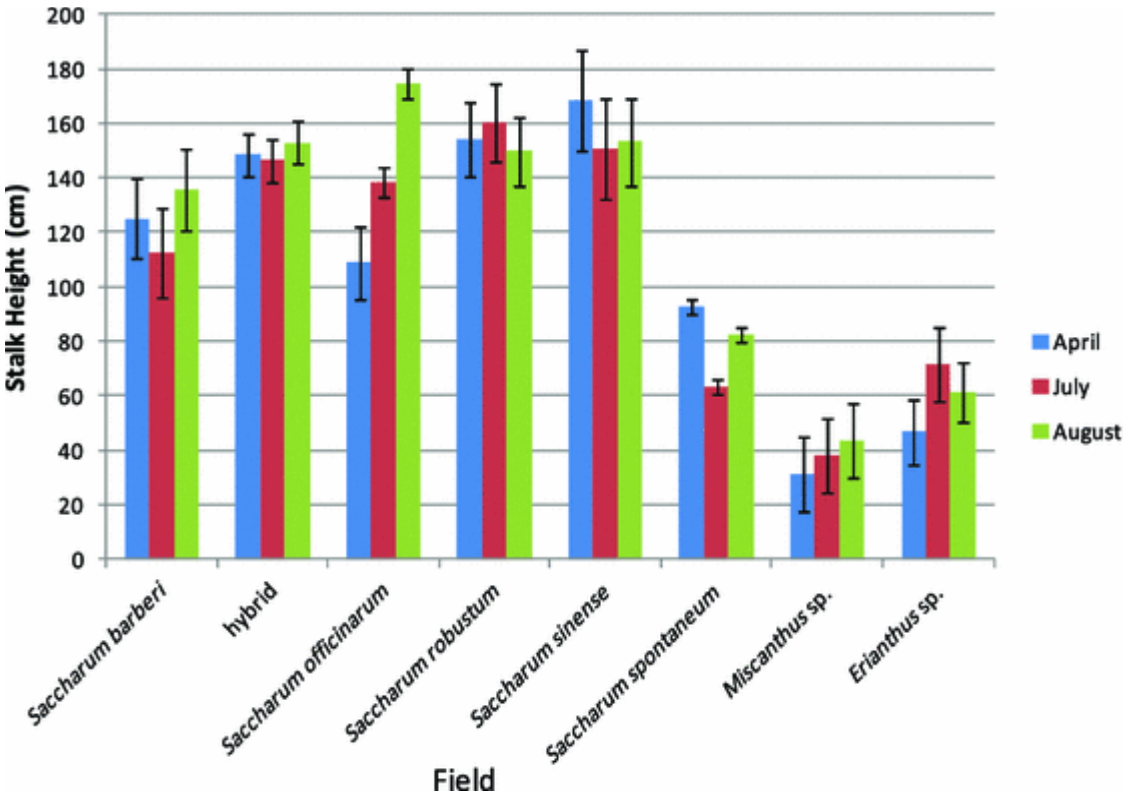

Fig. 1. Mean stalk height (cm) of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars*

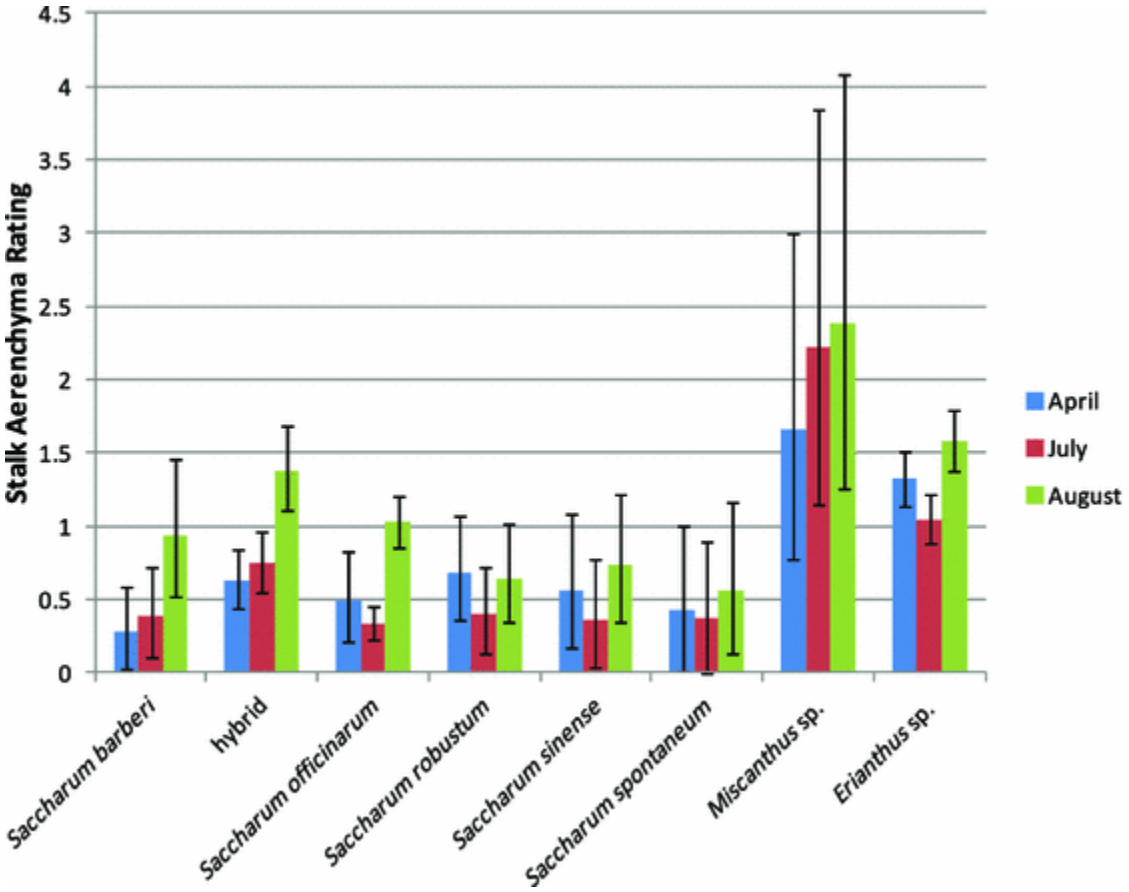

Fig. 2. Mean stalk aerenchyma rating of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars.*

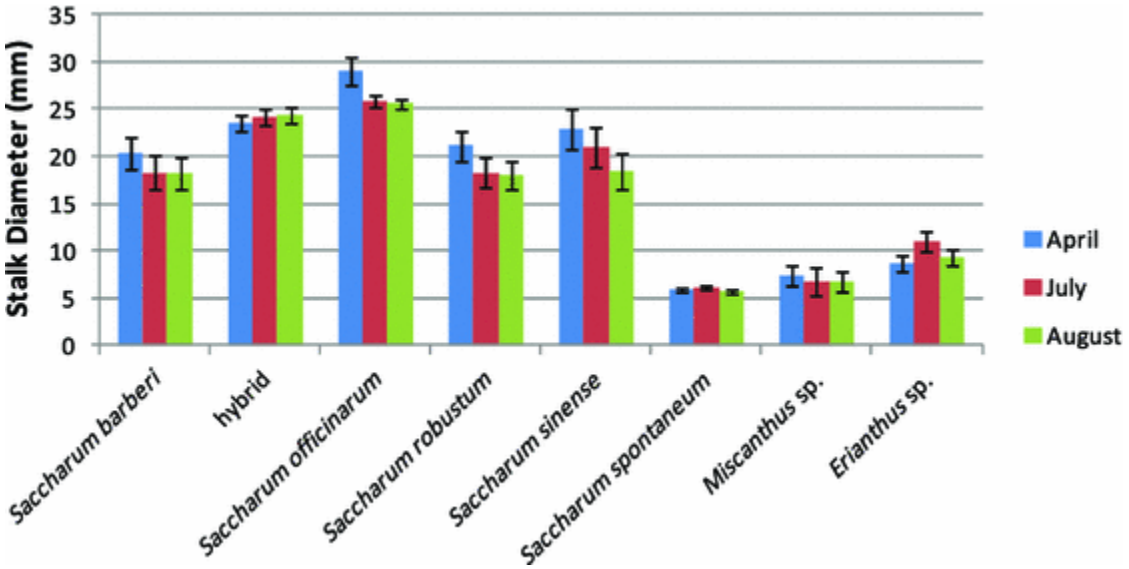

Fig. 3. Mean stalk diameter (mm) of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars*.

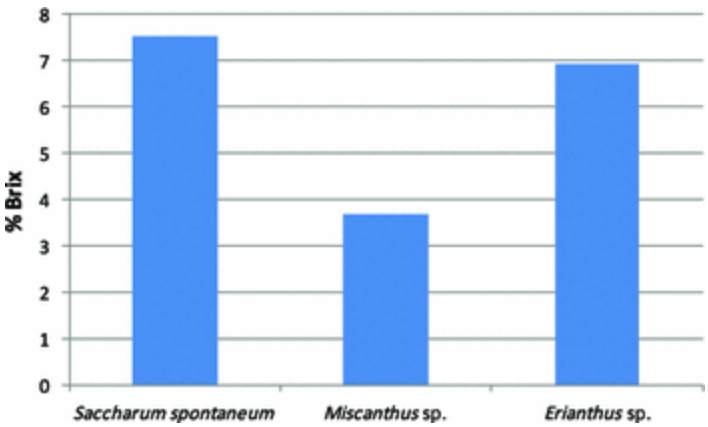

Fig. 4. Mean percentage Brix in April of juice extracted from stalks of species with >10 accessions within the container portion of the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL.

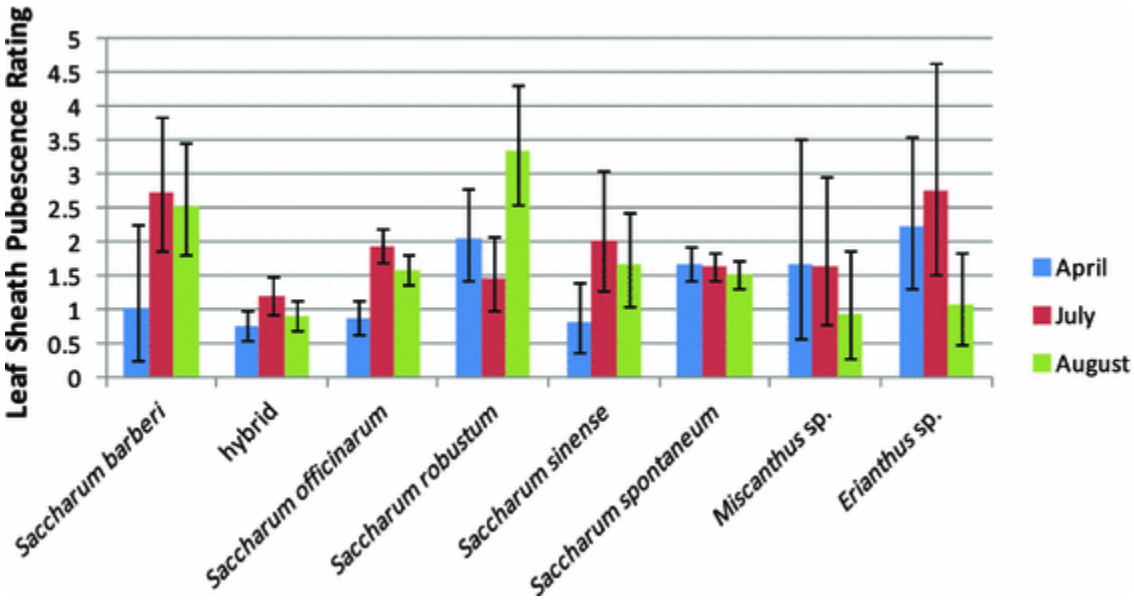

Fig. 5. Mean leaf sheath pubescence rating of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars*

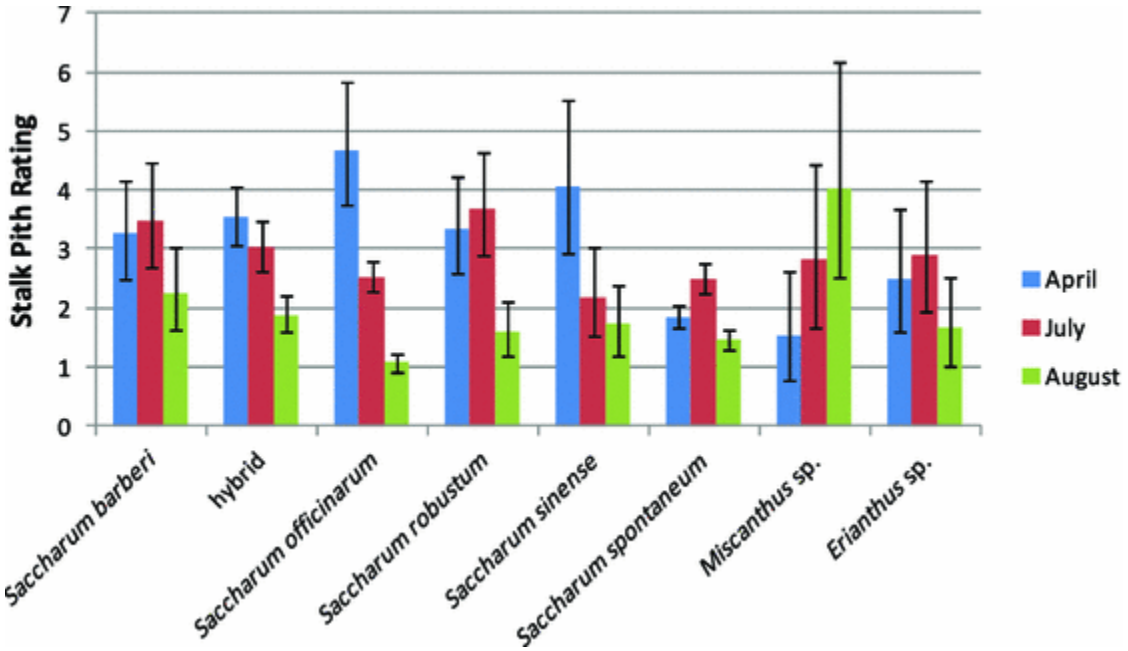

Fig. 6. Mean stalk pith rating of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars*

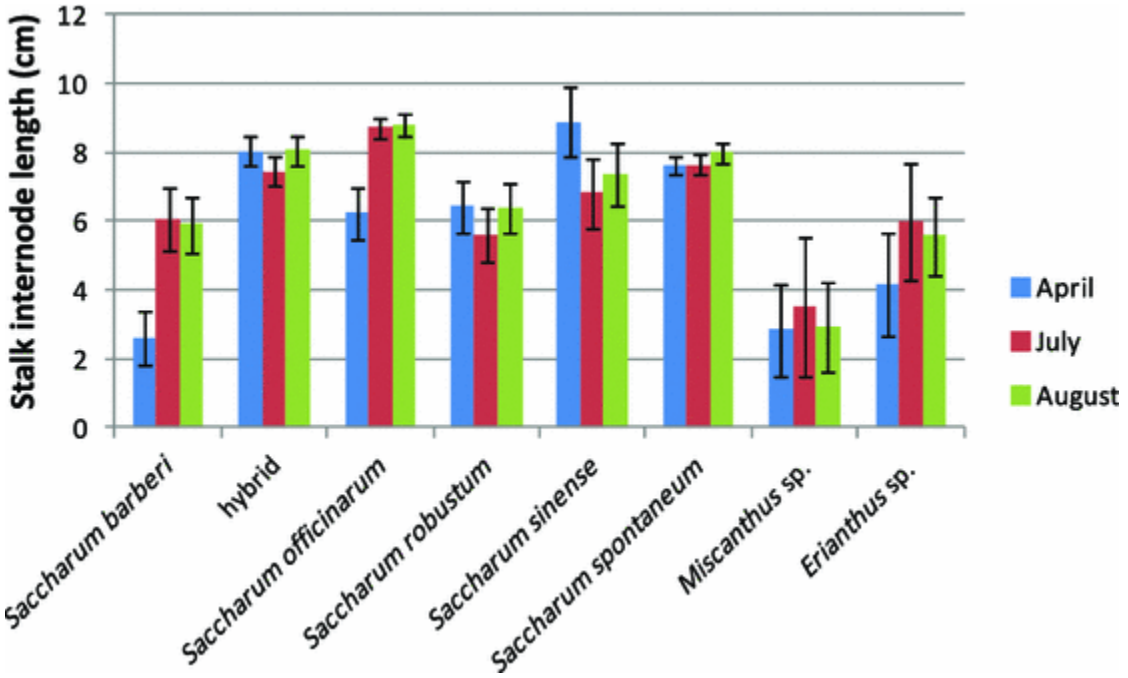

Fig. 7. Mean stalk internode length (cm) of species with accessions >10 within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL for all three measurement months with 95 % confidence interval *error bars*

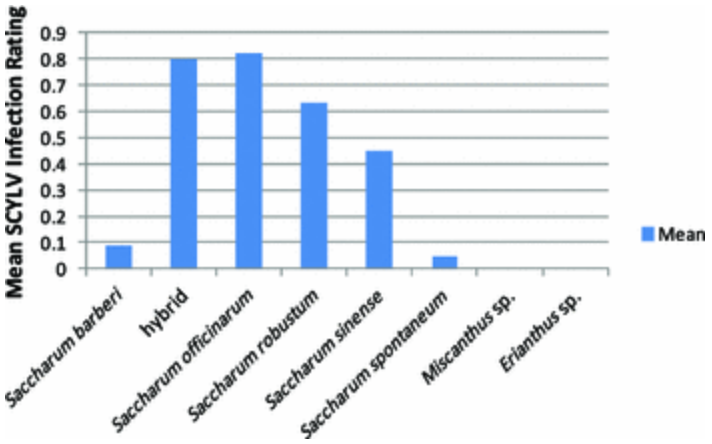

Fig. 8. Mean *SCYLV* infection of species with >10 accessions within the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses in Miami FL, and the higher the number, the more infection

Table 2. Comparison of descriptive statistics of the accessions growing in the field at the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses of sugarcane in Miami, Florida and the accessions from this collection that were selected for the core collection

Source	April				July				August			
	Mean	Median	Range	Variance	Mean	Median	Range	Variance	Mean	Median	Range	Variance
Stalk height (cm)												
WC	143.9	139.7	340.7	2,683	141.72	139.00	237.00	1,725.54	157.6	156.0	334.7	2,349
Core	154.2	146.5	322.7	3,813	147.10	147.17^{NS}	234.83	2,096.65^{NS}	165.1	161.7	300.3	2,264
Stalk diameter (mm)												
WC	23.34	23.54	45.57	41.39	23.67	23.53	33.06	30.97	22.65	22.80	34.76	33.51
Core	24.25	23.88^{NS}	41.90	53.54^{NS}	23.44^{NS}	23.53^{NS}	25.33	31.06^{NS}	22.40^{NS}	22.11^{NS}	31.36	34.12^{NS}
Stalk internode length (cm)												
WC	7.02	7.00	19.00	9.28	7.69	7.67	16.50	6.95	7.95	7.83	20.33	18.50
Core	7.61	7.86^{NS}	19.00	11.12^{NS}	7.77	7.58^{NS}	15.50	9.07†	8.27	8.00^{NS}	18.50	8.14^{NS}
Leaf sheath pubescence												
WC	1.30	1.00	9.00	2.77	2.35	2.00	10.00	4.84	2.54	2.00	10.00	6.72

Source	April				July				August			
	Mean	Median	Range	Variance	Mean	Median	Range	Variance	Mean	Median	Range	Variance
Core	1.53	1.00^{NS}	9.00	3.49^{NS}	2.77	2.00*	10.00	5.80^{NS}	2.76	2.00^{NS}	10.00	8.09^{NS}
Stalk aerenchyma												
WC	0.96	0.00	8.00	1.85	0.71	0.00	8.00	1.40	1.92	0.00	10.00	6.39
Core	1.18	0.00^{NS}	8.00	2.45^{NS}	0.73	0.00^{NS}	8.00	1.93^{NS}	1.85	0.00^{NS}	10.00	6.58^{NS}
Stalk pith												
WC	4.15	4.00	10.00	3.41	3.24	2.00	10.00	2.88	2.27	2.00	10.00	7.24
Core	4.34	4.00^{NS}	8.00	3.33^{NS}	3.46	4.00^{NS}	10.00	4.13†	2.67	2.00*	10.00	8.47^{NS}
SCYLV												
WC	0.73	1.00	1.00	0.20	–	–	–	–	–	–	–	–
Core	0.68	1.00^{NS}	1.00	0.22^{NS}	–	–	–	–	–	–	–	–

We also genotypically evaluated1002 WCRG accessions (Figure 9) to understand its genetic diversity and population structure and to form a core collection, which captures most of the genetic diversity in the WCSRG. We screened 36 microsatellite markers (Table 3) on 1002 genotypes and recorded 209 alleles. Genetic diversity of the WCSRG ranged from 0 to 0.5 with an average of 0.304 (Table 4). The population structure analysis and principal coordinate analysis revealed three clusters with all *S. spontaneum* in one cluster, *S. officinarum* and *S*. hybrids in the second cluster and mostly non-*Saccharum* spp. in the third cluster (Figure 10). A core collection of 300 accessions was identified which captured the maximum genetic diversity of the entire WCSRG which can be further exploited for sugarcane and energy cane breeding (Figure 10b). Sugarcane and energy cane breeders can effectively utilize this core collection for cultivar improvement. Further, the core collection can provide resources for forming an association panel to evaluate the traits of agronomic and commercial importance.

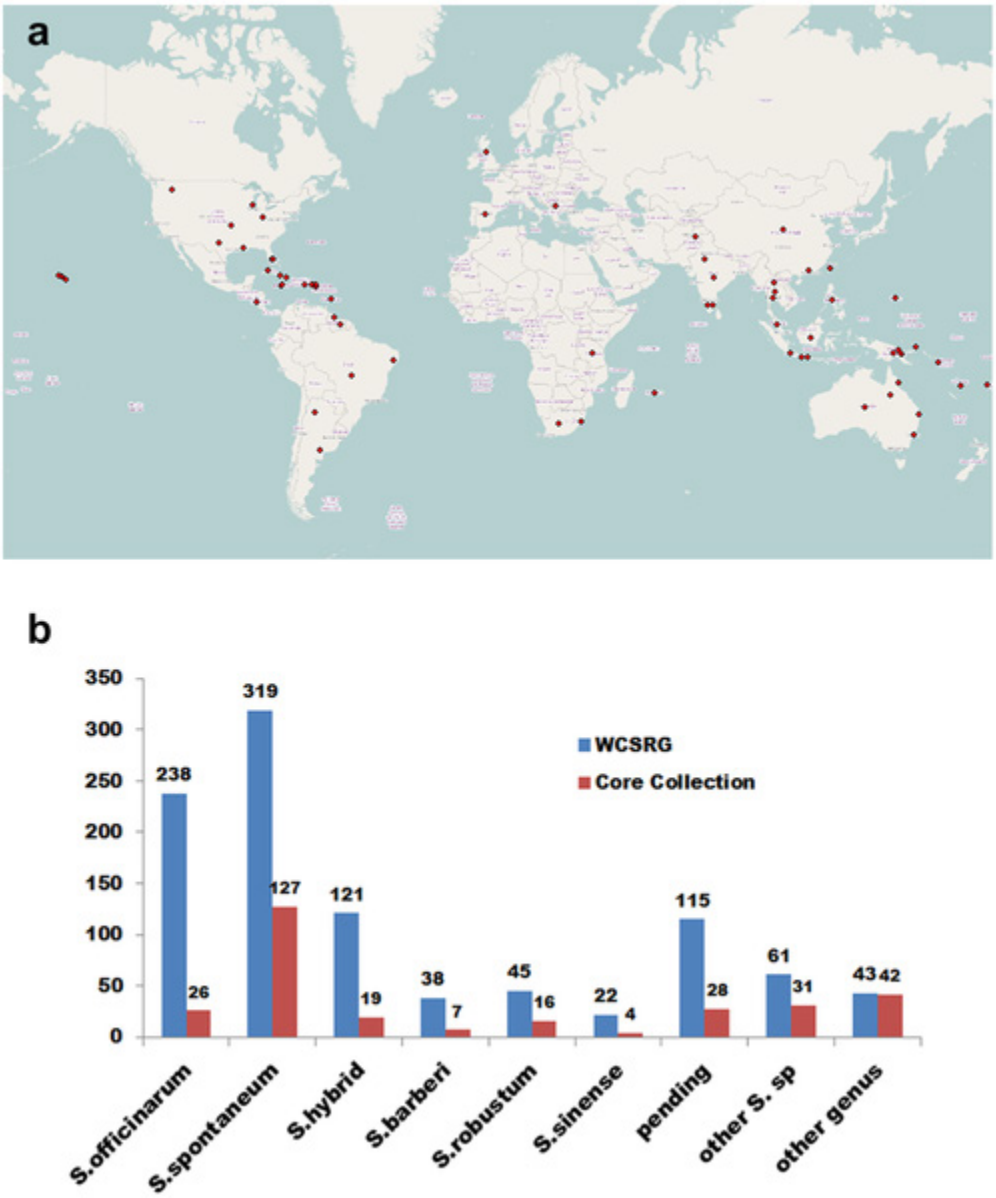

Figure 9. Distribution of the World Collection of Sugarcane and Related Grasses (WCSRG). (a) Geographic distribution of the accessions in the WCSRG. The 1002 accessions in the WCSRG were obtained from 45 countries. Each red dot represents a sugarcane collecting location. Global Mapper V14 software with OpenStreetMap was used to locate the accessions based on the latitude and longitude of origins. (b) Numerical distribution of the different species in the WCSRG and the core collection identified.

Table 3. Parameter list of 36 simple sequence repeat (SSR) primer pairs used for genotyping of 1002 accessions in the World Collection of Sugarcane and Related Grasses in Miami, FL.

Primer name	Product size (bp)	Alleles (No.)	Major allele frenquency	Range of PIC[a] values	Mean of PIC values	I[b]	Q[c]
SCAA8L15-12A	180–400	11	0.996	0.008–0.372	0.168	0.028	0.750
SC109D21-11-2	100–1000	8	0.918	0.102–0.375	0.317	0.038	0.706
SC159C20-21a	150–700	4	0.984	0.062–0.368	0.232	0.125	0.474
SMC24DUQ	140–150	3	0.975	0.048–0.327	0.166	0.188	0.367
SMC334BS	160–190	6	0.913	0.157–0.375	0.292	0.069	0.604
SMC569CS	180–300	4	0.959	0.076–0.273	0.192	0.291	0.281
SMC22DUQ	150–180	4	0.911	0.153–0.340	0.229	0.137	0.452
mSSCIR43	180–280	11	0.922	0.136–0.374	0.283	0.030	0.742
SMC31CUQ	120–500	2	0.567	0.371–0.372	0.372	0.375	0.187
SCA09	90–400	4	0.977	0.163–0.370	0.268	0.147	0.438
SCA10	200–250	15	0.997	0.006–0.374	0.240	0.016	0.811
SCB10	100–500	4	0.989	0.022–0.190	0.129	0.110	0.503
SCM21	160–450	3	0.940	0.136–0.334	0.256	0.189	0.366
UGSM60	200–900	5	0.909	0.123–0.343	0.220	0.073	0.592
UGSM585	160–1000	6	0.981	0.095–0.359	0.252	0.054	0.651
UGSM667	160–1000	17	0.984	0.029–0.375	0.272	0.012	0.840
SEGM2dot	180–100	2	0.747	0.307–0.369	0.337	0.395	0.178
SEGM285	290–800	8	0.873	0.197–0.374	0.315	0.042	0.692
UGSM459	380–600	10	0.989	0.022–0.374	0.263	0.036	0.716
UGSM594	200–700	3	0.928	0.143–0.360	0.246	0.339	0.229
UGSM629	300–1000	4	0.902	0.248–0.336	0.272	0.189	0.377
UGSM694	20–1000	3	0.883	0.217–0.292	0.244	0.187	0.369
UGSM399	120–600	3	0.871	0.199–0.373	0.314	0.205	0.348
UGSuM26	180–800	3	0.952	0.081–0.185	0.144	0.186	0.370
UGSuM349	400–1000	1	0.610	0.363	0.363	–	–
UGSuM43	250–1000	5	0.922	0.145–0.250	0.205	0.072	0.595
UGSuM15	220–420	3	0.921	0.134–0.353	0.278	0.240	0.314
UGSuM186	130–500	4	0.791	0.293–0.374	0.352	0.125	0.474
UGSuM197	120–500	9	0.998	0.002–0.374	0.171	0.064	0.620
UGSuM337	200–550	11	0.997	0.006–0.367	0.176	0.017	0.806
UGSuM97	120–380	6	0.690	0.360–0.375	0.357	0.055	0.647
UGSuM96	150–480	5	0.993	0.014–0.354	0.279	0.092	0.547
UGSuM56	120–800	4	0.968	0.059–0.348	0.241	0.189	0.381
UGSuM34	180–700	5	0.952	0.194–0.375	0.278	0.146	0.441
UGSuM21	500–900	7	0.937	0.014–0.354	0.279	0.065	0.618
UGSuM45	210–500	6	0.933	0.059–0.348	0.241	0.095	0.541

[a]PIC, polymorphism information content; [b]I, Probability of identity; [c]Q, Power of exclusion.
doi:10.1371/journal.pone.0110856.t001

Table 4. Gene diversity, Shannon's information index, and polymorphism status of six species of *Saccharum* and other categories.

Species	Gene diversity	I[a]	Polymorphic bands (%)
S. officinarum	0.2564	0.456	95.22
S. spontaneum	0.3032	0.492	99.52
S. hybrid	0.2531	0.452	93.30
S. barberi	0.2398	0.423	85.17
S. robustum	0.2670	0.427	85.65
S.sinense	0.2381	0.383	75.60
pending	0.2985	0.486	96.65
other *S. sp*	0.2756	0.462	91.39
other genus	0.3030	0.4470	91.39

a Shannon's Information Index

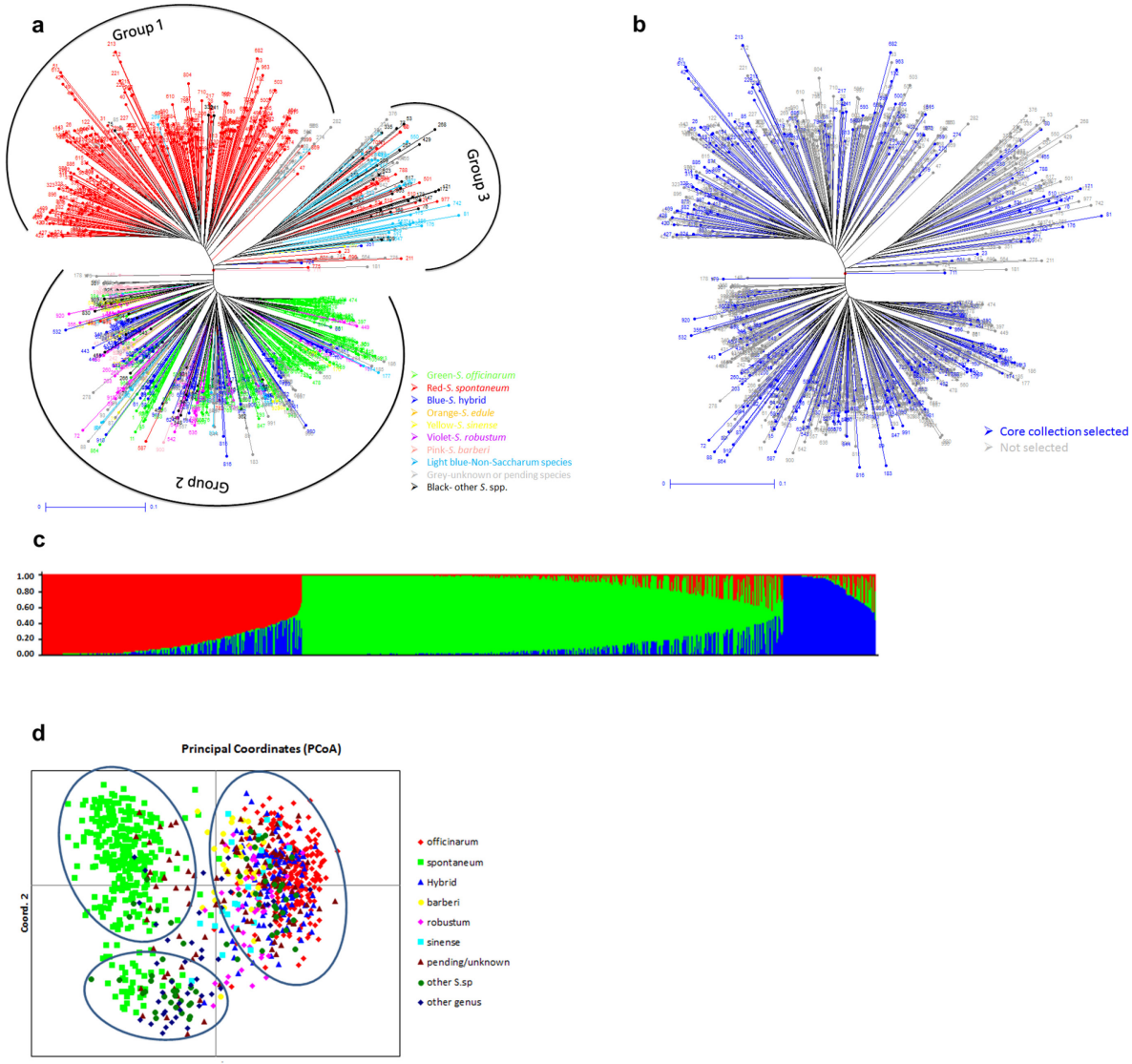

Figure 10. The clusters of the World Collection of Sugarcane and Related Grasses (WCSRG) and core collection in Miami, FL USA. (a) Phylogenetic tree of the WCSRG using neighbor-joining analysis. (b) Representativeness of the 300 accessions (colored blue) of the core collection selected from the WCSRG. Accessions not selected for the core collection are shaded grey. (c) The population structure of the WCSRG based on model-based estimation of 209 alleles. The WCSRG is grouped into three subgroups. Each individual is represented by a vertical line. Each color represents one subpopulation, and the length of the colored segment shows the proportion of membership for that accession. (d) Two-dimensional plot of the distribution of the WCSRG through principal coordinate analysis based on genetic distance generated from 209 alleles. The different colors represent nine pre-defined species.

2) Identify the alleles contributing the biomass traits

To identify the alleles contributing to biomass traits, we established a

representative diversity panel selected from the WCSRG of approximately 300 accessions by planting them at Canal Point, FL in three replications. This diversity panel was thoroughly phenotyped for morphological traits and fiber component traits. Firstly, they were measured for stalk height and stalk number multiple times throughout the growing season as well as for Brix and fresh biomass during the 2013 harvest (Figure 11 and Table 5). In the first-ratoon, stalk height, stalk number, stalk diameter, internode length, Brix and fresh and dry biomass were determined (Table 6). Hybrids and check varieties had higher fresh weight and Brix while *S. spontaneum* had higher stalk number and dry mass. There were 110 accessions not significantly different in Brix from the sugarcane commercial standards, including 10 *S. spontaneum* accessions. There were 17 and 6 accessions that were higher than the commercial standards in dry and fresh mass, respectively.

In addition, fiber component traits such as the acetyl groups, acid insoluble lignin, acid soluble lignin, arabinan, glucan, holocellulose, total lignin, structural ash, xylan, and nonstructural ash were quantified. Significant differences for holocellulose, lignin, acetyl, acid soluble lignin, nonstructural ash, and glucan were observed between species or hybrid groups of accessions. *S. spontaneum* had significantly more holocellulose, glucan, lignin, and nonstructural ash, and less acetyl and acid soluble lignin than the other groups (Figures 12, 13, and 14). Holocellulose had the lowest variability and ash traits had the highest variability. Ash negatively correlated with lignin and acetyl and positively with holocellulose. The fiber data evaluated in this study will allow breeders to make informed decisions about potential effects of parental selection on biomass traits of progeny.

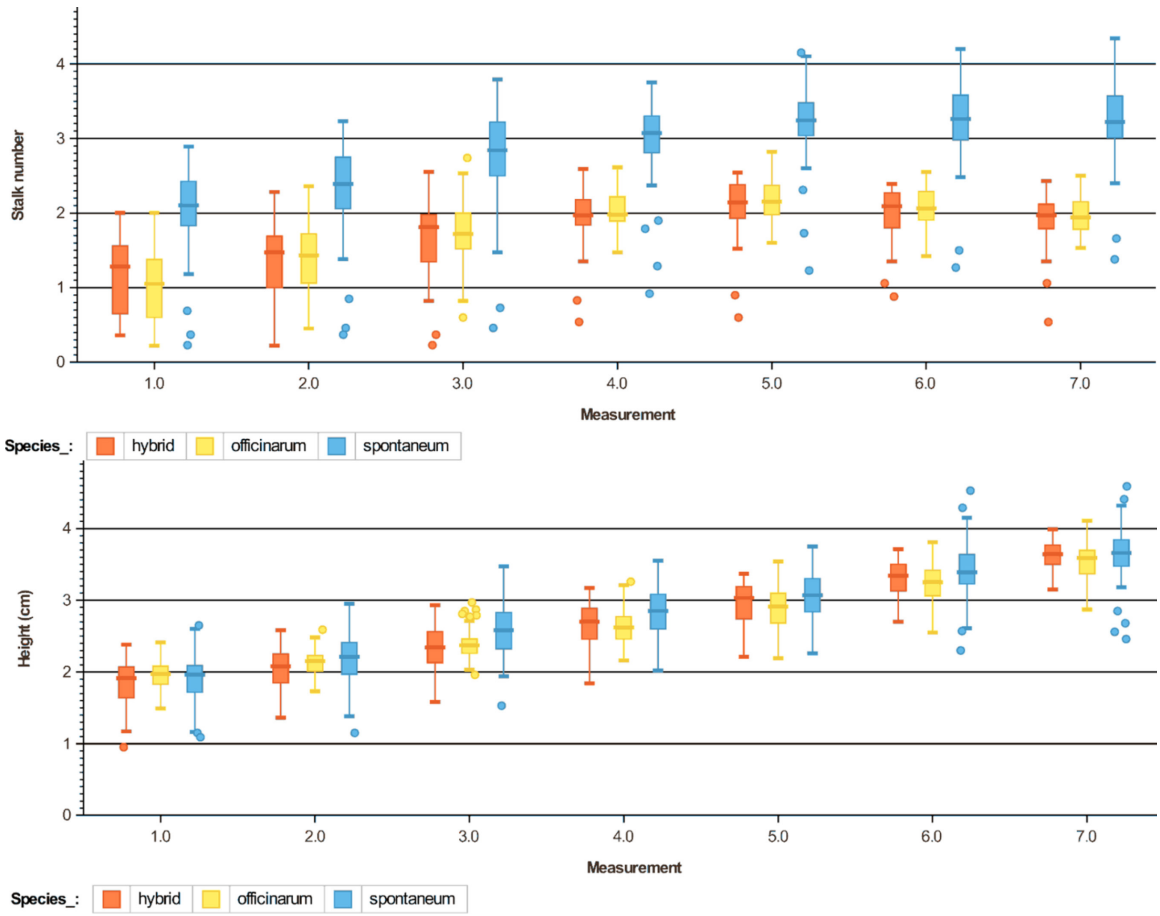

Figure 11. (a) Stalk number of the three major species of *Saccharum* measured at seven different times during the plant cane growing season. (b) Box plots of the height of the three major species of *Saccharum* diversity panel measured at seven different times during the plant cane growing season.

Table 5. The estimated mean, standard deviation, least significant difference (0.05) and coefficient of variation for plant cane phenotypic traits in the diversity panel.

		Fresh weight 2013				SH 2013				2013 Brix				2013 Stalk Number			
	2013	Mean	SE	diff	CV	Mean	SE	diff	CV	Mean	SE	diff	CV	Mean	SE	diff	CV
Over all	308	4.5	0.41		30.81	122.94	4.08		20.1	10.55	2.53		15.89	13.08	0.42,0.41		12.19
officinaru	78	4.75	0.42	BC	34.06	116.32	4.53	C	17.64	12.04	2.54	B	14.43	5.8	0.28,0.27	E	15.75
spontaneu	107	3.96	0.43	DE	26.1	128.42	4.42	AB	21.83	8.53	2.54	D	18.27	31.34	1.45,1.38	A	9.25
Hybrid	38	5.15	0.44	B	26.03	122.44	5.1	BC	14.71	12.42	2.54	B	14.31	6.24	0.39,0.37	E	16.15
er *Sacchai*	33	4.53	0.45	C	35.45	115.75	5.42	C	16.97	10.81	2.55	C	13.83	8.52	0.55,0.52	D	13.83
ther Gene	18	3.63	0.5	E	35.2	117.4	6.66	BC	30.78	8.8	2.56	D	18.9	24.7	2.04,1.89	B	8.22
Checks	10	7.15	0.54	A	19.77	131.77	7.51	AB	10.09	15.65	2.56	A	12.45	7.87	0.84,0.76	D	11.54
Unknown	24	4.47	0.5	BCD	26.05	133.95	5.95	A	24.46	10.53	2.55	C	19.79	13.23	1.11,1.03	C	13.01

Table 6. The estimated mean, standard deviation, least significant difference (0.05) and coefficient of variation for ratoon phenotypic traits in the diversity panel

	#	2014 Brix				Fresh weight 2014				Dry Weight 2014				2014 Moisture content			
		Mean	SE	diff	CV	Mean	SE	diff	CV	Mean	SE	diff	CV	Mean	SE	diff	CV
Over all	289	12.32	0.34		19	3.14	0.25		57.35	1.04	0.13		72.47	2.12	0.15		58.77
. *officinarur*	70	13.4	0.44	C	16.08	2.6	0.26	C	53.19	0.63	0.13	D	55.42	1.97	0.15	C	53.21
spontaneu	103	10.73	0.41	E	22.05	3.13	0.24	B	53.92	1.27	0.13	A	59.89	1.88	0.14	C	56.33
Hybrid	37	14.39	0.47	B	15.55	3.78	0.29	A	46.1	1.01	0.14	C	52.08	2.79	0.18	A	45.47
her *Sacchar*	28	12.28	0.49	D	15.17	3.62	0.31	AB	57.38	1.06	0.15	BC	53.58	2.65	0.19	AB	54.53
Other Gener	18	10.99	0.55	E	21.34	3.03	0.36	BC	82.33	1.31	0.17	AB	81.35	1.84	0.24	C	73.93
Checks	10	16.2	0.65	A	12.12	4.19	0.43	A	44.08	1	0.2	ABC	43.74	3.19	0.29	A	47.88
Unknown	23	12.9	0.52	CD	19.41	3.42	0.33	AB	61.49	1.35	0.16	A	108.77	2.18	0.21	BC	84.32
		2014 Stalk Diameter				2014 Stalk Number				2014 Stalk Height				2014 Stalk internode			
Over all	289	16.33	0.58		25.95	16.77	0.92,0.87		62.24	127.6	4.67		23.32	10.09	0.6		33.98
. *officinarur*	70	23.31	0.6	B	24.07	6.91	0.50,0.47	E	11.89	121.09	6.16	C	23.12	8.1	0.74	C	39.71
spontaneu	103	8.34	0.55	D	31.23	41.75	2.12,2.02	A	11.54	128.93	5.94	B	24.46	11.9	0.72	A	29.62
Hybrid	37	24.77	0.72	A	19.65	7.32	0.62,0.57	E	11.99	137.85	6.75	AB	19.59	9.78	0.79	B	28.65
her *Sacchar*	28	18.09	0.79	C	22.07	10.97	0.98,0.90	D	16.88	128.77	7.14	BC	20.77	8.41	0.82	C	35.36
Other Gener	18	10.06	0.95	D	26.88	28.08	2.85,2.60	B	16.2	100.33	7.97	D	30.79	10.59	0.94	AB	40.16
Checks	10	25.14	1.18	AB	20	8.71	1.30,1.15	DE	12.41	128.96	9.44	ABC	18.48	9.27	1.03	BC	40.69
Unknown	23	16.33	0.86	C	30.38	18.7	1.78,1.64	C	17.78	143.04	7.51	A	22.13	10.63	0.86	B	30.44
		2014 Leaf Width				2014 Leaf Length											
Over all	289	27.06	1.42		42.76	119.24	1.65		18.17								
. *officinarur*	70	40.4	1.55	A	33.42	125.16	2.18	B	15.77								
spontaneu	103	12.37	1.45	E	86.91	108.31	1.92	C	20.28								
Hybrid	37	36.96	1.8	B	28.22	133.09	2.82	A	12.08								
her *Sacchar*	28	35.2	1.96	B	27.79	122.27	3.18	B	21.05								
Other Gener	18	17.73	2.3	D	37.5	110.09	3.92	C	22.93								
Checks	10	34.37	2.85	B	33.57	139.21	5.1	A	12.73								
Unknown	23	27.45	2.12	C	57.08	120.58	3.52	B	21.94								

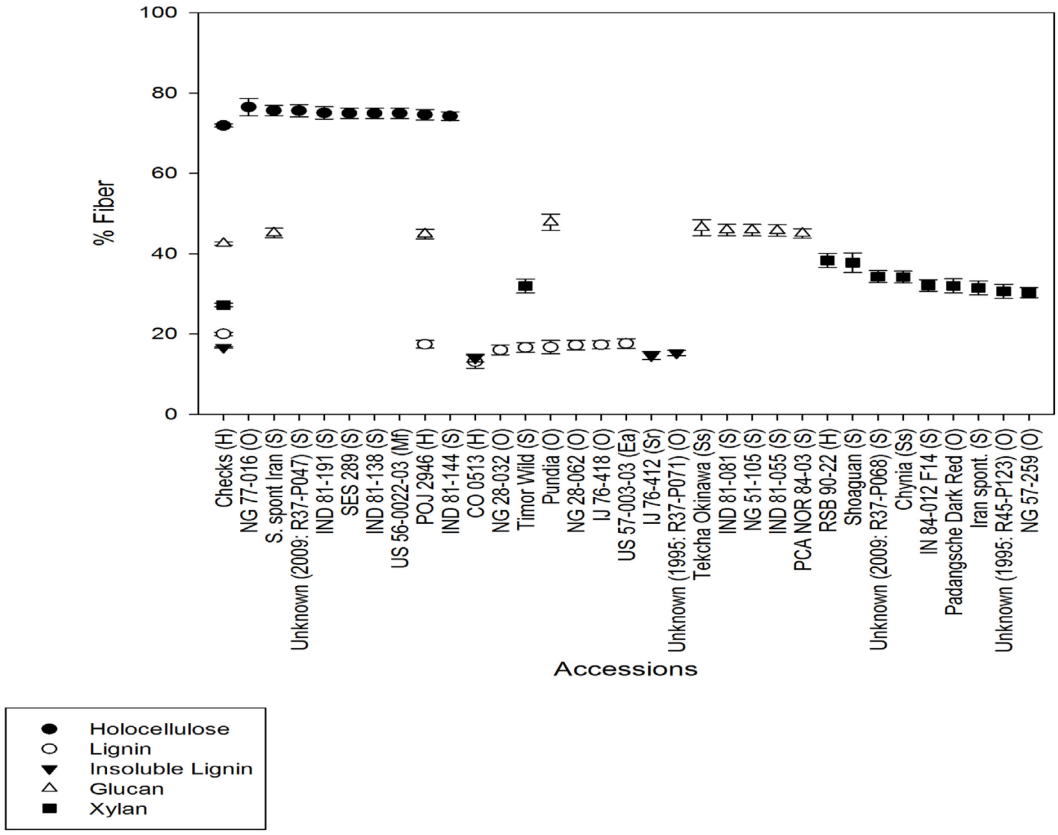

Figure 12. Scatterplot of the holocellulose, total lignin (lignin), acid insoluble lignin (Insoluble Lignin), Glucan and Xylan fiber components of *Saccharum* and related taxa accessions significantly different ($P < 0.05$) than the mean of the checks (CP 00-1101, CP

01-1372, CP 03-1912, CP 72-2086, CP 78-1628, CP 88-1762, CP 89-2143, CP 96-1252, CPCL 00-4111, and CP 01-2390). Species are indicated by the following abbreviations: H = hybrids, Ea *Erianthus arundinaceum*, Mf = *Miscanthus floridulus*, O = *Saccharum officinarum*, S = *Saccharum spontaneu*m, Sr = *Saccharum robustum*, Ss = *Saccharum sinense*.

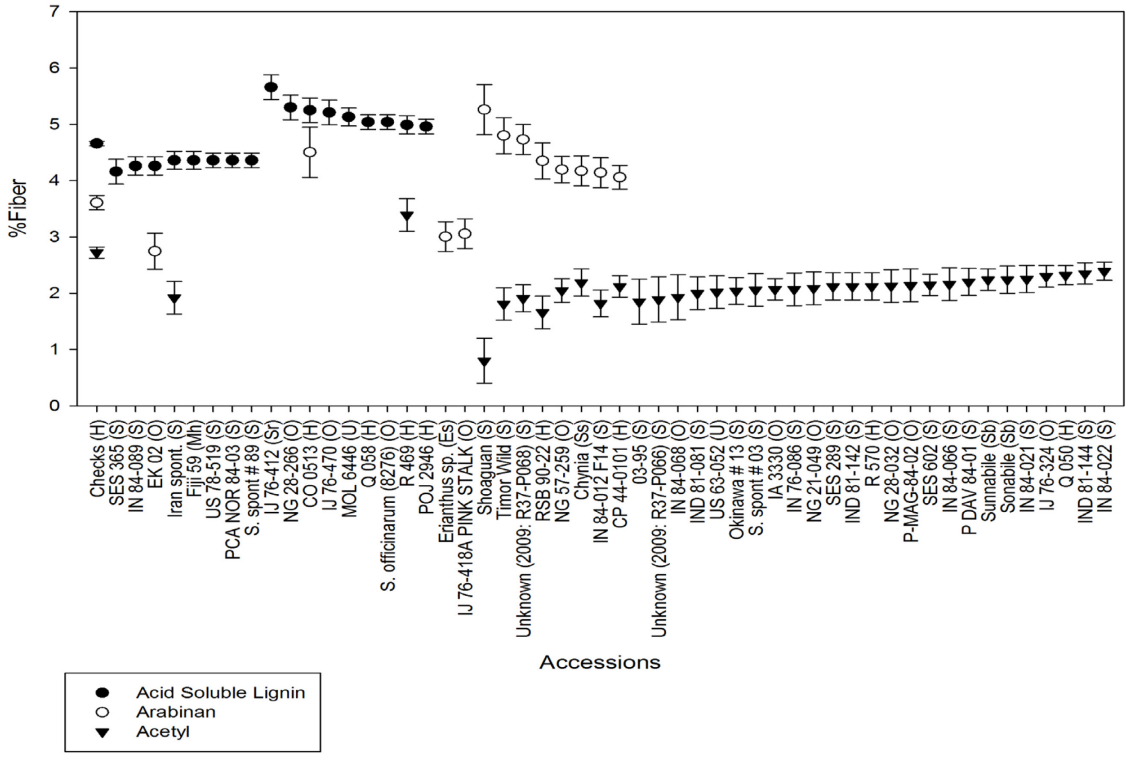

Figure 13. Scatterplot of the acid soluble lignin, arabinan, and acetyl fiber components of *Saccharum* and related taxa accessions significantly different ($P < 0.05$) than the mean of the checks (CP 00-1101, CP 01-1372, CP 03-1912, CP 72-2086, CP 78-1628, CP 88-1762, CP 89-2143, CP 96-1252, CPCL 00-4111, and CP 01-2390). Species are indicated by the following abbreviations: H = hybrids, Ea *Erianthus arundinaceum*, Mh = *Miscanthus hybrid*, O = *Saccharum officinarum*, S = *Saccharum spontaneu*m, Sr = *Saccharum robustum*, Ss = *Saccharum sinense*, Sb = *Saccharum barberi*, U = Unknown.

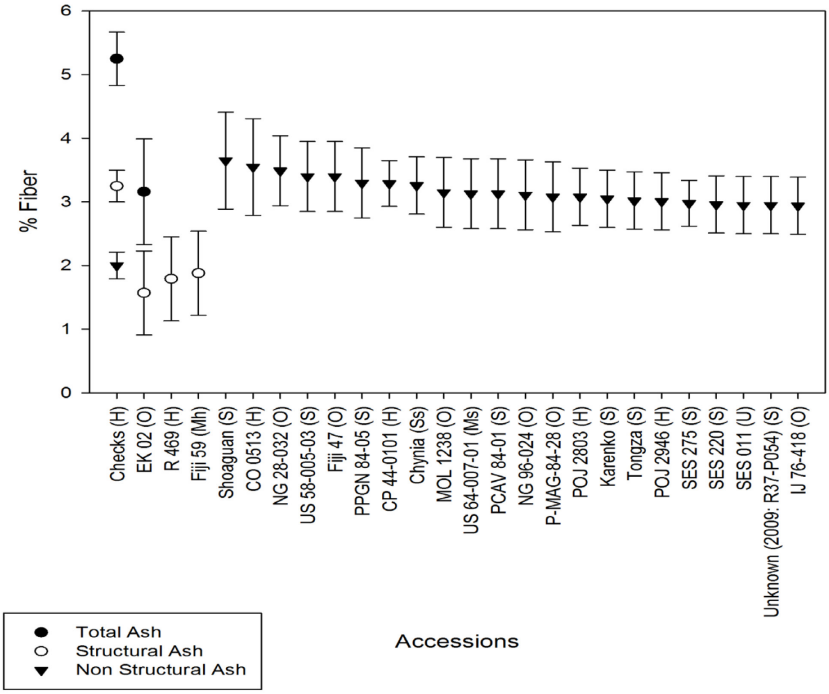

Figure 14. Scatterplot of the ash fiber components of *Saccharum* and related taxa accessions significantly different ($P < 0.05$) than the mean of the checks (CP 00-1101, CP 01-1372, CP 03-1912, CP 72-2086, CP 78-1628, CP 88-1762, CP 89-2143, CP 96-1252, CPCL 00-4111, and CP 01-2390). Species are indicated by the following abbreviations: H = hybrids, Mh = *Miscanthus hybrid*, Mh = *Miscanthus sinensis*, O = *Saccharum officinarum*, S = *Saccharum spontaneu*m, Ss = *Saccharum sinense*, U = Unknown.

This diversity panel was further genotyped using target enrichment sequencing. Firstly a target enrichment sequencing approach was used to assay 12 representative germplasm accessions as a pilot experiment. In total, 55,946 highly efficient probes were designed based on the sorghum genome and sugarcane unigene set targeting a total of 6 Mb of the sugarcane genome (Figure 15). A pipeline specifically tailored for polyploid sequence variants and genotype calling was established. BWA-mem and sorghum genome approved to be an acceptable aligner and reference for sugarcane target enrichment sequence analysis, respectively (Table 7). Genetic variations including 1,166,066 non-redundant SNPs, 150,421 InDels, 919 gene copy number variations, and 1,257 gene presence/absence variations were detected (Figure 15). SNPs from three different callers (Samtools, Freebayes, and GATK) were compared (Figure 16) and the validation rates were nearly 90%. Based on the SNP loci of each accession and their ploidy levels, 999,258 single dosage SNPs were identified and most loci were estimated as largely homozygotes (Figure 17). An average of 34,397 haplotype blocks for each accession was inferred. The target enrichment sequencing approach provided an effective way to discover and catalog natural allelic variation in highly polyploid or heterozygous genomes.

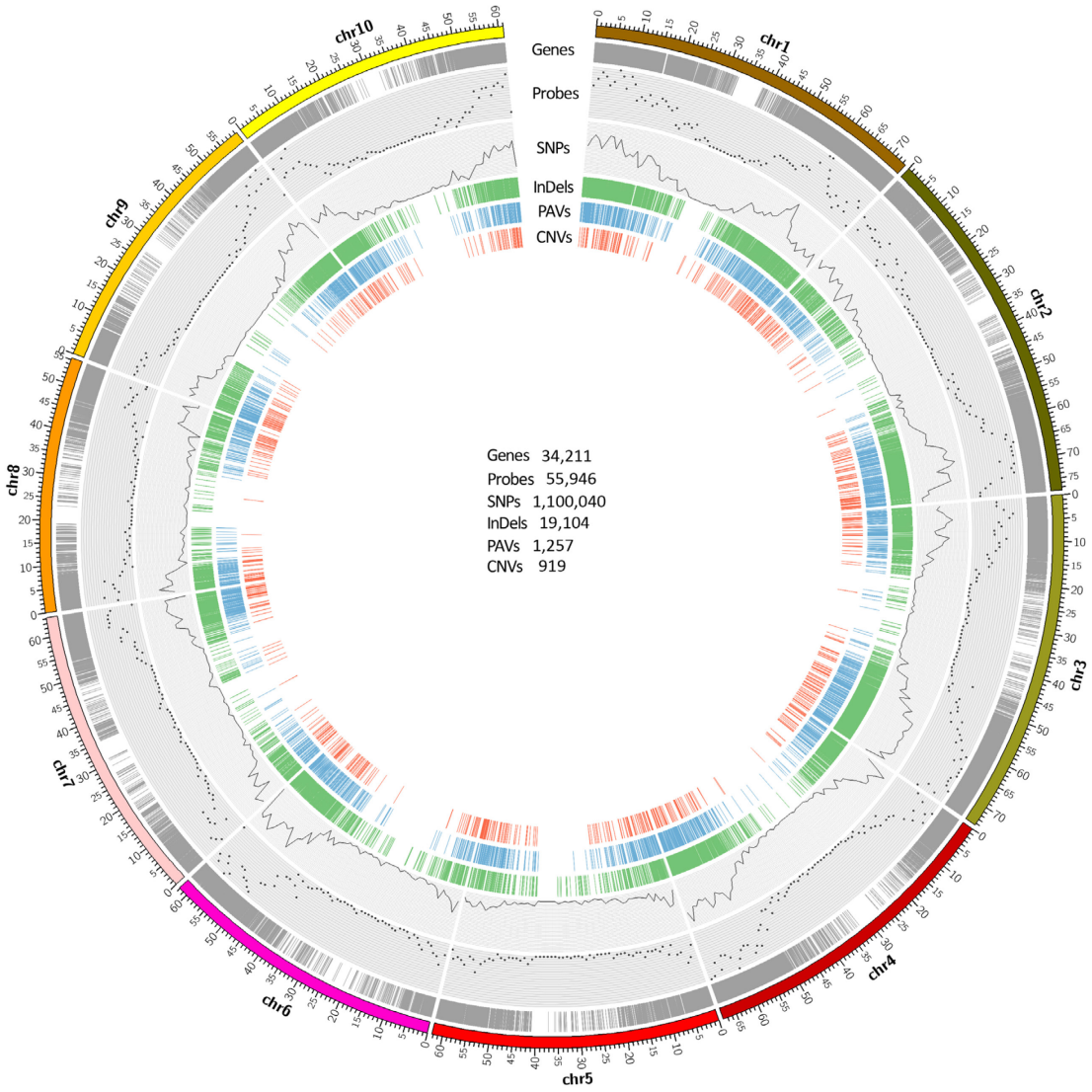

Figure 15. Genome-wide distribution of probes and sequence variants along sorghum genome. Gene models were displayed by grey ring showing the majority of gene regions on the chromosome arms. Scatter plot represents the number of probes within 1Mb window size across sorghum genome. The SNPs for 12 accessions were plotted by the black line which represented the SNP density within 1Mb window size across sorghum genome. The green color represented concordant InDels which were called by Samtools, GATK and Freebayes. The blue and red color represented copy number variations and presence and absence of variations identified in this study, respectively. The number of gene models and each variant category were showed in the circle.

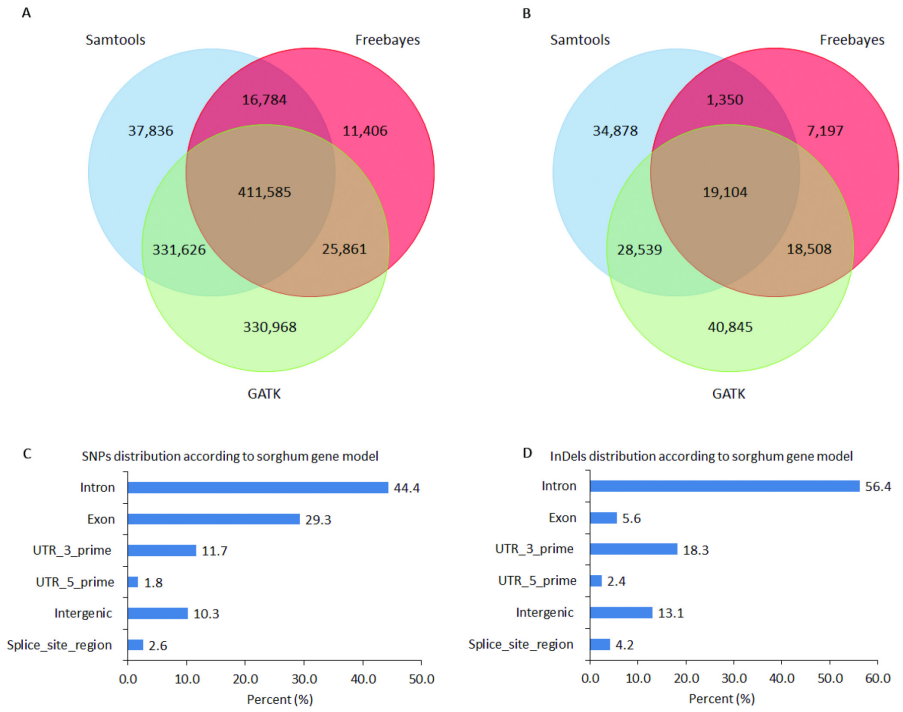

Figure 16. Number of single nucleotide polymorphisms (SNPs) and InDels. (A) Venn diagram showing overlapping SNP variants among Samtools, GATK and Freebayes callers; (B) Venn diagram showing overlapping InDel variant sites among Samtools, GATK and Freebayes callers; (C) SNPs distribution according to sorghum gene model by using SNP calls from GATK; (D) InDels distribution according to sorghum gene model by using common InDels generated by three callers

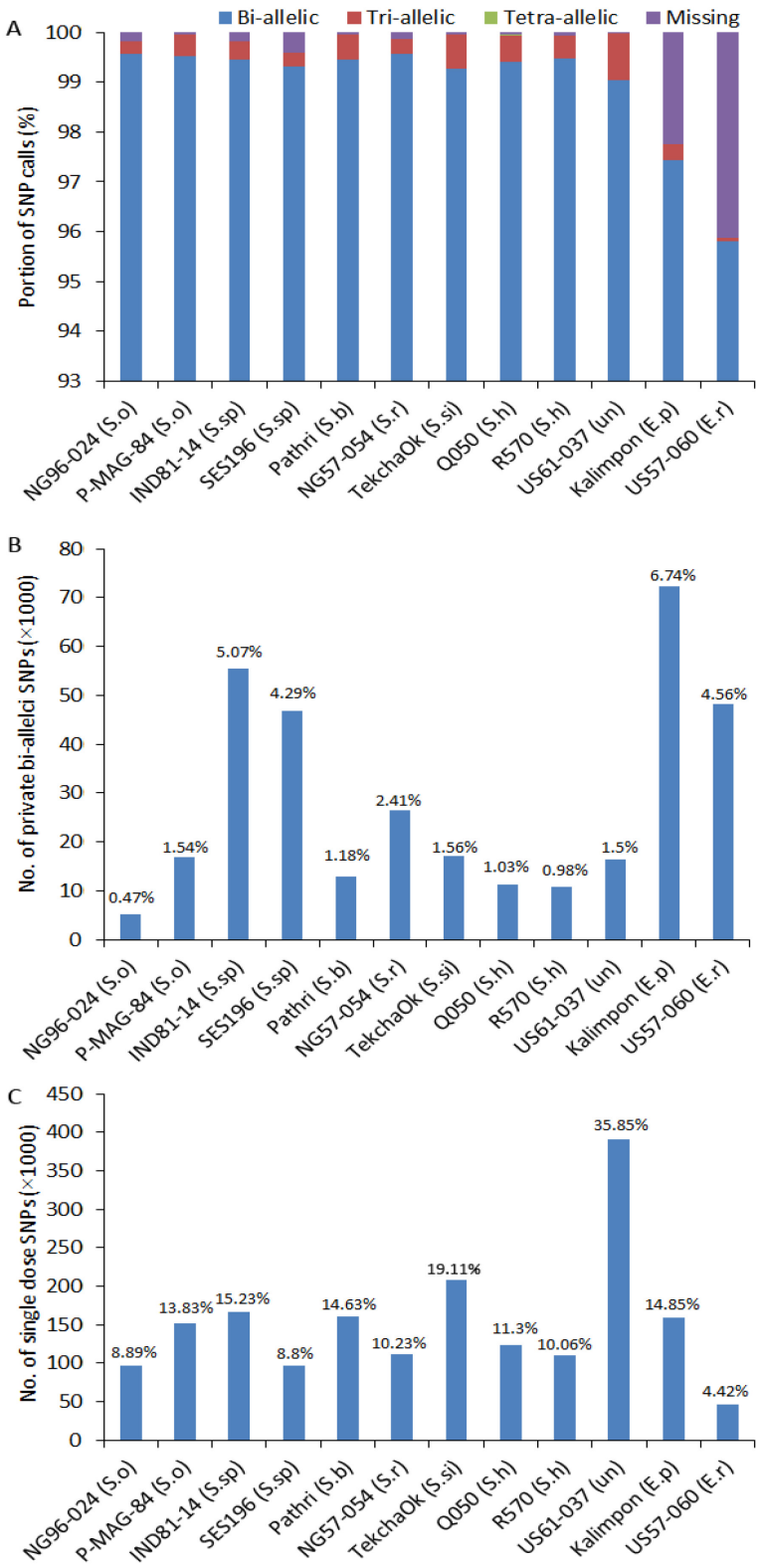

Figure 17. Distribution of SNP types and genotypes. (A) The percentage of various SNP calls including bi-allelic, tri-allelic, tetra-allelic SNPs and SNP loci with missing data for 12

accessions; (B) The percentage of the genotype-specific SNPs called for each accession; (C) The number of single-dose markers called for each accession

Table 7. Statistics of target sequencing reads alignment to four different sequence sources as reference.

			Sorghum genome		Probes		EST unigene set		Assembly (> 200 bp)	
	No. of raw reads (millions)	No. of clean reads (millions)	No. of reads mapped (millions)	Percentage of reads mapped	No. of reads mapped (millions)	Percentage of reads mapped (%)	No. of reads mapped (millions)	Percentage of reads mapped (%)	No. of reads mapped (millions)	Percentage of reads mapped (%)
NG96-024 (S. o)	28.4	25.6	21.9	85.6	13.8	53.9	18.1	70.7	18	70.5
P-MAG-84 (S. o)	46.5	40.2	34.3	85.5	21.2	52.9	28.3	70.5	28.2	70.4
IND81-14 (S. sp)	52.2	42	35.2	83.9	18.2	43.4	28.9	68.9	29.1	69.4
SES196 (S. sp)	35.4	28.4	24	84.3	12.6	44.2	19.9	69.8	19.8	69.6
Pathri (S. b)	40.7	36.5	31.5	86.3	20.3	55.6	25.8	70.6	25.7	70.4
NG57-054 (S. r)	35.5	31.4	27.4	87.1	18.7	59.5	22.5	71.5	22.2	70.6
TekchaOk (S. si)	44.9	35.9	30.3	84.4	17.5	48.8	25	69.7	25.1	69.8
Q050 (S. h)	56.7	47.5	40.4	85.1	22.5	47.5	33.3	70.3	33.3	70.2
R570 (S. h)	45.8	35.5	30.5	86	16.1	45.5	24.8	69.7	25.1	70.7
US61-037 (un)	55.2	48.5	42.7	88	26.7	55.1	34.2	70.6	34.3	70.7
Kalimpon (E. p)	22.8	19.7	16.1	81.9	9.4	47.7	12.4	62.8	13.6	68.9
US57-060 (E. r)	24	20.5	17.1	83.5	9.8	47.9	13.1	64	14.1	68.7
Total	488.1	411.5	351.3		206.9		286.2		288.5	
Average	40.7	34.3	29.3	85.1	17.2	50.2	23.8	69.1	24	70

S. o: *Saccharum officinarum*; S. sp: *Saccharum spontaneum*; S. b: *Saccharum barberi*; S. r: *Saccharum robustum*; S. si: *Saccharum sinense*; S. h: *Saccharum* hybrid; un: unknown; E. p: *Erianthus procerum*; E. r: *Erianthus rufipilus*

Target enrichment sequencing of 300 representative accessions, a condensed assembly of the WCSRG were further conducted. Genomics regions spanning a total of 6 Mb of 307 accessions, including 300 accessions from WCSRG and seven modern cultivars were deep sequenced. In total, 3,4 million SNPs across 307 clones were identified, which included 2.5 million bi-allelic and 0.9 million multi-allelic SNP loci. Out of 3.4 million SNPs, 1,192,505 (35.3 %) and 1,365,196 (40.4 %) were located in exons and introns of the gene models, respectively, according to sorghum genome. The SNPs located on the exon regions included 697,275 nonsynonymous and 457,712 synonymous SNPs. The rest SNPs, 358,213 (13%) and 465,803 (10%) located on intergenic regions and UTRs. A total of 32,467 SNPs with significant effects on gene functions were predicted. The population structure demonstrated four groups of this diversity panel. The structure and kinship information were applied in the mixed model

for association analysis using GWASPoly. In total, for the 24 biomass related traits, we identified 248 associations between SNPs and traits we measured. The further model refining and validation should be conducted for final conclusion. The findings provided deep insight of the sugarcane complex genome variations and substantial genetic information for sugarcane and energy cane improvement.

3) Breeding high biomass yielding energy cane cultivars through energy cane breeding program in Florida.

Eight advanced breeding lines were planted in relicated trials at three locations in Florida with four replications at each location. All three locations had mineral soil with low organic matter, also called marginal soils with low inherent soil fertility. Trial designs were randomized complete blocks (RCBD) with four replications at each site. Plots were 7.6 m long and 3-rows (4.5 m) wide and were arranged in tiers. Two plots were side by side (total 6 rows) in each tier with 4.5 m alleys on each side of the tier. High dry biomass yield (i.e. megagrams per hectare; Mg/ha) and resistance to natural disease infections in the field were the main genotype selection criteria. Smut was the most predominant disease in these energy cane trials. Smut stools and whips were counted in each plot, and the numbers of whips per $10m^2$ area were calculated to normalize the data. Plots were also rated for brown rust, orange rust, leaf scald and SCMV, but the infection was either absent or quite low.

Fresh biomass yields were estimated by cutting and weighing the center row of each plot. A 10-stalk subsample was collected from each plot to estimate dry biomass and fiber composition. The subsample was shredded using a modular sugarcane disintegrator (Codistil S/A Denini, Mod: 132S, Piracicaba-SP, Brazil)) and mixed thoroughly in a plastic container. To estimate dry biomass, a sample of approximately 1 kg shredded material was weighed from fresh biomass and then dried at 60^oC until constant weight. The dry samples were weighed and then ground to pass a 2-mm screen in a Wiley-Mill, standard model #3 (Thomas Scientific, Philadelphia, PA). A 40 g subsample of this finely ground material was collected into 20 ml vials (HDPE scintillation vials, Fisher Scientific, Atlanta, GA) for fiber compositional analysis.
Total dry biomass yields were calculated using the formula:
Dry biomass yield = Fresh biomass yield * Proportion of dry biomass yield: fresh biomass yield

Natural disease infection under field conditions was determined for brown rust and orange rust based on size and number of uredia. Field ratings for both rusts consisted of five classes: 0 (resistant), 1 (moderately resistant), 2 (moderately susceptible), 3 (susceptible), and 4 (highly susceptible). For natural infection of smut, smut whips were counted in each plot at each location. Smut whips per $10m^2$ were calculated to normalize the data and to make comparisons with the L 79-1002 check. Leaf scald and sugarcane mosaic were never observed in the field. Incidence of all field diseases was compared with the reference cultivar, L 79-1002. In addition to natural infection ratings, artificial disease inoculation tests were performed for smut, leaf scald

and mosaic in a greenhouse at USDA, Canal Point using standard methods developed for the CP sugarcane cultivar development program

Plant compositional analysis was performed via NIR (near-infrared reflectance spectroscopy) on dried and finely ground biomass subsamples. Analysis was performed at a contract laboratory (Hauser Division, Microbac Laboratories, Inc., Boulder, CO). Samples packed in 4 ml glass vials were scanned using a vial autosampler on a Fourier transform NIR analyzer (Antaris II, Thermo Scientific, Madison, WI, USA) fitted with an integrating sphere module. NIR chemometric software (OPUS, Version 7.0, Bruker Optik GmBH, Ettlingen, Germany) was used for spectral processing and analysis utilizing an independently developed calibration model. Compositional results are reported on a % of total H_2O extractives free basis (% fiber).

In addition, five new energy cane cultivars (UFCP74-1010, UFCP78-1013, UFCP82-1655, UFCP84-1047, and UFCP87-0053) were released for commercial cultivation The released cultivars have significantly lower smut infection (caused by *Sporisorium scitaminea*) and comparatively greater biomass yields than commercial check (L79-1002). Our data collected through 12 location-years (5 years at Citra, 3 years at Lykes Bros., and 2 years each at Tecan and 960 farms) indicate that L79-1002 (overall 26.28 whips/10m^2) had significantly greater smut infection than the released cultivars (overall 0.60 – 1.65 whips/10m^2), and there was no significant difference among the released cultivars (Table 1). The artificial lab inoculation data (Table 2) also indicate lower smut susceptibility (11-17.6% infection) in the released cultivars than L 79-1002 (37.8% infection) as well as smut susceptible sugarcane cultivar, CP 78-1628 (41.6% infection). CP 78-1628 is most widely grown sugarcane cultivar on mineral soils in South Florida. The released cultivars also have lower infection rates of leaf scald (*Xanthomonas albilineans* Ashby, Dowson) and mosaic (*Sugarcane mosaic virus* strain E) diseases compared to the susceptible commercial checks for these diseases.

The released cultivars have high biomass yields, which is a very important selection criteria for cellulosic ethanol production. In mineral soils, the cumulative (plant cane through ratoons) dry biomass yields of the released cultivars ranked higher than L 79-1002. Fiber analysis indicates that the released cultivars and a commercial check (L79-1002) had similar concentrations for cellulose, hemicellulose, lignin and ash composition.

6. Identify products developed under the award and technology transfer activities

Refereed Publications:

Sandhu, H. S., R. A. Gilbert, J. C. Comstock, V. Gordon, P. Korndorfer, N. El-Hout, and R. Arundale. 2016. Registration of 'UFCP 82-1655' sugarcane. *J. Plant. Reg.* 10: 22-27

Gordon, V., J. C. Comstock, H. S. Sandhu, R. A. Gilbert, P. Korndorfer, R. Arundale, and N. El-Hout. 2016. Registration of 'UFCP 87-0053' Sugarcane. *J. Plant Reg.* 10: 258-264.

Gordon, V., J. C. Comstock, H. S. Sandhu, R. A. Gilbert, P. Korndorfer, R. Arundale, and N. El-Hout. 2016. Registration of 'UFCP 84-1047' Sugarcane. *J. Plant Reg. 10: 251-257.*

Sandhu, H. S., J. C. Comstock, R. A. Gilbert, V. Gordon, P. Korndorfer, R. Arundale, and N. El-Hout. 2015. Registration of 'UFCP 78-1013' sugarcane. *J. Plant. Reg.* 9: 318-324.

Sandhu, H. S., R. A. Gilbert, J. C. Comstock, V. Gordon, P. Korndorfer, N. El-Hout, and R. Arundale. 2015. Registration of 'UFCP 74-1010' sugarcane. *J. Plant. Reg.* 9: 179-184.

Song J. (p) X. Yang (g), M. F. R. Resende Jr., L. G. Neves, J. Todd, J. Zhang, J. C. Comstock, J. Wang. Natural allelic variations in highly polyploid *Saccharum* complex. Frontiers in Plant Sciences. 2016. 7:804. doi: 10.3389/fpls.2016.00804.

Nayak S. N* (p)., J. Song* (p), A. Villa* (u), B. Pathak (&), T. Ayala-Silva, X. Yang (g) , J. Todd, N. C. Glynn , D.N. Kuhn, B. Glaz, R. Gilbert, J. C. Comstock, J. Wang. Promoting utilization of *Saccharum* spp. genetic resources though genetic diversity analysis and core collection construction. Plos One. 2014. 9 (10): e 10856. *: Equal contributors.

Todd J., J. Wang, B. Glaz, S. Sood, T. Ayala-Silva, S. Nayak (p), N. Glynn, O.A. Gutierrez, D. Kuhn, M. Tahir and J. Comstock. Phenotypic characterization of the Miami World Collection of sugarcane (*Saccharum* spp.) and related grasses for selecting a representative core. Genetic Resources and Crop Evolution. 2014. 61(8): 1581-1596.

Wang J., S. Nayak (p), K. Koch, and R. Ming. Carbon partitioning in sugarcane (*Saccharum* species). Frontiers in Plant Science. 2013. 4:201. doi:10.3389/fpls.2013.00201.

Conference papers/posters

Sandhu, H.S., J.C. Comstock, and R.A. Gilbert. 2015. Biomass yield and disease resistance in newly developed energy cane cultivars in Florida. World Congress of Bioenergy, 22-25 September, Xi'an, China.

Sandhu, H.S., V.S. Gordon, R.A. Gilbert, J.C. Comstock, and P. Korndorfer. 2015. Evaluation of energy cane clones as lignocellulosic ethanol feedstock in Florida. International Society of Sugar Cane Technologists (ISSCT) meeting, 24-28 August, Durban, SA.

Sandhu, H.S., R.A. Gilbert, J.C. Comstock, and V.S. Gordon. 2015. Effects of soil properties and environmental factors on biomass yield and disease incidence in energy cane. Annual meeting of ASA, CSSA and SSSA, 15-18 November, Minneapolis, MN.

Gordon, V.S., J. C. Comstock, H. S. Sandhu, R. A. Gilbert, N. El-Hout, and R. Arundale. 2015. Development of energy cane cultivars in Florida. 2015 International Plant and Animal Genome. 10-14 January, San Diego, CA.

Sandhu, H. S., R. A. Gilbert, J. C. Comstock, V. Gordon, N. El-Hout, and R. Arundale. 2014. New energy cane varieties in Florida. Annual meeting of ASA, CSSA and SSSA, 2-5 November, Long Beach, CA.

Todd, J., H.S. Sandhu, J. Wang, and B. Glaz. 2015. Evaluation of traits within the diversity panel of the World Collection of Sugarcane (*Saccharum* sp.) and related grasses. Joint Annual ASSCT meeting, 22-24 June, 2015, New Orleans, LA.

Sandhu, H. S., V. S. Gordon, J. Comstock, R. Gilbert, N. El-Hout, and R. Arundale. 2014. Development of energycane cultivars in Florida. Joint annual meeting of American Society of Sugar Cane Technologists, 18-20 June, Bonita Springs, FL.

Song J. (p) X. Yang (g), M. F. R. Resende Jr., L. G. Neves, J. Todd, J. Zhang, J. Comstock, J. Wang. Sequence variations in the *saccharum* complex revealed by next generation sequencing technologies. 2016. Sugar journal. 12.

You Q. (g), J. Song (p), X. Yang (g), L. Xu, J. Wang. Developing SNP array for sugarcane high throughput genotyping. 2016. Sugar journal. 20.

Song J. (p), S.N. Nayak,(p) A. Villa,(u) T. Ayala-Silva, X. Yang,(g) J. Todd, B. Glaz, J. Comstock, D.N. Kuhn, R. Gilbert, J. Wang. Characterizing and utilizing *Saccharum spp* genetic resources. Journal of American Society of Sugar Cane Technologists. 2014. 34: 57.

Todd J., B. Glaz, J. Wang, T. Ayala-Silva, N. C. Glynn, O. A. Gutierrez, D. Kuhn and J. Comstock. Phenotypic Characterization of the *Saccharum* spp. Diversity Panel Selected from the Miami World Collection of Sugarcane and Related Grasses. Journal of American Society of Sugar Cane Technologists. 2014. 34: 57.

Todd J., T. Ayala-Silva, B. Glaz, J. Wang, S. Sood, S. N Nayak (p), N.C. Glynn, O. A. Gutierrez, D. Kuhn, and J. C. Comstock. Core Selection from the Miami World

Collection of Sugarcane and Related grasses using phenotypic markers. Journal of American Society of Sugar Cane Technologists. 2013. 33: 60-61.

Nayak S. N. (p), A. Villa (u), J. Song (p), J. Todd, B. Glaz, T. Ayala-Silva, J. Comstock, D. N Kuhn, R. Gilbert, and J. Wang. Core Selection from the Miami World Collection of Sugarcane and Related Grasses using Genotypic Markers-Seeking Alleles for Biomass. Journal of American Society of Sugar Cane Technologists. 2013. 33: 61-62

Wang J., T.Ayala-Silva, S. Nayak (p), J. Todd, A. Villa (u), N. Glynn, X. Yang (g), B. Glaz, J. Comstock, D. Kuhn, O. Gutierrez, R. Gilbert. Exploring Broad Genetic Resources Available to Sugarcane. Proceedings of International Symposium on New Paradigms in Sugarcane Research, 2012. P172. Ed. by R. Viswanathan et al.

Nayak S. (p), B. Pathak (&), N. Glynn, B. Glaz, A.Villa (u), J. Patel (u), A. Bae (u), A. Vaccaro (u), T. Ayala-Silva, J. Comstock, R. Gilbert, and J. Wang. World Collection of Sugarcane and Related Grasses: Utilizing a Vast Genetic Resource. Sugar Journal. 2012. P26.

Wang J., J. Song (p), X. Yang (g), J. Todd. Target enrichment sequencing of 300 representative accessions in the sugarcane world germplasm collection. Plant & Animal Genome Conferences. Sugarcane Genomics Workshop. San Diego, California. 2017

Song J. (p) X. Yang (g), M. F. R. Resende Jr., L. G. Neves, J. Todd, J. Zhang, J. Comstock, J. Wang. Target sequencing reveals natural allelic variations in *Saccharum* Complex. XXIV Plant & Animal Genome Conferences. San Diego, CA. USA. 2016.

Song J. (p), X. Yang (g), S. Nayak (p), J. Todd. H. Sandhu, J. C. Comstock. J. Wang Sugarcane diversity and exome sequence variation. The ISSCT 11th Germplasm and Breeding Workshop/8th Molecular Biology Workshop. Reunion Island, France. 2015.

Wang J., J. Song (p), X. Yang (g), S. Nayak (p), J. Todd. H. Sandhu, J. C. Comstock. Diversity and exome variation in *Saccharum* spp. XXIII Plant & Animal Genome Conferences. San Diego, CA. USA. 2015.

Song J. (p), X. Yang (g), J. Wang. Single nucleotide polymorphism (SNP) discovery in the large and complex *Saccharum* spp. genomes through exome sequencing. XXIII Plant & Animal Genome Conferences. San Diego, CA. USA. 2015.

Song J. (p), S. Nayak (p), A. Villa (u), B. Pathak (&), T. Ayala-Silva, X. Yang (g) , J. Todd, N. C. Glynn , D.N. Kuhn, B. Glaz, R. Gilbert, J. C. Comstock, J. Wang. Promoting utilization of *Saccharum spp.* genetic resources. Florida Genetics 2014 symposium. Gainesville, Florida. 2014.

Wang J., J. Song (p), J. Todd, S. Nayak (p), J. Zhang, A. Villa (u), B. Glaz. Genetic diversity and genomic complexity in sugarcane. XXII Plant & Animal Genome Conferences. San Diego, CA. USA. 2014.

Todd J., N. C. Glynn, T. Ayala-Silva, B. Glaz, J. Wang, S. Sood. Nayak. S (p). Gutierrez O.A.,

K. D., T.M. and Comstock J. Phenotypic evaluation of the *World Collection of Sugarcane and Related Grasses.* XXI Plant & Animal Genome Conferences. San Diego, CA. USA. 2013.

Nayak S. N. (p), A. Villa (u), B. Pathak (&), J. Todd, T. Ayala-Silva, B. Glaz, J. Wang Genetic diversity of world collection of sugarcane and its related grasses. Florida Genetics 2012 Symposium. Gainesville, FL. 2012.

b. Web site or other Internet sites that reflect the results of this project;
https://www.wanglabuf.com/projects

c. Networks or collaborations fostered;
Collaboration with Drs. Noa Lincoln and Michael Kantar at University of Hawaii, Dr. Qianchun Zeng at Yunnan Agriculture University.

d. Technologies/Techniques;

Sugarcane target enrichment sequencing with probe sets

e. Inventions/Patent Applications, licensing agreements;
None

f. Other products, such as data or databases, physical collections, audio or video, software or netware, models, educational aid or curricula, instruments or equipment.