Final report
March 22, 2018

DOE project: Pyramiding genes and alleles for improving energy cane biomass yield

PI: Ray Ming, University of Illinois at Urbana-Champaign, Urbana, IL 61801
Co-PIs: Chifumi Nagai, Hawaii Agriculture Research Center, Kunia, HI 96759
Qingyi Yu, Texas A&M AgriLife Research, Dallas, TX 78596

Summary

The overall goal of this project is to identify genes and gene interaction networks contributed to the extreme segregants with 30 folds biomass yield difference in sugarcane F2 populations. Towards achieving this goal, yield trials of 108 F2 extreme segregants from *S. officinarum* LA Purple and *S. robustum* MOL5829 (LM population) were carried out in two locations in three years. A yield trial of the second F2 population from *S. officinarum* LA Purple and *S. spontaneum* US56-14-4 (LU population) was installed in the summer of 2014 and the first set of yield component data was collected. For genotyping, transcriptomes from leaves and stalks of 70 extreme segregants of the LM F2 population and 119 individuals of the LU F2 populations were sequenced. The genomes of 91 F1 individuals from the LM populations are being sequenced to construct ultra-high density genetic maps for each of the two parents for both assisting the LA Purple genome assembling and for testing a hypothesis of female restitution. The genomes of 110 F2 individuals from single F1 in the LU population, a different set from the 119 F2 individuals used for transcriptome sequencing, are being sequenced for mapping genes and QTLs affecting biomass yield and for testing a hypothesis of female restitution. Gene expression analysis between extreme segregants of high and low biomass yield showed up-regulation of cellulose synthase, cellulose, and xylan synthase in high biomass yield segregants among 3,274 genes differentially expressed between the two extremes. Our transcriptome results revealed not only the increment of cell wall biosynthesis pathway is essential, but the rapid turnover of certain cell wall polymers as well as carbohydrate partitioning are also important for recycling and energy conservation during rapid cell growth in high biomass sugarcane. Seventeen differentially expressed genes in auxin, one in ethylene and one in gibberellin related signaling and biosynthesis pathways were identified, which could potentially regulate biomass yield. Differentially expressed genes, PIF3 and EIL5, involved in gibberellin and ethylene pathway could play an important role in biomass accumulation. Differential gene expression analysis was also carried out on the LU population. High-biomass yield was mainly determined by assimilation of carbon in source tissues. The high-level expression of fermentative genes in the low-biomass group was likely induced by their low-energy status. The haploid (tetraploid) genome of *S. spontanium* AP85-441 was sequenced with chromosome level assembly and allele defined annotation. This reference genome along with the upcoming *S. officinarum* genome will allow us to identify genes and alleles contributed to biomass yield.

Keywords: Energy cane, extreme segregants, gene expression profiling, genomics, sugarcane

1. *Phenotyping extreme segregants of the F2 population for exploring the molecular basis of high biomass yield from transgressive segregation.*

Yield trials of extreme segregants from F2 population of LA Purple x MOL 5829:

From the yield trial of 261 F2 plants derived from interspecific cross *S. officinarum* LA Purple x *S. robustum* MOL5829 on Maui, 108 extreme segregants (54 plants from each end) along with 12 plants with special features were planted with three replications at Kunia Station on Oahu and yield data including a ratoon crop were collected in two years. Four of the top 7 clones didn't perform well in ratoon crop, perhaps reflecting poor root development associated with ratoon crop yield. The yield performance of the top 7 clones in three years combined ranking remained in top 10, although the order has been altered (Table 1).

Table 1. Biomass yield trials of extreme segregants of an F2 population at two locations in three years in Hawaii.

F2 clones	Maui	Oahu	Oahu ratoon		Maui	Oahu	Oahu ratoon	
	T/ha	T/ha	T/ha	Average	Rank	Rank	Rank	Rank
09-9011	122.1	163.8	86	123.9	1	1	30	**1**
09-9180	89.1	157.3	97.2	114.5	3	2	21	**2**
09-9188	30.2	151.1	141.5	107.6	42	3	2	**3**
09-9183	70.6	113.9	135	106.5	4	21	5	**4**
09-9006	63.8	118.2	136.7	106.2	6	15	4	**5**
09-9002	82.9	117.8	116.2	105.6	5	16	10	**6**
09-9019	77.5	112.6	107.9	99.3	7	24	14	**7**
09-9178	79.1	125.7	91.3	98.7	2	7	26	**8**
09-9078	25.1	108.2	159.3	97.5	66	33	1	**9**
09-9126	27.3	124.9	120	90.7	49	9	8	**10**
LA Purple	36.1	54.9	51.7	47.6				

Yield trial of *S. officinarum* LA Purple x *S. spontaneum* US56-14-4 F2 population

A yield trial with three replications using 164 F2 individuals along with LA Purple and the F1 plants started in July 2014. In March of 2015, 8.5-month yield component morphological data were collected, including diameter of three stalks, height of three stalks, and stalk number. Stalk volume was estimated for each plot using: 0.25 x 3.14 x (three stalk mean diameter)^2 x (three stalk mean height) x stalk number. The mean stalk volume was calculated for each clone and ranked. Average stalk volume of the top 3 F2 clones (126,385 cm^3) was 44% and 700% higher than that of F1 and the high yielding *S. officinarum* parent LA Purple, respectively. We don't have data for *S. spontaneum* parent since we cannot plant *S. spontaneum* in the field under the regulation by the State of Hawaii.

Generating additional F2 clones in five crosses

We continue to produce additional F2 individuals during the crossing season in November every year and 3 F2 populations have sufficient number of individuals for field trials (Table 2). All the F1 are maintained at Mawnawili Breeding Station.

Table 2. Five additional interspecific crosses among three *Saccharum* species (2010-2014)

	Female	Male	Species	F1 #	F2 2010	F2 2011	F2 2012	F2 2013	F2 2014	# Selected F2 in replicated trials
1	LA Purple	US56-14-4	OFF x SPT	9	62	50	11	357	340	164
2	LA Purple	MOL6136	OFF x ROB	49	0	10	198	7	198	180
3	MOL6081	MOL6136	ROB x ROB	5	0	0	11	30	37	0
4	MOL6081	US56-14-4	ROB x SPT	3	0	0	0	500	309	0
5	MOL6081	MPTH97-2	ROB x SPT	6	0	0	0	1	1	0

Figure 1. Sugarcane F2 field trials at Kunia station on Oahu, Hawaii

Field Trial in Texas

Based on the phenotype data collected in 2011 and 2012 in Hawaii, 86 F2 clones derived from the cross between LA Purple (*S. officinarum*) and MOL5829 (*S. robustum*) were selected and shipped to Weslaco, Texas for additional field trial in 2013. Due to harsh winter weather in 2013 and 2014, 38 clones were lost and 4 clones were very weak and couldn't generate enough seed pieces for field trial. Additional 30 F2 clones from the cross made in 2010 between LA Purple (*S. officinarum*) and MOL5829 (*S. robustum*) were selected (15 from high end and 15 from low end) and shipped to Texas to replace the lost clones for field trial in Texas in April 2015. USDA-APHIS requires that "distributed cane sets are subjected to a long hot water treatment prior to shipping". However, hot water treatment significantly reduced the germination rate, especially for the low-end clones. Therefore, field trial in Texas was not successful.

2. Mapping genes affecting biomass yield by transcriptome sequencing of the extreme segregants.

To achieve this objective, two F2 populations were used for field experiments and genotyping by transcriptome sequencing. The first F2 population was from a cross between *S. officinarum* LA Purple (2n = 8x = 80) and *S. robustum* MOL5829 (2n = 8x = 80) with 835 F2 individuals. This is the only F2 population in sugarcane with balancd chromosome numbers and

ploidy level from the two parents, and the population yielded extreme segregants showing 338% biomass yield increase, the best population for mapping genes affecting biomass yield.

We are taking a more aggressive genomic approach for characterizing this population by sequencing the genome of LA Purple, and constructing ultra-high density genetic maps for both parents. We have isolated DNA from 91 F1 individuals of this cross and each F1 individual will be sequenced using HiSeq2500 with 2 x 250bp paired-end reads at 20x coverage. We expect to map 1 million SNPs for each parent. With such ultra-high density linkage maps, we will be able to identify parental origin of recombination blocks in the F2 population.

While waiting for the completion of sequencing the F1 population this summer, RNA was extracted from leaf and internodes (3^{rd}, 9^{th}, and 15^{th}) of 70 F2 individuals and parents. RNAseq libraries were constructed and multiplexed libraries were sequenced by HiSeq2500 using paired-end reads. A total of 1.1 billion raw reads were obtained from sequencing, with an average of 15 million reads per individual. The SNP calling is in progress. We will compare the transcriptome-based F2 linkage map with the ultra-high density map of each parent to examine the completeness of the transcriptome-based map. We expect the transcriptome-based genetic map to be sufficient for mapping genes affecting biomass yield.

The second F2 population was from a cross between *S. officinarum* LA Purple (2n = 8x = 80) and *S. spontaneum* US56-14-4 (2n = 10x = 80). This F2 population is with balanced chromosome numbers but not balanced ploidy from the two parents. The LA Purple was a commercial cultivar in Louisiana (hence the LA) a century ago before disease resistant hybrid cultivars were used. US56-14-4 was widely used in the sugarcane breeding program in Louisiana, so the genes associated with biomass yield and sugar content will be highly relevant to Louisiana. This is a typical bi-parental cross used in sugar exhibiting female restitution with 2n + n transmission from the parents to F1 individuals and n + n transmission of F2 individuals. There are 697 F2 individuals in this population.

The leaf and internode (3^{rd}, 9^{th}, and 15^{th} internodes) tissues were harvested from the 119 F2 individuals and two parents of this second F2 population. Total RNA was isolated from each individual sample, and purified RNA samples were used for RNA-Seq library construction. A total of 242 RNA-Seq libraries were constructed and every 24 RNA-Seq libraries were pooled and sequenced by one lane of Illumina HiSeq2500. For the samples that fail to produce sufficient numbers of sequences, we re-constructed the RNA-Seq libraries, pool libraries, and re-sequenced. After 3 rounds of sequencing, a total of 3.2 billion reads were obtained, including 1.65 billion reads from leaf tissue and 1.57 billion reads from stalk internodes. SNP calling from the RNAseq data is ongoing.

We also took a more aggressive genomic approach to genotype this F2 population for mapping genes affecting biomass yield and for testing a hypothesis of female restitution. We previously sequencing transcriptomes from F1 individuals with 100 bp reads available then, but the reads are too short for separating haplotypes. With the current 2 x 250 bp paired ends reads, we can get 450-500 bp reads, long enough to separate haplotypes. Leaf samples from additional 110 F2 individuals originated from single F1 individual were collected for sequencing the genome of each individual using 2 x 250 bp paired-end reads by Illumina HiSeq2500 at 20x coverage. We hypothesized that the 2n transmission from female parent was doubling of the female gamete during or post fertilization and the 80 chromosomes from the female parent in each F1 individual are 2 sets of 40 identical chromosomes, but the 40 chromosomes in each F1 individual are randomly sorted from the 80 chromosomes. An ultra-high density genetic map of

the F2 population in comparison with ultra-high density map of LA Purple from the F1 population described above will accept or reject our hypothesis.

3. Identifying differentially expressed genes and alleles through analyses of transcriptomes of extreme segregants from the F2 population.

To elucidate the molecular mechanism contributing to high biomass phenotype in sugarcane, we analyzed the transcriptome of the extreme segregants from the LM F2 population, using the same RNAseq dataset generated from objective 2. Fourteen high biomass individuals and eight low biomass individuals were chosen and four tissues (first dew leaf, 3rd, 9th, and 15th internodes) were collected from field for RNA extraction and sequencing library construction. Multiplexed libraries were sequenced using Illumina HiSeq2500 paired end 150nt mode.

The trimmed reads from each individual were aligned to *Sorghum bicolor* gene annotation version 2.1 using Novoalign CS and aligned read count per gene was calculated by HtSeq-count. Differential expressed gene analysis was performed between high biomass group and low biomass group using edgeR, an R package. The principle component analysis plot showed that the transcriptome of high and low biomass groups was clustered separately (Fig 2a).

Out of 33,032 annotated sorghum genes, 3,274 genes were differentially expressed (DE; log2 fold change≥2, p-value ≤ 0.01 and FDR ≤ 0.05) between high and low biomass group. GO enrichment test of these DE genes revealed that they were over-represented in biological processes related cell wall modification (p=0.0177), cell wall macromolecule catabolic process (p=0.0444) and carbohydrate catabolic process (p=0.0010) (Fig 1b). This finding led us to study the gene expression difference of cell wall metabolism pathway between two groups.

Cell wall constitutes of five major components: cellulose, hemicelluloses, pectin, lignin and structural protein. Cellulose is composed of β (1→4) linked glucose and anabolism and catabolism is controlled by cellulose synthase and cellulase, respectively. We found that 5 cellulose synthases and 7 cellulases were significantly strongly expressed in high biomass group (Table 3). A similar phenomenon was observed in xylan biosynthesis, which produces hemicelluloses precursors. In this pathway, four genes including xylan synthase, were up-regulated in high biomass group. Cellulose and hemicelluloses are major components of cell wall and their active metabolism suggests that the plants are undergoing rapid cell division, which could contribute to faster growth rate and higher total plant weight.

However, in two pectin precursors (homogalacturonon and xylogalacturonan) biosynthesis pathway, no significant gene expression difference between two groups was observed. Instead, a vast number of up-regulated pectinesterase and polygalaturonase transcripts were detected in high biomass individuals (Fig 2b), indicating a rapid breakdown of pectin, which is required for loosening the adhesion between cells during cell expansion and development. We also observed a decrease of transcript level in genes involved in production of lignin precursors, which are essential secondary cell wall components.

Our transcriptome results revealed not only the increment of cell wall biosynthesis pathway is essential, but the rapid turnover of certain cell wall polymers is also important for recycling and energy conservation during rigorous cell growth in high biomass sugarcane. At the same time, high biomass cane restricts the secondary cell wall growth by suppression genes involved in lignin precursor biosynthesis, in order to allow continuous cell expansion and division. This study enhances our understanding on the gene interaction and expression that leading to plant growth, which will facilitate the selection of potential sugarcane varieties for biomass production.

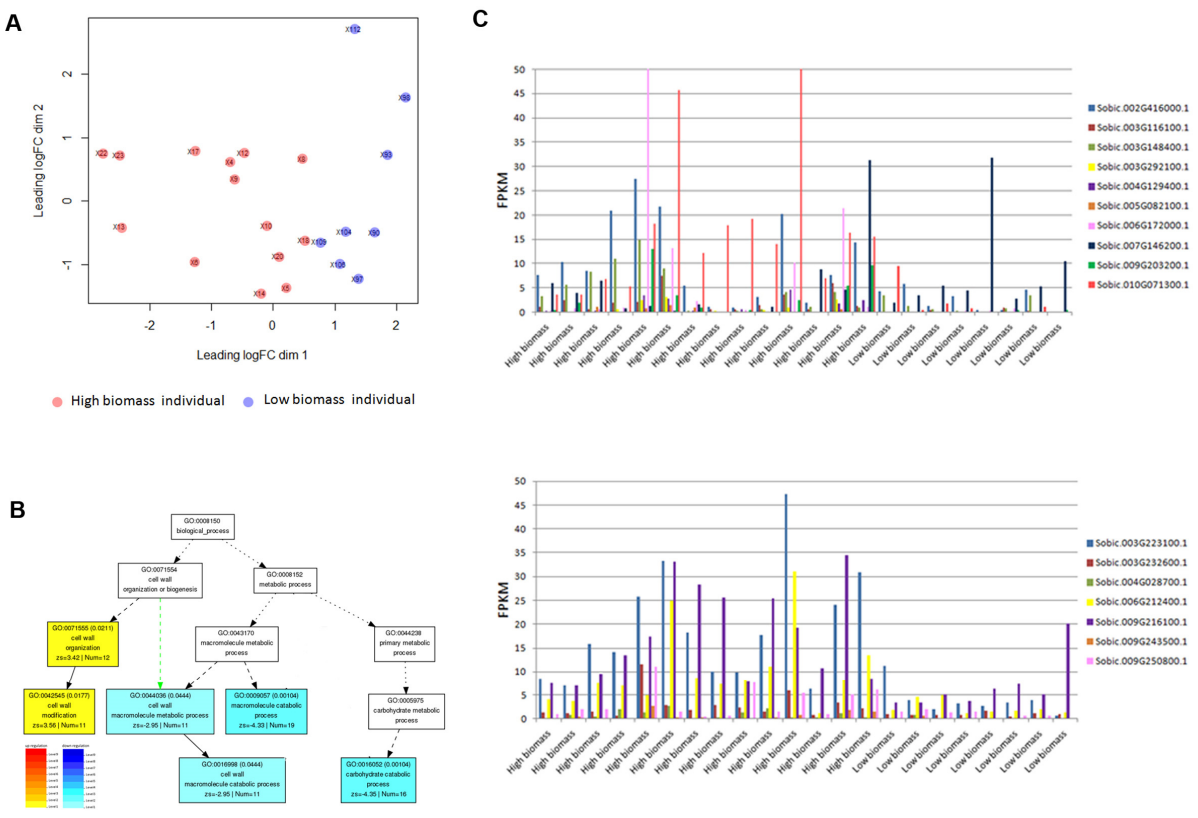

Figure 2. Differential expression analysis of transcriptomes between high and low biomass group. A) PCA plot of transcriptomes of high biomass group (red circle) and low biomass group (blue circle). B) Gene Ontology (GO) enrichment analysis of 3,274 differentially expressed genes. C) Expression level of polygalacturonase (upper panel) and pectinesterase (lower panel) genes in high biomass and low biomass individuals.

Table 3. Number of differentially expressed genes involved in cell wall biosynthesis and breakdown pathways between high and low biomass individuals.

Cell wall components	Biosynthesis pathway		Catabolism pathway	
	Up-regulate	Down-regulate	Up-regulate	Down-regulate
Cellulose	5	0	7	0
Hemicellulose	4	0	N/A	N/A
Pectin - Homogalacturonan	2	0	17	0
Pectin - Xylogalacturonan	2	1	0	0
Lignin - Monolignol	3	7	N/A	N/A

N/A: not studied

Gene expression profiling was used to detect genes involved in three important hormone-related pathways, auxin, ethylene and gibberellin, to find out how they are differently regulated between the extreme segregants of high and low biomass yield groups. We identified seventeen differentially expressed genes in auxin, one in ethylene and one in gibberellin related signaling and biosynthesis pathways, which could potentially regulate biomass yield. Differentially expressed genes, PIF3 and EIL5, involved in gibberellin and ethylene pathway could play an important role in biomass accumulation. These plant hormone-related genes could serve as candidate genes in genetic modification and breeding programs to develop high yielding energy cane.

Differential gene expression analysis was also carried out on the LU population. We sequenced transcriptomes of source and sink tissues from 12 selected extreme segregants to explore the molecular basis of high biomass yield for future breeding of high-yielding energy canes. Among the 103,664 assembled unigenes, 10,115 and 728 showed significant differential expression patterns between the two extreme segregating groups in the top visible dewlap leaf and the 9th culm internode, respectively. The most enriched functional categories were photosynthesis and fermentation in the high-biomass and the low-biomass groups, respectively. Our results revealed that high-biomass yield was mainly determined by assimilation of carbon in source tissues. The high-level expression of fermentative genes in the low-biomass group was likely induced by their low-energy status.

4. Developing gene- and allele-specific markers for implementation of marker-assisted selection in energy cane breeding programs.

We have sequenced an *S. spontaneum* haploid (tetraploid) genome of AP85-441, the simplest genome of *Saccharum* and the assembly, annotation, and analysis have completed. From a bacterial artificial chromosome (BAC) library of AP85-441, 35,156 BAC clones were pooled into 712 libraries (mostly of 48 BACs) and sequenced by Illumina HiSeq-2500 with PE150, yielding 267.5 Gbp data that were assembled using ALLPATH-LG[8], yielding a 2.56 Gbp assembly with contig N50 of 7.38 kb. To reduce fragmentation, 295 Gbp data from Pacific Biosciences (PacBio) RS II system was used for self-cgorrection and assembly by CANU, incorporating assembled BAC sequences, correcting and polishing with 90x Illumina paired end sequences, yielding 3.13 Gbp with contig N50 of 45 kb. The hybrid assembled contigs and BAC contigs correspond with ~ 99.3% accuracy.

To provide a scaffold for contig assembly, four Hi-C libraries were constructed from young leaves of AP85-441. Chimeric fragments representing the original cross-linked long-distance physical interactions were processed into paired-end sequencing libraries, then 1 billion of 150 bp paired-end Illumina reads were produced and uniquely mapped onto the draft assembly contigs. Due to polyploidy, existing Hi-C scaffolding programs such as LACHESIS and SALSA link *S. spontaneum* allelic haplotypes together and are no longer suitable for this autopolyploid genome. We developed a Hi-C-based scaffolding algorithm (ALLHIC) that integrates four functions, pruning, partition, optimization and building, to select contigs specific for assembly.

A high-density genetic map of 998,370 SNPs in 32 linkage groups was combined with the Hi-C-based physical map to assemble 32 pseudo-chromosomes that anchor 2.9 Gbp of genome including 97% of gene content. The 32 pseudo-chromosomes comprise 8 homologous groups with four sets of monoploid chromosomes, A, B, C and D. A total of 219 (88.3 %) complete gene models among 248 ultra-conserved core eukaryotic genes (CEGs) in CEGMA [11] and 1,397 (97.1 %) among 1,440 conserved genes in BUSCO were recalled in our assembly. Further, 1,624 million (97.01%) of 1,674 million Illumina short reads were alignable and

covered 97.3 % of the assembly. The high quality assembly allowed us to predict 28 potential centromeric regions along the 32 chromosomes, with length ranging from 0.25 Mb to 11.85 Mb.

Using two rounds of MAKER followed by manual annotation to separate genes and alleles, we annotated 35,525 genes with alleles defined, including 4,289 (12.7%) genes with four alleles, 9,792 (27.6%) with three, 14,797 (41.7%) with two, and 6,647 (18.7%) with one. The total number of alleles is 82,773 with an average 2.3 alleles per gene. In unanchored sequences, 3,130 gene/alleles were annotated and this could represent 1,361 genes based on 2.3 alleles per gene. We annotated 1,256 tandemly duplicated genes and 3,375 dispersedly duplicated genes (Table 4). The AP85-441 genome contains 36,886 non-redundant genes and 4,631 duplicated genes. The cytochrome P450 gene families illustrated the importance of annotating alleles in polyploid genomes, with total of 1,465 manually annotated alleles in 387 genes (Figure 3). Without allele specific annotation, the number of P450 genes in this genome would be 1,465, not 387.

We used SNP markers to distinguish different alleles of each gene and then conducted allele-specific expression analysis. A total of 643 SNP loci from 423 genes were identified to show group-specific expression between the two extreme biomass groups. Among the 423 genes with group-specific expression alleles, 184 could be assigned functions by annotation. Seven alleles of 7 photosynthesis-related genes showed group-specific expression patterns. Five of them showed expression only in the high-biomass group while two of them showed expression only in the low-biomass group. All the 7 photosynthesis-related genes have functions in light reactions of photosynthesis. A total of 4 alleles of 4 fermentation-related genes were identified to be group-specific expression alleles. Interestingly, all the 4 alleles were only expressed in the low-biomass group. These four fermentation-related genes encode key enzymes of the fermentation process, namely, an alcohol dehydrogenase (ADH) and three pyruvate decarboxylase (PDC). Alleles of 11 stress-related genes showed group-specific expression patterns. Seven of them were only expressed in the low-biomass group and four were only expressed in the high-biomass group.

Table 4. Allele annotation in AP85-441 genome

Chromosomes	Total No. of genes	No. of genes with 4 alleles	No. of genes with 3 alleles	No. of genes with 2 alleles	No. of genes with 1 allele	No. Of dispersely duplicated genes	No. of tandem duplicated genes
Chr1	6,677	682	1,663	2,903	1,429	654	211
Chr2	5,961	784	1,717	2,438	1,022	558	225
Chr3	5,097	525	1,419	2,158	995	443	180
Chr4	4,081	529	1,112	1,687	753	374	165
Chr5	4,325	476	1,077	1,852	920	391	145
Chr6	3,800	483	1,069	1,556	692	365	132
Chr7	4,013	516	1,135	1,643	719	427	139
Chr8	1,571	294	600	560	117	163	59
Genes with annotated alleles	35,525	4,289	9,792	14,797	6,647	/	/
Duplicated genes	4,631	/	/	/	/	3,375	1,256
Unanchored genes/alleles	3,130	/	/	/	/	/	/

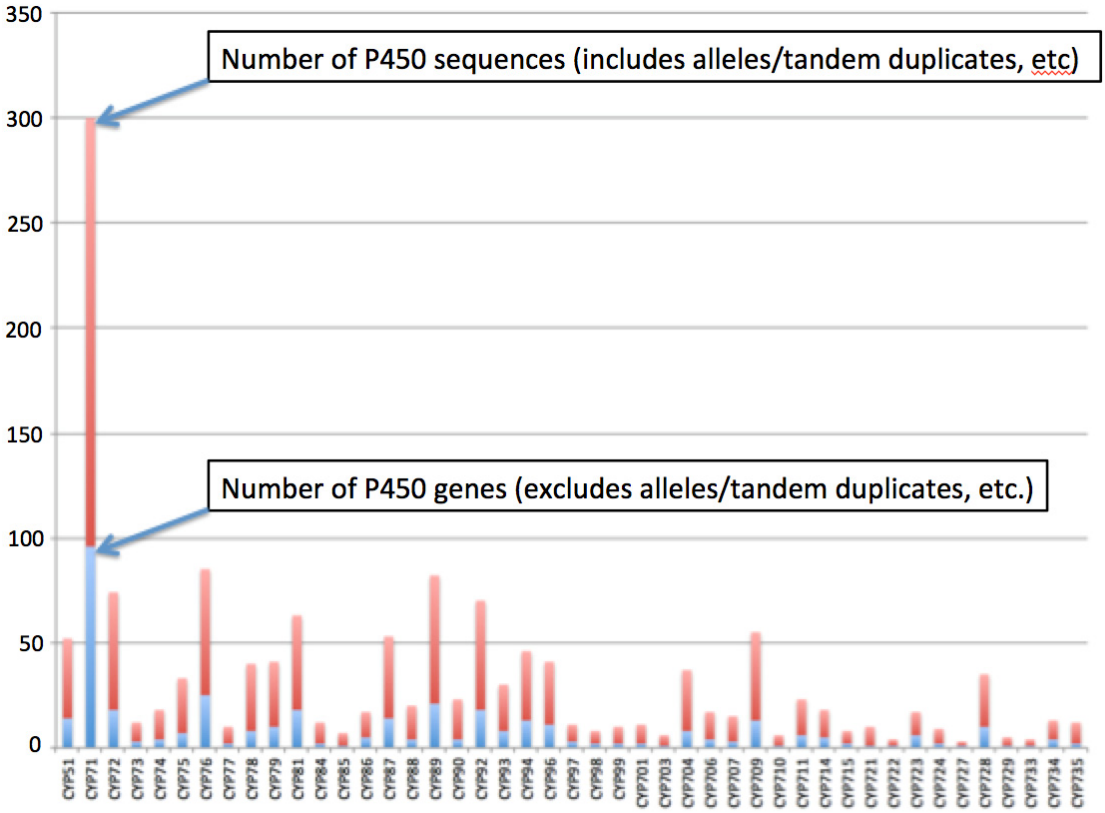

Figure 3. The number of cytochrome P450s by family. Red is the total number of P450s. Blue is the number of unique genes not counting alleles, tandem duplicates and other very close duplicates.

We are also sequencing the *S. officinarum* LA Purple genome using PacBio, Hi-C mapping, and ultra-high density genetic mapping to reach a chromosome level assembly, followed by allele defined annotation. The completion of these two genomes are essential for accomplish objectives 2 and 4, as SNP calling using its own genome as reference is more accurate and abundant, and for object 3, as allele specific expression analysis is only possible for allele define progenitor genome as references.

Publications:

1. Zhang, J. X. Zhang., H. Tang, Q. Zhang, X. Hua, X. Ma, F. Zhu, T. Jones, X. Zhu, J. Bowers, C. M. Wai, C. Zheng, Y. Shi, S. Chen, X. Xu, J. Yue, D. Nelson, L. Huang, Z. Li, H. Xu, D. Zhou, Y. Wang, W. Hu, J. Lin, Y. Deng, N. Pandey, M. Mancini, D. Zerpa, J.K Nguyen, L. Wang, L. Yu, Y. Xin, L. Ge, J. Arro, J.O. Han, S. Chakrabarty, M. Pushko, W. Zhang, Y. Ma, P. Ma, M. Lv F. Chen, G. Zheng, J. Xu, Z. Yang, F. Deng, X. Chen, Z. Liao, X. Zhang, Z. Lin, H. Lin, H. Yan, Z. Kuang, W. Zhong, P. Liang, G. Wang, Y. Yuan, J. Shi, J. Hou, J. Lin, J. Jin, P. Cao, Q. Shen, Q. Jiang, P. Zhou,Y. Ma, X. Zhang, R. Xu, J. Liu, Y. Zhou, H. Jia, Q. Ma, R. Qi, Z. Zhang, S.R. Dhungana, S.E. Huss, X. Yang, A. Sharma, J.H. Trujillo, M.C. Martinez, M. Hudson, J.J. Riascos, M. Schuler, L.Q. Chen, D.M. Braun, L. Li, Q. Yu, J. Wang, K. Wang, M.C. Schatz, D. Heckerman, M.-A. Van Sluys, G.M. Souza, P.H. Moore, D. Sankoff, R. VanBuren, A.H. Paterson, C. Nagai, R. Ming 2018. Allele defined genome of the usingautopolyploid *Saccharum spontaneum* L. (in review)
2. Zhang, J., Q. Zhang, L. Li, H. Tang, Q. Zhang, Y. Chen, J. Arro, X. Zhang, A. Wang, C. Miao, R. Ming. 2018. Recent Polyploidization Events in Three *Saccharum* Founding Species. (in review)
3. Sharma, A., J. Song, Q. Lin, R. Singh, N. Ramos, K. Wang, J. Zhang, R. Ming, Q. Yu. 2018. Comparative analysis of homologous sequences of *Saccharum officinarum* and *Saccharum spontaneum* reveals independent polyploidization events. Frontiers in Plant Science (*in review*)
4. Singh, R., T. Jones, C.M. Wai, J. Jifon, C. Nagai, R. Ming, Q. Yu. 2018. Transcriptomic analysis of transgressive segregants revealed the central role of photosynthetic capacity and efficiency in biomass accumulation in sugarcane. Scientific Report. 8:4415
5. Zhu, F., C. M. Wai, J. Zhang, T.C. Jones, C. Nagai, R. Ming. 2018. Differential expression of hormone related genes between extreme segregants of a *Saccharum* interspecific F2 population. Euphytica 214: 55. https://doi.org/10.1007/s10681-018-2137-z
6. Wai, C.M., J. Zhang, T. Jones, C. Nagai, R. Ming. 2017. Cell wall metabolism and hexose allocation contribute to biomass accumulation in high yielding extreme segregants of a *Saccharum* interspecific F2 population. BMC Genomics. 18:773

7. Zhang, J., A. Sharma, Q. Yu, J. Wang, L. Li, L. Zhu, X. Zhang, Y, Chen, R. Ming. 2016 Comparative structural analysis of Bru1 region homeologs in *Saccharum spontaneum* and *S. officinarum*. BMC Genomics. 17:446.
8. Arro, J., J.-W. Park, C. M. Wai, R. VanBuren, Y.-B. Pan, C. Nagai, J. da Silva, R. Ming. 2016. Balancing selection contributed to domestication of autopolyploid sugarcane (*Saccharum officinarum* L.). Euphytica. 209:477-493.