Napiergrass (*Cenchrus purpureus* Schumach.) Genome Survey and High-Density Genetic Map Construction

Dev Paudel[1], Baskaran Kannan[1], Xiping Yang[1], Karen Harris-Shultz[2], Mahender Thudi[3], Rajeev K. Varshney[3], Fredy Altpeter[1], and Jianping Wang[1*]

[1]Agronomy Department, Plant Molecular and Cellular Biology Program, Genetics Institute, University of Florida, IFAS, Gainesville, FL; [2]Crop Genetics and Breeding Research Unit, USDA-Agricultural Research Service, 115 Coastal Way, Tifton, GA; [3]Center of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

*Author for correspondence: wangjp@ufl.edu

ABSTRACT

Napiergrass (*Cenchrus purpureus* Schumach.) is a tropical forage grass and a promising lignocellulosic biofuel feedstock due to its high biomass yield, persistence, and nutritive value. However, its utilization for breeding has lagged behind other crops due to limited genetic and genomic resources. In this study, next-generation sequencing was first used to survey the genome of napiergrass. Napiergrass sequences displayed high synteny to the pearl millet genome and showed expansions in the pearl millet genome along with genomic rearrangements between the two genomes. Furthermore, to construct a high density genetic map of napiergrass, genotyping-by-sequencing (GBS) was employed in a bi-parental population of 185 F_1 hybrids. A total of 512 million high quality reads were generated and 287,093 SNPs were called by using multiple *de-novo* and reference-based SNP callers. Single dose SNPs were used to construct the first high density linkage map that resulted in 1,913 SNPs mapped to 14 linkage groups, spanning a length of 1,410 cM and a density of 1 marker per 0.73 cM. This map can be used for many further genetic and genomic studies in napiergrass and related species.

INTRODUCTION

- Napiergrass (elephant grass) is a tropical perennial C4 grass
- It is an allotetraploid ($2n = 4x = 28$, A'A'BB)
- Cultivated primarily for forage and widely used by smallholder dairy farmers in Africa
- Important cellulosic energy crop due to its high dry biomass yield
- Dry matter yield is highest in napiergrass compared to other bioenergy crops like sorghum, switchgrass, or sugarcane [1].
- Breeding of napiergrass for desirable traits is lagging
  - Little genetic information available
  - Absence of genetic map
  - Absence of reference genome
  - Molecular tools haven't been deployed in breeding programs

OBJECTIVES

- Compare software for SNP calling in napiergrass
- Construct high-density genetic map of napiergrass
- Investigate genomic and genetic architecture of napiergrass

MATERIALS & METHODS

- Contrasting napiergrass parents N190 and N122 were crossed
  - **N190** ♀
    - ✓Late flowering
    - ✓Reduced number of thick tillers
    - ✓High biomass but not persistent
  - **N122** ♂
    - ✓Early flowering
    - ✓Prolific tillering with thin stalks
- Mapping population of 185 F1 hybrids was developed
- DNA was extracted from leaves using Dellaporta protocol [2]
- DNA samples were submitted to Cornell University for Genotyping by Sequencing (GBS) [3]
- Enzyme used for digestion in GBS: *PstI*
- Sequenced on 2 lanes of Illumina HiSeq 2000
- SNPs were called according to the software used
- Linkage map was constructed using JoinMap 4.1
  - Marker grouping: LOD 20
  - Regression-mapping algorithm
  - Kosambi function

N190 ♀ N122 ♂

Figure 1. Napiergrass parents used to develop mapping population

RESULTS

- Number of SNPs called by software varied considerably (Fig. 2)

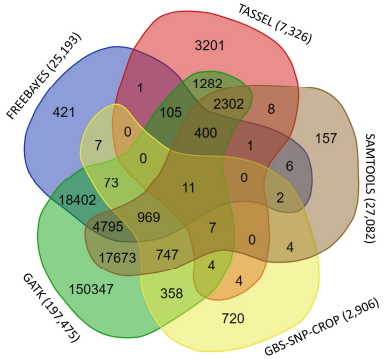

Figure 2. Venn diagram showing concordant napiergrass SNPs called by five reference-based SNP callers, SAMtools, GBS-SNP-CROP, GATK, FreeBayes, and TASSEL. Numbers in parenthesis after the program name shows the total number of SNPs called by each program.

- 1,913 single dose SNPs were mapped to 14 linkage groups, spanning a length of 1,410 cM and a density of 1 marker per 0.73 cM (Fig. 3)

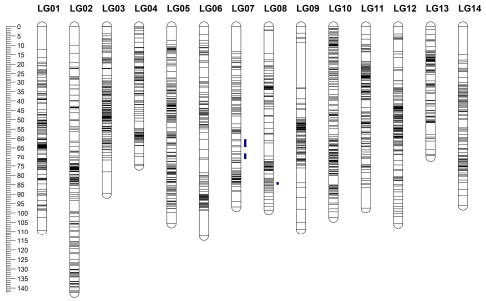

Figure 3. Genotyping by sequencing single nucleotide polymorphism (GBS-SNP) marker distribution for the 14 linkage groups of napiergrass. A black bar means a GBS-SNP marker. A blue bar represents segregation distortion region. The left scale plate represents genetic distance (centiMorgan as unit).

- TASSEL *de-novo* UNEAK showed the highest percentage of SNPs mapped followed by Stacks and SAMtools (Fig. 4)

Software	SNPs called	Useful SNPs	Mapped SNPs	Mapped SNPs (%)
FreeBayes	25,193	6	0	0
GATK	197,475	52	5	0.26
SAMtools	27,082	3,377	151	7.89
GBS-SNP-CROP	2,906	115	52	2.72
TASSEL	7,326	116	56	2.93
Stacks	4,920	447	257	13.43
GBS-SNP-CROP *de-novo*	4,521	96	51	2.67
Stacks *de-novo*	6,871	339	185	9.67
TASSEL *de-novo* UNEAK	10,799	2523	1156	60.43
Total	**287,093**	**7,071**	**1,913**	

Figure 4. Summary of napiergrass single nucleotide polymorphism (SNP) markers mapped on the combined linkage map using 9 different software pipelines

RESULTS

- Collinearity was observed between the napiergrass and pearl millet genome [4]
- For each pearl millet pseudomolecule, two corresponding regions in the napiergrass linkage groups were identified (Fig. 5)

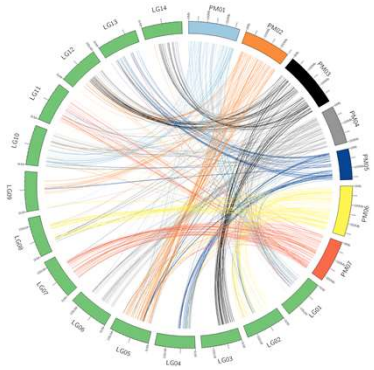

Figure 5: Syntenic regions between napiergrass linkage groups and the pearl millet genome. PM01 to PM07 are pearl millet pseudomolecules, L G01 to LG14 are napiergrass linkage groups. The small dots represent significant BLAST hits of mapped UNEAK tags to the pearl millet genome (>80% identity and >50 bp length).

CONCLUSION

- GBS was successfully used to construct a high density genetic map of napiergrass
- The genetic map had 1,913 SNP markers grouped into 14 linkage groups, spanning a length of 1410 cM and a density of 1 marker per 0.73 cM.
- TASSEL *de-novo* UNEAK pipeline retained highest number of SNPs on genetic map followed by STACKS.
- Tags of napiergrass showed high collinearity to pearl millet genome.
- For most of the pearl millet chromosome, there were two corresponding regions on the napiergrass genome.

REFERENCES

[1] Byrt, Caitlin S., Christopher PL Grof, and Robert T. Furbank. "C4 Plants as Biofuel Feedstocks: Optimising Biomass Production and Feedstock Quality from a Lignocellulosic Perspective." *Journal of integrative plant biology* 53.2 (2011): 120-135.

[2] Dellaporta, Stephen L., Jonathan Wood, and James B. Hicks. "A plant DNA minipreparation: version II." *Plant molecular biology reporter* 1.4 (1983): 19-21.

[3] Elshire, Robert J., et al. "A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species." *PloS one* 6.5 (2011): e19379.

[4] Varshney, R. K., *et al.* Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* (2017).

ACKNOWLEDGEMENT

- Plant Breeders Workgroup, University of Florida
- USDA National Institute of Food and Agriculture

CONNECT

- Email: dpaudel@outlook.com

devRpaudel