Fine-scale variation of nuclear and chloroplast genomes in a Johnsongrass (*Sorghum halepense*) population

Andrew Busby, Caitlin McCann, William Laycock, Kayln Pate, Nathaniel P. Hofford, M'Kayla G. Motley, Michael R. McKain
Department of Biological Sciences, The University of Alabama

Introduction

- Invasive species may have a particularly large impact on native ecosystems as well as exerting financial pressure on infrastructure and agriculture. Understanding the dynamics by which invasions occur aids further efforts to combat the spread of invasive species.
- Both whole chloroplast genome and transposome analyses can provide insight into population dynamics.
- We analyzed a population of Johnsongrass (*Sorghum halepense*), an especially invasive grass known for its ability to spread prolifically through North American grasslands and agricultural areas, from a site on Rice Mine Road in Tuscaloosa, Alabama.
- Johnsongrass can reproduce both asexually, forming clonal stands via rhizomatous growth and sexually through the mass dispersal of seeds.
- Johnsongrass' propensity for rapid spread raises question to its dominant reproductive method and how it may explain Johnsongrass' invasiveness.
- Our analysis provides insight to how we may answer this question, provides a model for how Johnsongrass invasions occur, and raises further questions to be answered in the future.

Questions

- How might the Rice Mine population fit in a phylogeny of other populations?
- How many unique chloroplast genomes exist in the population?
- What is the degree of clonality in the population, and how does this vary throughout the population?
- How does transposon composition and abundance vary within and among populations?

Methods

Sampling

- Sampling from Rice Mine population was conducted using a transect-based approach using quadrats spaced 10 meters apart.

Preparation

- DNA was isolated from silica dried leaves using a modified CTAB protocol adapted from Doyle and Doyle [1].
- DNA quantity was checked using a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and a Qubit 2.0 (Thermo Fisher Scientific). Double-stranded DNA concentrations were used to determine starting material for library preparation.
- Sequencing libraries were made using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, MA, USA.)
- Libraries were size selected for 300-400 base pair insert sizes. Libraries were sequenced using paired-end 150 base pair reads on an Illumina NovaSeq 6000 at Vanderbilt University.

Analysis

- Fast-Plast [2] was used to assemble whole chloroplast genomes, which were then aligned using MAFFT v.7.313 [3].
- Phylogenetic relationships were reconstructed using under the GTR+gamma model in RAxML v.8.2.4 [4] with 500 bootstraps and PopART v.1.0 [5] was used to estimate a haplotype network.
- Organellar reads were removed by mapping trimmed reads to the chloroplast and mitochondrial genomes of *Sorghum bicolor* using bowtie v.2.3.0 [6] under the "very-sensitive-local" option. Unmapped reads were used to estimate transposon composition and abundance using Transposome v.0.12.1 [7] with a percent identify of 90 and a fraction coverage of 0.55.

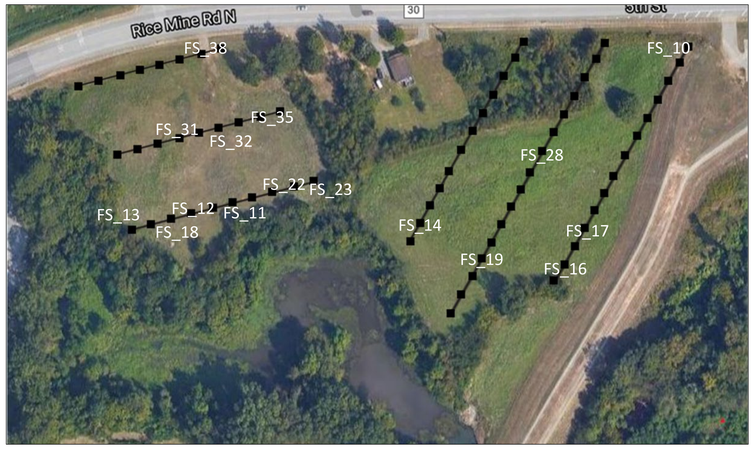

Figure 1. Transect sampling scheme and position of samples sequenced for this study from the Rice Mine population. A total of 1,185 samples were taken across the transects. Sixteen samples were sequenced for this study as a preliminary analysis of chloroplast diversity in the population.

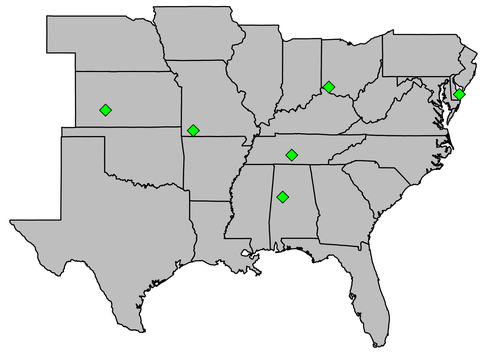

Figure 2. Locations where all samples added for this study were collected. All sites, except for Tuscaloosa, AL, represent a single individual. Location information provided in names in **Figure 4**.

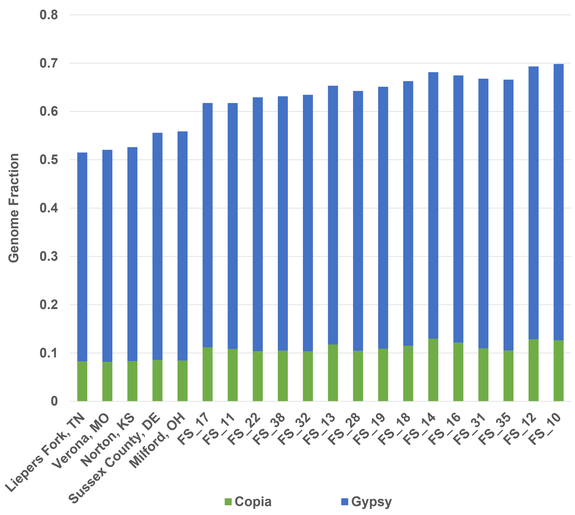

Figure 3. Genome fraction of *Copia* and *Gypsy* retrotransposons per sample. These transposons make up 51.5% (Liepers Fork, TN sample) to 69.8% (FS_10 sample) of the total genome in Johnsongrass accessions sampled.

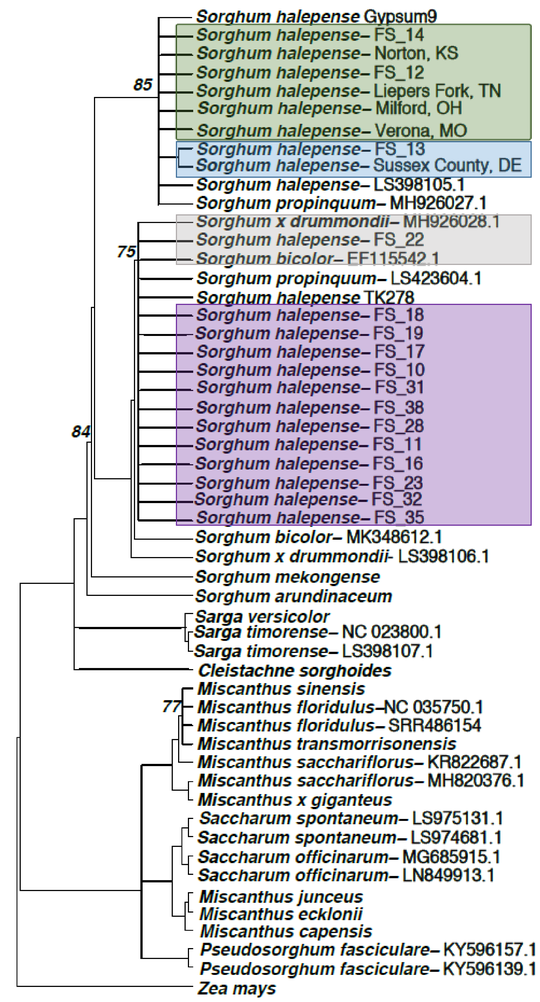

Figure 4. Maximum likelihood phylogeny of Johnsongrass and close relatives estimated from chloroplast genomes. **Green**, **blue**, grey, and **purple** blocks represent three sets of samples with identical chloroplast genomes. Nodes with less than 75 BSV collapsed.

Results

- The field at Rice Mine road consists mainly of a single chloroplast haplotype.
- A few individuals with alternate haplotypes exist in the interior of the population (Individuals **FS_12**, **FS_13**, and **FS_14** in **Figure 1**).
- Of the 24 Johnsongrass accessions sampled for this study, two unique haplotype groups (represented by clades in the phylogeny) were found. Each group has two to three derived haplotypes demonstrating mutational variation.
- No geographical relationships are identified based on haplotype occurrence. More sampling is needed to verify this.
- Both parental species of Johnsongrass, *Sorghum bicolor* and *Sorghum propinquum*, share chloroplasts with Johnsongrass.
- Transposable element composition was found to be variable both within and among different populations of Johnsongrass and notably between individuals with identical chloroplast haplotypes.

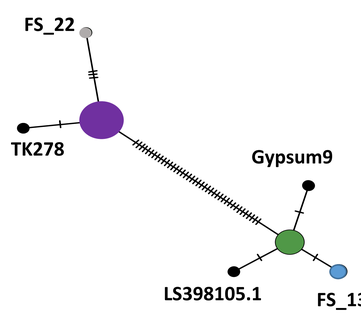

Figure 5. Haplotype network of all *Sorghum halepense* samples, both newly generated ones and those available publicly. Each circle represents a unique haplotype. Colors correspond to those highlighted in **Figure 4.** Hatch marks represent mutations between haplotypes.

Conclusions

- The majority of the population consists of a single haplotype, suggesting the population is either clonal or stems from a single maternal line; however, transposable element composition is variable.
- The outlying individuals represent wholly distinct haplotypes from the rest of the population and are evidence of multiple invasion events at the Rice Mine site.
- The spatial distribution of those distinct haplotypes may indicate that invasions have occurred in waves with newly introduce Johnsongrass pushing the established population deeper into the field away from the road over time.
- Haplotype diversity in the United States is very low based on the samples used in this study; this may be an indication of a low number of successful introductions with much of the diversity coming from just a few maternal lines.
- Future sampling will include more individuals from across the US range and look at the nuclear genome.

Acknowledgements

Funding for this project provided by CPING NSF Grant to MRM (Award #1920858).

References

1. Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11– 15.
2. Michael McKain, & afinit. (2017, September 23). mrmckain/Fast-Plast: Fast-Plast v.1.2.6 (Version v.1.2.6). Zenodo. http://doi.org/10.5281/zenodo.973887
3. Katoh, K., & Standley, DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.
4. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.
5. Leigh, JW, & Bryant, D. (2015). popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution, 6(9), 1110-1116.
6. Langmead, B. and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2.Nature Methods, 9:357-359.
7. Staton, SE, and Burke, JM. 2015. Transposome: A toolkit for annotation of transposable element families from unassembled sequence reads Bioinformatics, 31:1827-1829.