Title: Leveraging *Setaria viridis* and *Sorghum bicolor* pan-genomic variation for gene mining and crop improvement

Scott J. Lee[1]* (slee@danforthcenter.org), Adam Healey[2], Zhenbin Hu[3], Philip Ozersky[1], Jeremy Schmutz[2], Geoffrey P. Morris[3], **Ivan Baxter**[1], and Todd C. Mockler[1]

[1]Donald Danforth Plant Science Center, 975 North Warson Rd., Saint Louis, MO, USA,

[2]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806

[3]Kansas State University, 1712 Claflin Rd., Manhattan, KS 66506

http://www.foxmillet.org

Project Goals: This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock Sorghum bicolor to enhance water use and photosynthetic efficiencies.

The emergence of *Sorghum bicolor* as premier bioenergy crop is based on its naturally high yield, growth habit, and abiotic stress resistance. A key untapped avenue for sorghum breeders is the characterization of novel and wild alleles. Low coverage sequencing is inadequate to capture all the possible variation within diverse germplasm, especially for genes known to exhibit presence/absence variation. Model grass systems such as *Setaria viridis* can provide additional insights and diversity towards the discovery of novel and useful alleles for the development of bioenergy sorghum. The *de novo* assembly of a diverse collection of *Setaria viridis* and *Sorghum bicolor* lines will provide a novel resource in order to accelerate the development and breeding of sorghum as a bioenergy crop.

We performed high coverage Illumina-based sequencing of a diverse collection of 363 *S. bicolor* lines from the Sorghum Bioenergy Association Panel (BAP), along with de novo assembly of a subset of lines in order to begin to characterize the sorghum pan-genome. These BAP lines are a phenotypically, geographically, and racially diverse collection of sorghum, and were grown in both field and controlled greenhouse conditions. Following phenotyping and tissue collection, the lines were sequenced using the Illumina platform and were then aligned to the reference sorghum genome as well as de novo assembled to capture pangenomic variation. Additional genetic resource development for *Setaria viridis* will aid in these approaches. We performed genome-wide association studies (GWAS) of the phenotypes collected in sorghum, as well as analysis of observed polymorphisms, population structure, genomic selection, and presence/absence variation. We identified several known genes that show signs of selection as well as exhibit PAV within distinct subpopulations of the BAP panel. Additionally, we have identified several potentially novel genes that may be utilized to enhance biomass accumulation and water-use efficiency.