Using Natural Variation in Switchgrass to Identify Genes Relevant for Biomass Production and Improved Breeding Strategies

Joseph Evans1, Megan Bowman1, Jeongwoon Kim1, Kevin Childs1, Emily Crisovani1, Brieange Vaillancourt1, Aruna Nandety2, Shawn Kaeppler, Michael Casler4, and C. Robin Buell1*
(buell@msu.edu)

1Michigan State University, Dept. of Plant Biology, East Lansing MI; 2University of Oklahoma, Department of Plant Biology and Microbiology, Norman, OK; 3University of Wisconsin, Dept. of Agronomy, Madison, WI; 4USDA-ARS, U.S. Dairy Forage Research Center, 1925 Linden Dr., Madison, WI

https://www.glbrc.org

Project Goals

Switchgrass (Panicum virgatum L.) is a polyploid, perennial grass species that is native to North America, and is being developed as a biofuel feedstock crop. Switchgrass is present primarily in two ecotypes: a northern upland ecotype composed of mixed tetra- and octoploid switchgrasses, and a southern lowland ecotype composed of primarily tetraploid switchgrasses. While the ecotype designations are relatively well understood due to previous work using genotyping-by-sequencing and chloroplast simple sequence repeat markers, the development of switchgrass as a biofuel feedstock crop will require more detailed genetic information that can drive improvement efforts including accelerated and targeted breeding efforts. We developed a custom repeat library to mask repetitive sequences in the genome, and then utilized a high-coverage exome capture sequencing approach with the switchgrass v.1.1 genome to sequence a diversity panel composed of 537 individuals spanning 45 upland and 21 lowland switchgrass populations. From these data, we identified over 12 million polymorphic loci, from which, we selected a highly filtered and robust subset of 1.5 million high fidelity loci. From these polymorphisms, we were able to identify distinct population groups within the upland and lowland ecotypes, a result that was further supported through genetic distance analysis. We were able to also identify several large effect SNPs in flowering time genes that appeared to be ecotype specific, which may help contribute to the large phenotypic differences between ecotypes. We also identified 16,971 up-copy number variants, 108,475 down-copy number variants, and 14,178 presence-absence variants. Using these copy number variants in a genetic distance analysis, we were able to confirm our earlier population clustering results, which provides evidence that there may be large groups of both ecotype and population specific copy number variants. These polymorphisms are being incorporated into genome wide association studies and genome selection experiments to identify individuals with improved biomass yield and biofuel feedstock traits.

This work was funded by the Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).