Title: Impact of genome assemblies, genotyping methods, variant types, and ploidy levels on genomic prediction in switchgrass

Authors: Peipei Wang[1,2]* (peipeiw@msu.edu), Fanrui Meng[1,2], Christina B Azodi[4], Michael D. Casler[5], and **Shin-Han Shiu**[1,2,3]

Institutions: [1]Department of Plant Biology; [2]DOE Great Lakes Bioenergy Research Center; [3]Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA; [4]St.Vincent's Institute of Medical Research, Fitzroy 3065, Victoria, Australia; [5]USDA, U.S. Dairy Forage Research Center, Madison, WI and DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706, USA

Project Goals: Assess factors influencing genomic prediction accuracy in switchgrass; optimize predictions of 20 traits; probe the genetic basis underlying multiple target traits.

Abstract text: Genomic prediction where genotype information is used to predict phenotypes has accelerated breeding processes[1-3] and can provide mechanistic insights into phenotypes of interest. Switchgrass (*Panicum virgatum* L.) is a perennial biofuel feedstock with multiple traits targeted for accelerated breeding using genomic prediction approaches. To optimize switchgrass genomic prediction, we assessed the impact of genome assembly versions, marker sequencing strategies, marker types, marker allelic complexities, and polyploidy levels to predict 20 traits in a switchgrass association panel with 510 individuals[4-6]. We found that genomic prediction models performed similarly regardless of the genome assembly: v.1 or the recently available v.5 assembly. This occurred because the majority of variants—e.g., 70.7% of the single nucleotide polymorphism (SNP) markers from Genotyping-By-Sequencing (GBS)—are shared between these assemblies. Models using markers generated with exome capture outperformed those based on GBS markers for five traits. But in four traits, GBS marker-based models had higher prediction accuracy because the variants underlying the polymorphisms of these four traits tend to be located in intergenic regions. We also found that SNP-based models performed better than Insertion/Deletion (indel) based models for 12 traits, and biallelic marker-based models outperformed models using multiallelic markers for 17 traits. This was due to more SNPs and biallelic markers than indels and multiallelic markers, respectively, as models built with the same numbers of markers had similar accuracy. The most significant model improvement was observed when tetraploids were separated from octoploids, which can be partially explained by the higher trait variances in tetraploid populations. By considering the population structures and factors mentioned above, we present improved genomic prediction models for each of the 20 traits. Finally, we identify candidate genes that are the genetic basis underlying multiple target traits by studying the markers that have the greatest impact on model performance. Our study provides insights into the best practices for performing genomic prediction, which can be used for improving switchgrass agronomic traits through selective breeding.

References/Publications

1 VanRaden, P. M. *et al.* (2009) Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92(1), 16-24.

2 Lorenzana, R. E. & Bernardo, R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120(1), 151-161.
3 Crossa, J. *et al.* (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22(11), 961-975.
4 Lu, F. *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9(1), e1003215.
5 Lipka, A. E. *et al.* (2014) Accelerating the Switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS One* 9(11), e112227.
6 Evans, J. *et al.* (2015) Diversity and population structure of northern switchgrass as revealed through exome capture sequencing. *Plant J.* 84(4), 800-815.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018409.