

B-209

Population and quantitative genomic properties of the USDA soybean germplasm collection

*Katy Rainey\**, Department of Agronomy, Purdue University, Indiana, USA

*Alencar Xavier*, Dow AgroSciences, Indiana, USA

*Rima Thapa*, Syngenta Seeds, Oklahoma, USA

*William Muir*, Department of Animal Sciences, Purdue University, Indiana, USA

This study is an assessment of the entire soybean collection of the United States Department of Agriculture National Plant Germplasm System (USDA) that reports quantitative- and population genomic parameters along with phylogenetics. It also provides a new insight into soybean germplasm structure. Germplasm studies enable plant breeders to incorporate novel genetic resources into breeding pipelines to improve valuable agronomic traits. We conducted comprehensive analyses on the 19,652 soybean accessions in the USDA-ARS germplasm collection, genotyped with the SoySNP50K iSelect BeadChip SNP array, to elucidate the quantitative properties of existing subpopulations. Inferred through Hierarchical clustering performed with the Ward's D agglomeration method with and Nei's standard genetic distance. We found the effective population size to be approximately 106 individuals based on the linkage disequilibrium of unlinked loci. A cladogram indicated the existence of eight major clusters (C1–C8). Each cluster displays particular properties with regard to major quantitative traits. Among those, Cluster 3 represents the tropical and semi-tropical genetic material, Cluster 5 displays large seeds and may represent food-grade germplasm, and Cluster 7 represents the undomesticated material in the germplasm collection. The average  $F_{ST}$  among clusters was 0.22 and a total of 914 SNPs were exclusive to specific clusters.

Our classification and characterization of the germplasm collection into major clusters provides valuable information about the genetic resources available to soybean breeders and researchers.