# Four Haplotypes, One Genome: Uncovering Hidden Structural Diversity in Alfalfa RegenSY27x Genotype

*Harpreet Kaur, University of Minnesota*
*Melinda Dornbusch, USDA-ARS*
*Cameron Connor, National Center for Genome Resources*
*Adam Gomez, National Center for Genome Resources*
*Joann Mudge, National Center for Genome Resources*
*Andrew Farmer, National Center for Genome Resources*
*Laura Shannon, University of Minnesota*
*Deborah Samac, USDA-ARS*

Polyploid genome assembly is frequently confounded by high heterozygosity and complex haplotype structures. We report a chromosome-scale, haplotype-resolved assembly of RegenSY27x, an autotetraploid alfalfa genotype known for its superior regenerative capacity in tissue culture. By integrating PacBio high-fidelity long reads, Omni-C scaffolding, and a phased genetic linkage map, we generated a 3.2 Gb assembly partitioned into four distinct haplotypes.

Linkage maps were instrumental in phasing the four haplotypes and correcting chimeric scaffolds, ensuring the structural integrity of the final assembly. Linkage maps were based on genotyping-by-sequencing (GBS) markers. We evaluated the effects of reference genome choice and population size on mapping using two biparental F1 populations with RegenSYx27 crossed with UMN3988, a non-lodging genotype or with FL99, a fall dormancy 9 genotype. While a graph-based pangenome improved marker discovery in regions with complex structural variation (SV), population size was the most significant factor in maintaining genetic map stability across different reference genomes.

A critical finding of this study is the presence of extensive SV among the RegenSY27x homologs. Comparative analysis revealed tens of thousands of duplications, inversions, and translocations between haplotypes, demonstrating that a single autotetraploid genome captures a vast reservoir of structural diversity which is often collapsed in consensus assemblies. Beyond structural diversity, gene annotation identified 221,688 protein-coding genes, with repetitive elements, primarily long terminal repeats retrotransposons and helitrons, comprising 62.7% of the genome. The spatial clustering of helitrons, a class of transposable elements representing 5.52% of the genome, in gene-dense regions suggests their role as key drivers of genomic innovation and gene family expansion. Furthermore, the identification of 3,696 disease-resistant (R) genes revealed large tandem R-gene clusters. We have identified genes with allelic copies in this haplotype-resolved genome. These genes are being used to conduct preferential allele-specific expression analyses across multiple tissues, and these findings will also be discussed.