

Structural Genomics of Neotetraploid Alfalfa to Explore the Process of Polyploidization

Aaron Anderson, University of California-Davis & University of Perugia

Danilo Fabrizio Santoro, University of Perugia

Nooshin Alavi, University of California-Davis & University of Perugia

Emidio Albertini, University of Perugia

E. Charles Brummer, University of California-Davis

Daniele Rosellini, University of Perugia

Matthew Davis, University of California-Davis

Autotetraploid species, such as alfalfa (*Medicago sativa* L.), contain four homologous copies of each chromosome, with the potential of four unique alleles at any given locus. This complexity hinders genomic analyses because identifying alleles across loci that reside on each individual chromosome, that is, identifying the sequence of each individual haplotype has been difficult. However, assembling allele-aware chromosome level genomes with current, third generation sequencing technologies, which produce long, highly accurate reads, is now feasible.

We are using PacBio HiFi (Revio) sequencing to identify the immediate structural changes occurring in the alfalfa genome due to polyploidization. We are assembling the genomes of two meiotic mutant diploid parental genotypes and six of their offspring, three neotetraploids arising from bilateral sexual polyploidization and three diploids. These genotypes have been previously evaluated for differences in their phenotypes, gene transcription levels, and genomic methylation patterns (Rosellini et al. 2016). Using the parental genotypes as references for the progeny, we will be able to identify haplotypes and immediate structural changes that occurred as a result of polyploidization. Mapping differences within and between ploidy levels will provide further insight into the genomic effects of hybridization vs polyploidization. These data can then be compared to established tetraploid alfalfa genotypes to explore long term genome response to polyploidization. This information will be useful to breeders looking to incorporate genes or traits from diploid alfalfa germplasm into breeding programs. Preliminary sequencing results for the parents have shown high accuracy with Q values greater than 30 and reads averaging over 10,000 bp in length, which will be assembled both *de novo* and against current diploid reference genomes to determine best assembly methods as we progress with this project.