Application of Genotyping-By-Sequencing (GBS) in Alfalfa

Yanling Wei, Ananta Acharya, Xuehui Li, E. Charles Brummer
The Samuel Roberts Noble Foundation, Forage Improvement Division, Ardmore, OK 73401
ywei@noble.org; ecbrummer@noble.org

Genotyping of a mapping or breeding population can now be done in a short time at low cost by directly using next-generation sequencing, a methodology termed genotyping-by-sequencing (GBS). The alfalfa (*Medicago sativa* L.) genome is ~800 Mbp and carries a high level of heterozygosity. Alfalfa is an allogamous autotetraploids, and inbred lines are not available due to self-incompatibility and inbreeding depression. In order for GBS to work effectively, the genome complexity needs to be reduced so that only a subset of regions are sequenced. Given the heterogyzous, autotetraploid nature of alfalfa, imputation of missing data will be challenging and determining allele dosage at any given sequence site requires higher read depth than for annual, inbreeding species. Thus, sufficient read depth of the sequenced regions needs to be generated for each individual. The objective of our project is to optimize GBS in alfalfa for application to genome wide association studies (GWAS) and for genome-wide selection (GS).

We created reduced representation libraries using four different enzyme sets that provide variation in the number of sequenced sites. First, we used *ApeK*I (Elshire et al., 2011) and *Pst*I/*Msp*I (Poland et al., 2012) to create libraries of six genotypes, each of which was tagged with eight different barcodes. In the last two libraries (*Sal*I/*Msp*I and *BssH*II/*Msp*I), the same six genotypes were included and tagged with one barcode each, along with individuals from a biparental mapping population, DM3 × DM5. All of the four libraries were sequenced by Illumina HiSeq 2000 at either NCGR or the University of Texas. Preliminary analyses with *Stacks* software (Catchen et al., 2011) to compare the number of targeted sites indicated that the more sites sequenced (less genome reduction), the more reads that did not form stacks (tags) and hence could not serve for SNP genotyping. Conversely, the fewer sites sequenced (more genome reduction), the more reads wasted from deeper-than-needed stacks. We are currently determining the best enzyme combination and the multiplexing possibilities for use in alfalfa mapping and breeding programs.