

Title: The Soybean HapMap

Author: Dave Hyten, Soybean Genomics and Improvement Lab, USDA-ARS

Abstract:

This year the soybean genome sequence was published containing an estimated 46,430 protein-coding genes. While this genome sequence gives a list of protein-coding genes, the next step is to map the variation that occurs within these genes between different soybean accessions and how this variation relates to agronomically important traits. We are in the process of building a next generation genome map called a “HapMap” that charts out the genetic variation in the soybean genome. To build this map we have used next generation sequencing to discover 177,347 single nucleotide polymorphisms (SNPs) spread throughout the genome. From this set of SNPs we selected 52,041 SNPs distributed across the genome to create a 52K SNP chip designed on the Illumina Infinium Assay platform. From the 52,041 SNPs we estimate that 45,000 of the SNPs are polymorphic in diverse germplasm. Current results with the first five chromosomes indicate that 77% of the SNPs have a minor allele frequency >10% in diverse germplasm. In addition, on the same chromosomes we found that the 52K SNP chip is able to assign 78% of the genome sequence to distinct haplotype blocks for the diverse landrace accessions. This 52K chip will be used to genotype the USDA Soybean Germplasm Collection maintained at Urbana, Illinois, which consists of 19,854 soybean and wild soybean accessions. The high coverage the 52K chip allows for tagging soybean haplotype blocks and creation of a high quality HapMap for the entire germplasm collection will be useful for new gene discovery through association analysis. In addition the HapMap will be used to help decipher signatures of selection that have been created due to past selection by breeders for seed yield in soybean.