M-11

Identification of large deletions in soybean lines using a genotyping-by-sequencing (GBS) approach

*Marc-André Lemay**, Department of Phytology, University of Laval, Quebec, Canada
Davoud Torkamaneh, Department of Plant Science, Laval University, Quebec, Canada
Adrian O. Stec, Department of Agronomy and Plant Genetics, University of Minnesota, Minnesota, USA
Robert M. Stupar, University of Minnesota, Minnesota, USA
François Belzile, Pavillon Charles-Eugène Marchand, Quebec, Canada

Genotyping-by-sequencing (GBS) allows efficient identification and genotyping of a large number of molecular markers by sequencing a constant subset of the genome captured through the use of a restriction enzyme. However, technical developments regarding this approach have been almost entirely focused on the discovery and genotyping of single nucleotide polymorphisms (SNPs), while neglecting larger-scale variation that occurs among different cultivars. We present here a new analytical approach that allows the detection of large deletions in soybean plants from standard GBS data. Our approach relies on the identification of genomic regions with significantly lower read coverage than would be expected by chance. We initially tested our pipeline on a set of 96 mutant soybean lines derived from cultivar M92-220 and obtained using fast neutron mutagenesis. Deletions identified using our pipeline were validated for 8 of the 96 lines by comparing the GBS calls to deletions identified by comparative genomic hybridization (CGH). Results from this first stage indicate that our analysis can consistently and reliably identify deletions over 100 kb in size. In a second stage, we identified a set of possible deletions from GBS data generated for 24 Canadian soybean lines and validated these deletions using a dataset of structural variants obtained for these same lines using whole-genome sequencing. Results again showed that deletions over 100 kb in size could be called with a high degree of confidence from GBS data. These results show that GBS data can be used to rapidly call large deletions, in addition to SNPs. Furthermore, our GBS analysis pipeline is a less expensive alternative compared to other existing methods, making it a feasible approach for large variant detection among many individuals in a population.