

## **Genotyping-by-sequencing to identify SNP markers that vary in frequency in response to recurrent selection for freezing tolerance in alfalfa**

Yves Castonguay<sup>1</sup>, Solen Rocher<sup>1</sup>, Martine Jean<sup>2</sup>, David Gagné<sup>1</sup>, and François Belzile<sup>2</sup>

<sup>1</sup> Centre de Recherche et de Développement sur les Sols et les Grandes Cultures, Agriculture et agroalimentaire Canada, Quebec City, QC, Canada

<sup>2</sup> Département de Phytologie and Institut de biologie intégrative et des systèmes, Université Laval, Quebec City, QC, Canada

Methods allowing the simultaneous identification of multiple loci linked to phenotypic variation would help accelerate the improvement of complex traits. Genotyping-by-sequencing (GBS) is a relatively low-cost next-generation sequencing technology for high-throughput genotyping in orphan species with no reference genome. Combining reduced genome complexity and multiplexing with DNA barcoding provides a simple and affordable strategy to investigate variations in allele frequencies between genotypes or populations.

We used GBS to identify alleles under selection pressure in populations of outbreeding alfalfa recurrently selected for superior tolerance to freezing (TF) within the cultivar Apica. Sequencing was performed on *Ape*KI-digested libraries using DNA from 48 genotypes of each of two populations: A-TF0 (initial population) and A-TF5 (obtained after 5 cycles of recurrent selection). Nearly 400 million reads were obtained and analyzed with the UNEAK pipeline of the TASSEL software. After filtering, 4294 high-quality single nucleotide polymorphisms (SNPs) were retained. After genotype calling, allele frequencies were calculated for A-TF0 and A-TF5. Reads supporting each allelic state were also pooled within each population to compare allelic read frequencies.

About 100 SNPs were found to display a large change in allele and/or read frequency ( $\geq 25\%$ ) in response to recurrent selection. Fourteen SNPs with a single perfect match onto the *Medicago truncatula* syntenic genome were selected for validation by 454-FLX titanium sequencing. Fragments of about 600 bp overlapping the SNP of interest were amplified from eight genotypes. Most of the GBS genotype calls (homozygote vs. heterozygote) were confirmed by 454 sequencing. Discrepancies in the allelic ratio determined by GBS and 454 sequencing were frequently associated with lower GBS read counts and would therefore be eliminated by a more stringent filtering threshold. Our results show that GBS analysis of populations is an effective approach for the identification of alleles under selection pressure in recurrently selected populations of outbreeding species.