Novel Targeted Genotyping By Sequencing (tgbs): A Promising Tool For Lentil Genetics Research And Breeding

Ruwani Dissanayake[1,2]
Luke Pembleton[2], Noel O.I. Cogan[2,3], Kevin Smith[1,4] and Sukhjiwan Kaur[2]

[1] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3052, Australia
[2] Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Department of Jobs, Precincts and Regions, 5 Ring Road, Bundoora , Victoria 3083, Australia
[3] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3086, Australia
[4] Agriculture Victoria Research, Department of Jobs, Precincts and Regions, PB 105, Hamilton, Victoria, 3300, Australia

Crop improvement is highly dependent on genetic variation being present and identified from within a breeding pool, often through the use of DNA markers. Currently, SNPs are the dominant form of molecular markers used for genetic and genomic analysis due to their abundancy in genome, and relative ease of identification in a cost-effective manner. Although whole genome sequencing facilitates the identification of all SNPs within a genome, it is an expensive strategy when applied to plant species with large and complex genomes, including pulse crops. Therefore, development of low depth and cost-effective novel Genotyping by Sequencing (GBS) methods is required. For pulse crops, various complexity reduction-based GBS approaches have already been developed and applied. However narrow genetic diversity present within pulse cultivars, limits the number of SNPs detected, and the efficiency of the approach. Therefore to overcome this constraint, targeted GBS (tGBS) methods are needed that build from the genome sequences and known variant positions.

The present study has focused on the development and optimization of a GBS method applicable to lentil, based on target capture where 65,630 custom probes were designed to 47,367 targets that are distributed uniformly across the reference genome of lentil CDC Redberry (version 1.2). The experiment was performed using the NuGEN® Allegro tGBS protocol. The approach was further optimized to enable miniaturization of the reaction volume to reduce the costs further. The protocol was evaluated on 24 samples processed as a full reaction, and then 72 genotypes processed as half and quarter reaction, to a total of 168 samples. The sequence data for each reaction volume has revealed that there were no-significant differences in the number of targets detected (c. 47,000) or missing data percentages (6.82-7.83%), under data filtration on depth (dp) ≤5. Overall, the majority of probes successfully bound and were sequenced in all scales of the reaction. Further evaluation of data on optimal balance between number of mapping reads and missing data percentages has revealed that, c. 25% of targets were undetected (dp≤5), with one million of mapping reads. However, cost-effective bioinformatics tools can be applied to impute missing data rather than increasing the sequencing outputs. The developed tGBS method will facilitate a wide range of opportunities in different fields including, genome-wide association (GWAS) and genomic selection (GS). The optimised tGBS method will be exploited for c. 280 advanced lentil breeding lines targeting the genomic regulation of salt tolerance.