Comparison of Four Pipelines to Develop an Efficient Analysis Pipeline for Large GBS Data

Seo-Hyeon Ham¹, Da-Hye Jeon¹, Dong-Jun Lee¹, Tae-Ho Lee¹*

¹Genomics Division, National Institute of Agricultural Sciences, Jeonju 54874, Republic of Korea

[Introduction]
With cheap NGS technology, resequencing is increasingly used for genotyping, but GBS is still a good choice for large genomes. Genotyping-By-Sequencing (GBS) technology has become a useful technology in agricultural plant research with the announcement of the TASSEL-GBS pipeline in 2014. The genotyping pipelines used in GBS can be divided into two types: the reference-based method and the de novo method determining genotype without a reference genome. Many agriculturally important crops had no reference genome so the de novo method pipelines have been used for the plants. However, there were reports even genotyping result by a reference-based method based on a poor quality draft genome is better than a de novo method. In order to check whether the reports are also correct for a huge genome, we tested two GBS pipelines, GBS-SNP-CROP and Stacks, which support both GBS methods with onion.

[Materials and Methods]
To sequence GBS data, a population of 250 F2 individuals derived from a cross of 'Umjinara' x 'Sinsunhwang' was used. In this analysis, four pipelines were used. A) Stacks pipeline without the reference B) Stacks pipeline with the reference C) GBS-SNP-CROPS pipeline without the reference D) GBS-SNP-CROPS pipeline with the reference.

[Results and Discussion]
As a result, genotyping result by the reference-based method shows better result even that we used a very first draft genome as the reference genome in both pipelines. The cost of genome sequencing is getting to decrease sharply. Thus, the reference-based GBS can be the better choice in a genotyping of a huge genome plant.

[Acknowledgement]
본 성과물은 농촌진흥청 연구사업(사업번호: PJ01355801)의 지원에 의해 이루어진 것임

*Corresponding author: Tel. +82-63-238-4558, E-mail, thlee@korea.kr