#024

Long-read sequencing uncovers genome structural variations associated with protein content in *Brassica napus*

***Harmeet Singh Chawla*[1]**

Mohamed S Youssef[1]
Sean Walkowiak[2]
Robert W Duncan[1]

[1] University of Manitoba, Winnipeg, Canada

[2] Canadian Grain Commission, Winnipeg, Canada

Background:

Brassica napus is an important oilseed crop that is widely cultivated for its high-quality canola oil. During oil extraction, approximately 1.2 million tons of rapeseed meal are produced globally each year as a by-product. However, the breeding of rapeseed varieties with enhanced protein content has not received much attention, primarily due to the lack of financial incentives. As plant-based protein gains popularity and the demand for alternative protein sources increases, understanding the genetic factors that determine protein content in *B. napus* is essential.

Objective:

The main objective of this study was to identify genetic polymorphisms associated with protein content in *Brassica napus*. Our study aimed to pinpoint the crucial candidate genes, as well as the single nucleotide polymorphisms (SNPs) and structural variations (SVs) that play a major role in regulating protein content in this crop species. By employing a comprehensive genetic and genomic approach, our investigation aimed to provide insight into the underlying genetic mechanisms that control protein content in *B. napus*.

Methods:

To identify quantitative trait loci (QTL) associated with protein content in *B. napus*, we developed two bi-parental doubled haploid (DH) populations and genotyped them using the Brassica 90K Illumina Infinium array. We then phenotyped the populations under three different environmental conditions to obtain a comprehensive understanding of the genetic determinants of protein content. To further elucidate the molecular mechanisms underlying this trait in *B. napus*, we generated long-read PacBio HiFi data for the parental genotypes of the two bi-parental populations.

Results:

We identified 24 QTL associated with protein content in two populations. Among these, a QTL located on chromosome C09 was consistently found in all three environments in both populations. By anchoring the genetic markers flanking this QTL region on to the Westar genome assembly (Song et al. 2020), we pinpointed a 7 Mb region on chromosome C09 containing 1079 protein coding transcripts. Through integration of QTL data and long-read PacBio HiFi data for the parental lines of the two bi-parental populations, we discovered a potential candidate gene that may play a significant role in protein synthesis in *B. napus*.

Conclusions:

Our study identified multiple QTL associated with protein content in *B. napus* and uncovered a promising candidate gene on chromosome C09 that may be involved in protein synthesis. Our findings highlight the complexity of the genetic regulation of protein content in *B. napus* and provide valuable insights for future studies aimed at expanding end use of this critical crop. The integration of genetic data and advanced sequencing technologies is a powerful approach to unravelling the genetic basis of complex traits in crops and can pave the way for more targeted and efficient crop breeding programs.

Reference:

Song J-M, Guan Z, Hu J, *et al.* (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants* 6:34–45. https://doi.org/10.1038/s41477-019-0577-7