#045

The construction of tetraploid *B. napus* pan-genome based on genetic variations of 2,882 germplasm accessions

***Chaobo Tong*[1]**

Ming Hu[1]
Xiaobo Cui[1]
Yuanyuan Zhang[1]
Lijiang Liu[1]
Shengyi Liu[1]

[1] Oil Crops Research Institute, Chinese Academy of Agricultural Science, Wuhan, China

Background:

The construction of a pan-genome for B. napus is crucial for revealing the genetic diversity and identifying the genetic variations (SNP and SV) associated with phenotype innovation. However, comprehensive information on the B. napus pan-genome is lacking, hindering gene discovery and study in genomic evolution of polyploidy.

Objective:

Our aim is to construct a B. napus pan-genome by integrating 26 de-novo assembled reference genomes based on Pacbio or Nanopore technologies and the re-sequencing data for 2,884 diverse accessions, and rich annotation on mutation, gene expression and epigenetic modification.

Methods:

Firstly, we individually assembled 4 representative B. napus genomes by integrating PacBio, BioNano and Hi-C data. Secondly, a total of 26 de-novo assembled genomes, high-depth re-sequencing data of 381 representative accessions and 2,501 low-depth re-sequencing accessions were used to construct B. napus pan-genome. Finally, the Iso-seq for several tissues and RNA-seq for hundreds of accessions were conducted for accurately annotating the B. napus pan-genome.

Results:

We integrated 26 reference genomes and re-sequencing data from 2,884 diverse B. napus accessions, resulting in a comprehensive B. napus pan-genome that is 1.99 Gb in size and contains 159,099 high-confidence genes. A total of 8,736,523 high-quality SVs and 112,886 high-quality presence/absence variants (PAVs) were identified as well as 41,018 core genes, 16,839 softcore genes, 92,180 dispensable genes and 18,396 private genes. Compared to the dispensable genes, core genes may have greater selective adaptability due to their higher expression and maintain sequence conservation under purifying selection pressure. In addition, SVs were found to be responsible for the loss of gene sequences to create dispensable genes, which have a very large degree of population expression variability. The gene dosage effect was found to exist in many accessions as the total gene expression was compensated by other homologous copies of the lost gene. Finally, we developed a comprehensive database called BnaOmics (https://bnaomics.ocri-genomics.net/) that includes genome sequences, gene annotations, gene expression profiles in multiple tissues and populations, synteny blocks, epigenomes, and common bioinformatics tools for analysis and visualization.

Conclusions:

This study constructed a high-quality aligned pan-genome of *B. napus* and highlighted the polyploidy evolution and genomic plasticity, which provided valuable resources and a foundation for guiding genetic breeding of B. napus. The constructed BnaOmics database enables researchers to search for and visualize results in a pan-genome context and provides a good example for research into other species.

References:

Yao, S. et al., (2020). A global survey of the transcriptome of allopolyploid *Brassica napus* based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. The Plant Journal 103(2): 843-857.

Chen, X. et al., (2020). A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. Plant Biotechnology Journal 19(3): 615-630.