#106

Super pan-genome and RNA-Seq SNPs - GWAS to genetically characterise resistance to sclerotinia stem rot in *B. juncea* introgression lines

***Chhaya Atri*[1]**

Javed Akhatar[1]
Kusum Rana[1]
P.S, Sandhu[1]
Martin Barbetti[2]
Surinder Banga[1]

[1] Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

[2] School of Agriculture and Environment and UWA Institute of Agriculture, University of Western Australia, Crawley, Australia

Background:

Sclerotinia stem rot, caused by Sclerotinia sclerotiorum, is a major disease of Indian mustard (*Brassica juncea* L.), with no heritable or transferable source of resistance documented in the primary gene pool of crop. Our past efforts had allowed introgression of variation for resistance from a wild *Brassicaceae* species, *B. fruticulosa*. Introgressed resistance was later shown to be quantitatively inherited. However, the genetic variations in these studies were identified against a single accession *B. juncea* reference genome. Accordingly, DNA sequences that are absent from the reference genome, especially those belonging to the genomic information introgressed from the wild donor species, are likely to have remained unmapped.

Objective:

Investigations were conducted to create a super pangenome of *B. juncea*, carrying wild species-specific genomic content available in *B. juncea-B. fruticulosa* introgression lines (ILs) by using the developed genomic resource to conduct RNA-Seq SNP-GWAS and investigating the association of ePAV with trait variation.

Methods:

Eleven re-sequenced *B. juncea-B. fruticulosa* introgression lines (IL) were initially used to construct a super pan-genome. De-novo assemblies were first developed for each IL using MEGAHIT and MaSuRCA tools, and merged using CD-HIT. Merged contigs were anchored to the reference genome using RagTag. Repeat Masker was used to detect repeats and low complexity DNA sequences, followed by a BUSCO analysis to assess the completeness of each accession. Pseudo-chromosomes were constructed with local-blastn. RNAseq reads obtained from 104 ILs were then mapped individually on *B. juncea* super pangenome using HI-SAT2. SNPs were identified based on the vcf files generated by bcftools-mpileup and used for GWAS as implemented in GAPIT3.

Results:

Analysis of genome sequencing data allowed us to identify 9365 contigs (15.1 Mb) which could not be aligned with the *B. juncea* genome. These contigs were merged for all 11 accessions and blasted against the in-house *B. fruticulosa* de-novo genome assembly with a percent identity >95% and an e-value 1e-3. Blast hits with alignment length 100 bp were filtered out and overlapping hits merged by bedtools (--merge) tools and further consolidated into scaffolds and ordered on the available *B. fruticulosa* assembly. This allowed addition of 15.1 MB of genomic information (about 2908 genes) into the *B. juncea* genome [*B. rapa* Z1 (AA-genome) + *B. nigra* NI100 (BB-genome)] to create a super pan-genome reference. RNAseq reads obtained from 104 ILs were then mapped to this pangenome for SNP calling. High quality SNPs (50730) remaining after filtration were used for GWAS analyses as implemented in GAPIT3. This analysis enabled the identification of more than 15 disease related genes for association with sclerotinia rot resistance in these introgression lines. These included plant ribosomal proteins, disease resistance gene analogues (RGAs), including nucleotide-binding site-leucine-rich repeat (NLR), receptor-like kinase (RLK), and oxidoreductase protein family proteins.

Conclusions:

A preliminary superpan genome was developed by incorporating sequencing reads from 11 *B. juncea-B. fruticulosa* introgression lines. Mapping of RNAseq reads on this pan-genome resource and subsequent GWAS allowed for the identification of the genes associated with resistance to Sclerotinia stem rot. An extended pan-genome with the addition of genomic information from 72 ILs will be presented, along with ePAV and gPAV and their association with trait variation.