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

Chhaya Atri[1], Javed Akhatar[1], Kusum Rana[1], P.S. Sandhu[1], Martin J. Barbetti[2] , S.S. 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

SCLEROTINIA STEM ROT (SSR)

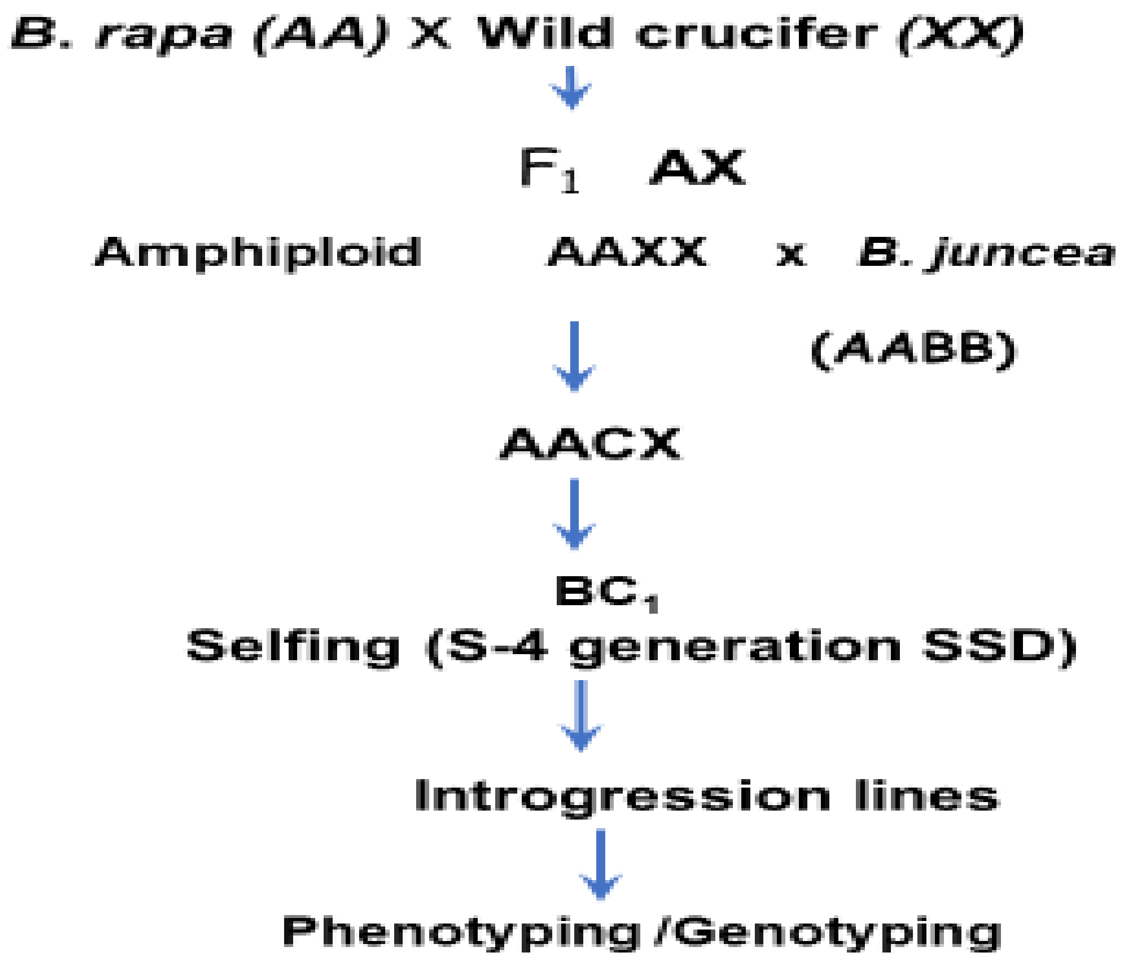
Sclerotinia stem rot (*Sclerotinia sclerotiorum*) causes extensive yield losses to the mustard crops in India

While available cultural and chemical controls can reduce the severity of the disease, management practices do not provide effective and reliable control

Host resistance is not available in the primary gene pool of oilseed brassicas

Introgressed variation for SSR resistance into cultivated *B. juncea* from related wild *Brassicaceae* species, *B. fruticulosa*

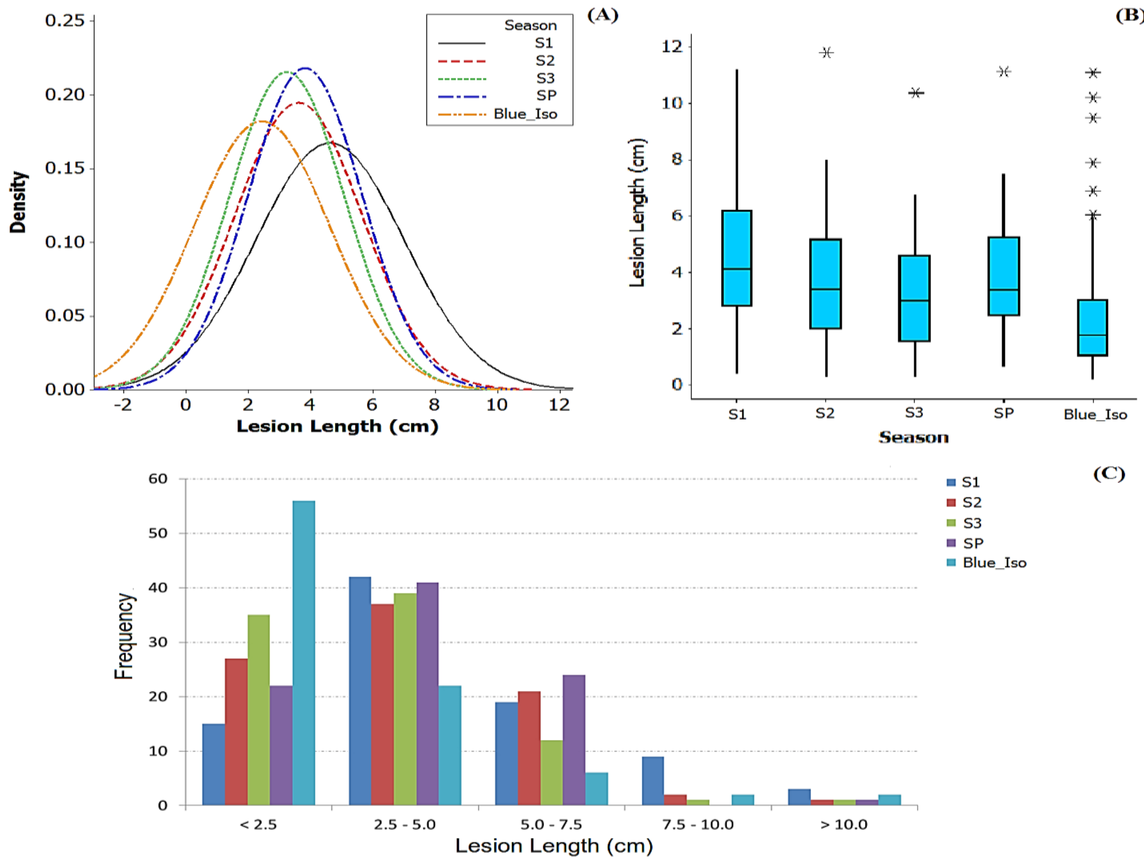
BREEDING STRATEGY TO INTROGRESS SCLEROTINIA RESISTANCE IN *BRASSICA JUNCEA* FROM *BRASSICA FRUTICULOSA*

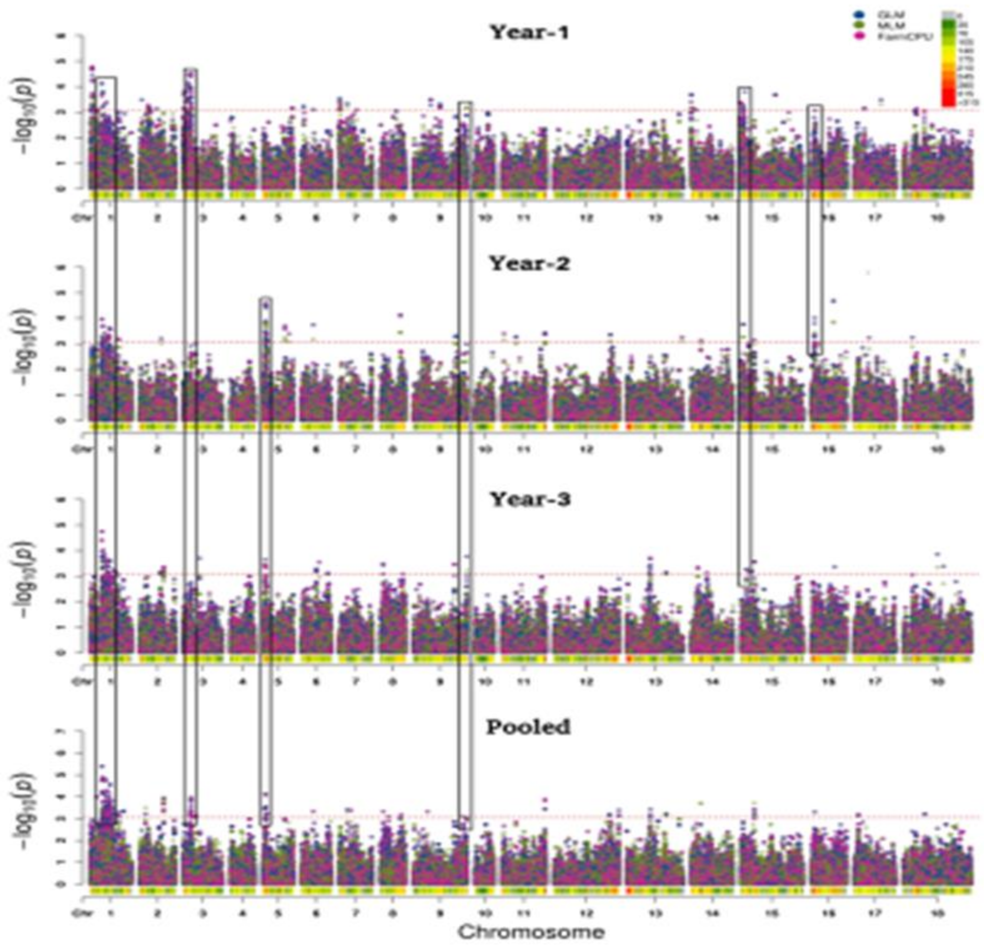


SCLEROTINIA STEM ROT RESISTANCE IN *B. JUNCEA-B. FRUTICULOSA* INTROGRESSION LINES

Stem inoculation method of Buchwaldt et al. (2005) and Li et al. (2006) with minor modifications

PHENOTYPIC DISTRIBUTION OF LESION LENGTH (CM) AND GBS-GWAS FOR WHITE ISOLATE OVER THREE CROP SEASONS





Phenotypic distributions and frequency histogram of stem lesion length in *Brassica juncea* – *B. fruticulosa* intro in seasons I, II and III and pooled across three seasons

Scientific Reports | *(2019) 9:17089* | https://doi.org/10.1038/s41598-019-53444-3

CREATING *B. JUNCEA-B. FRUTICULOSA* INTROGRESSION SPECIFIC SUPER-PANGENOME BY INTEGRATING VARIATION: RATIONALE

A major problem of using single accession reference genomes for mapping sequencing reads from introgression lines is that a large chunk of sequences are left unmapped resulting in missing the genomic regions that are not present in the reference genome

Traditional mapping approaches based on a single accession reference genomes also miss out on structural variations (SVs), which are pretty common in introgression lines. These include presence/absence variations (PAVs), copy number variations (CNVs), and many others like inversions, transversions, and inter/intrachromosomal translocations

Creating super-pangenome using several accessions of two different species for a given genus can be a very expensive and computational intensive

We therefore used a novel approach of developing super-pangenome using sequencing reads from stable *B. juncea-B. fruticulosa* introgression lines. *B. fruticulosa*, is a known donor for resistance to key biotic (aphids, sclerotinia stem rot) and abiotic (salinity) stresses

Genome sequence and resequencing data from 72 *B. juncea-B. fruticulosa* introgression lines, which were carefully selected to represent genetic and phenotypic diversity present in a large number of introgression lines, was used to produce a quality super-pan genome and allow detection of structural variations

CREATING *B. JUNCEA-B. FRUTICULOSA* INTROGRESSION SPECIFIC SUPER-PANGENOME BY INTEGRATING VARIATION - THE PROCESS

Sequencing reads (140Gb) from 11 ILs, were first assembled *de novo* separately, corrected and patched using *B. juncea* reference genome and merged after removing duplicate scaffolds

The left-over scaffolds were blasted on *B. fruticulosa* draft genome, selecting the best hits with Identity >95% and length >100bp. This allowed addition of 15.1Mb sequence data from wild donor

Next, we took sequencing reads 75 ILs [~800Gb]. These were first mapped to the pangenome 1 and then unmapped reads were assembled using *de novo* assemblers

The assembled contigs/scaffolds are then anchored to the developed pangenome 1, using Ragtag software

The left-over unanchored scaffolds were collectively anchored to in house draft genome assembly of *B. fruticulosa,* based on the best hits (Identity 100 %, length > 500bp) and then added to existing *B. juncea* genome. This step allowed addition of another 33.36 Mb of wild species genomic information to pangenome 1. Creating a super pangenome AABB+FF

PANGENOME CONSTRUCTION BASED ON *BRASSICA JUNCEA – B. FRUTICULOSA* ILS-I

Sequencing reads from 11 samples [~140Gb]

↓

***De-novo* assembly for each sample using MEGAHIT & MaSuRCA assemblers**

↓

Merge *De-novo* assemblies of each sample (CD-HIT)

↓

Correct, Scaffolding & Patch assemblies with reference genome (He et al.2021) separately for each sample (RagTag)

↓

Merge all patched scaffolds (n= 1,961,557, size =2,654Mb, N50=3027bp)
Remove duplicate scaffolds (n= 1,069,034, size = 1,706Mb, N50=4601bp)

↓

Blast final scaffolds with an in house *B. fruticulosa*(FF) draft assembly
Best hits (Identity > 95%, length > 100bp)

↓

Extract sequences from wild species based on best hits
Pan-sequence1 (n=53,482, size = 15.1Mb)

↓

Anchoring and ordering sequences with reference genome
Pangenome1 (AABBFF): *Brassica juncea* reference + Pan-sequences1

↓

Genome annotation (Augustus: genes 62,143)

In-House Process

PANGENOME CONSTRUCTION BASED ON *B. JUNCEA – B. FRUTICULOSA* ILS-II

Genome sequences from 75 samples [~800Gb]

↓

Genome mapping (bowtie2) on Pangenome1 (AABBFF)

↓

Save un-mapped FASTA reads and merge into 10 sets (~ 517 Gb)

↓

De-novo Analysis (MEGAHIT)

↓

Correct, Scaffolding & Patch assemblies with reference genome of each set (RagTag)

Ref: Pangenome1 (AABBFF)

↓

Merge all sets patched scaffolds (n= 11,082,359, size =7,002Mb, N50=727bp)

Remove duplicate scaffolds (n= 3,395,090, size = 2,869Mb, N50=996bp)

↓

Blast final scaffolds with *B.fruticulosa* draft genome

Best hits (Identity 100 %, length > 500bp)

↓

Extract sequences from wild species based on best hits

Pan-sequence2 (n=25,150, size = 33.36Mb)

↓

Anchoring and ordering sequences with reference genome

Fruti-Pangenome (AABBFF): *Brassica juncea* reference + Pan-sequence2

↓

Genome annotation (Augustus genes: 67,181)

In-House Process

GENOME WIDE ASSOCIATION STUDIES USING SVs and RNAseq

Sv-GWAS

SVs are longer (>100 bp) and likely to have greater influence on gene expression and protein function as compared with SNPs and small indels (insertions and deletions)

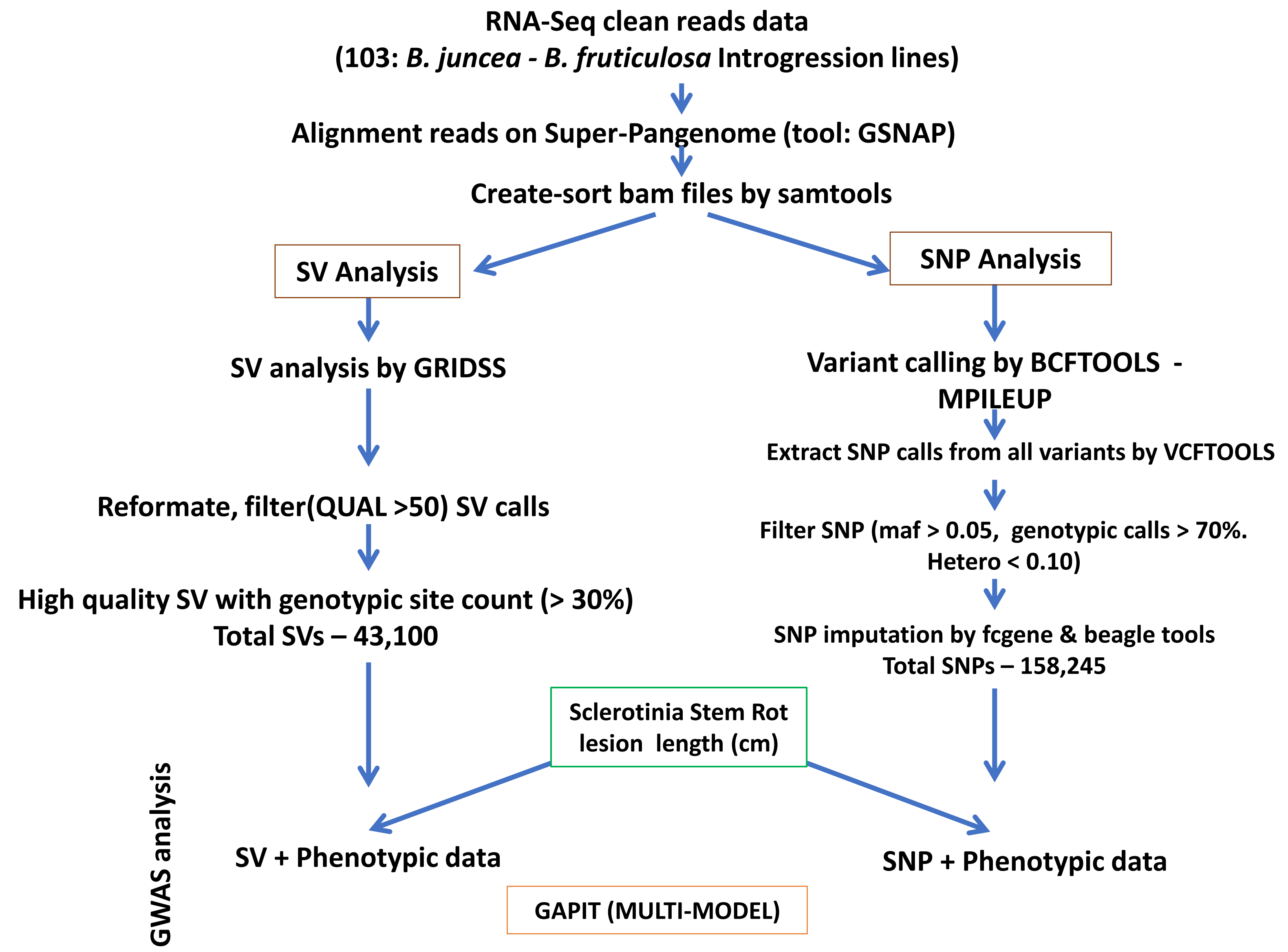
Moreover, many of the disease resistance-associated genes have been reported to evolve quickly through duplications, chromosomal rearrangements, and unequal crossing. We, therefore, also used data from SVs in ILs for conducting Sv GWAS

RNA-SEQ SNPs-GWAS

RNA-seq SNPs based GWAS was conducted following transcriptome sequencing of 111 introgression lines(Ils) and modern cultivars, using newly developed pan genome as a reference

The data obtained from thorough phenotyping ILs for defensive responses and the SNP diversity from adult plant transcripts of the diversity set were jointly analyzed to identify genomic regions associated with the expressed phenotypic variation

GWAS ANALYSIS USING SV SNP AND RNA-SEQ DATA OF *B. JUNCEA - B. FRUTICULOSA* ILS



CONCLUSIONS

Development of alien introgression based super pan genome is a major step forward to optimally identify introgressed genomic regions, identify structural variation and provide scientific framework to optimize experimental designs to reduce linkage drag, especially using CRISPER based segment deletions

Our results allowed validation of our approach of using a species-specific pan genome that not only helped to validate previously reported MTAs, but also highlighted the role of previously uncharacterized chromosomal regions, including on added pan genome

SV and RNA seq based GWAS allowed prediction of 38 genes for resistance to sclerotinia pathogen

Comparison of transcriptome profiles of inoculated and uninoculated introgression lines have indicated substantive gene expression variations for 15-19 of 38 predicted genes

***RCA*, *CHI* , *PP2-A1* , *CHI*, *AT1G72180*, *LECRK-4*,*CYP98A3* and *ACX1* appeared very promising candidates. All of these are known defensive genes**

Roles of jasmonate pathway, phloem proteins, defense related Pep peptides, plant receptor-like kinases, phenylpropanoid pathway and lignin biosynthesis in inducing defensive responses was emphasized

ACKNOWLEDGEMENTS

Research presented was carried out with the financial support of :

Centre of Excellence and Innovation in Biotechnology "Germplasm enhancement for crop architecture and defensive traits in *Brassica*" Department of Biotechnology, Government of India

National Professor Project "Broadening the genetic base of Indian mustard (*Brassica juncea*) through alien introgressions and germplasm enhancement. Indian Council of Agricultural Research

Project associated with Raja Ramanna Fellowship, Department of Atomic Energy, Government of India

Punjab Agricultural University, Ludhiana, India