


Finding function in the canola genome

Reece Tollenaere


Canola Pathology Meeting
25 March, 2013



Resources

Reference Genomes

- A and C genomes
- A – *B. rapa*
- C – *B. oleracea*

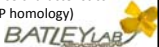


Rlm4 QTL Mapping

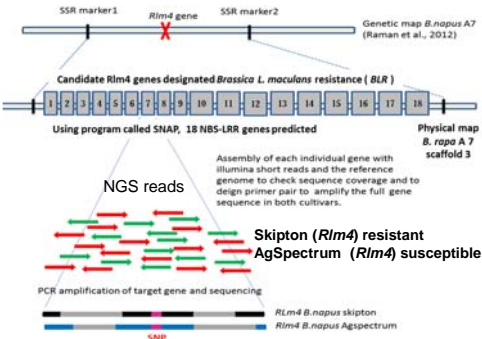
- 12 SSR markers underlying the *Rlm4* QTL from a Skipton x Ag Spectrum mapping population were BLASTed against the reference genome – Scaffold 3 of A7 (Raman *et al.*, 2012)

SNAP Resistance Gene prediction

- The program SNAP used to predict and annotate gene-coding sequences in *Rlm4* QTL
- 18 genes with disease resistance characteristics (BLASTP homology)



Sequence based identification of candidate genes – BLR genes



SSR marker1 *Rlm4* gene SSR marker2 Genetic map *B. napus* A7 (Raman *et al.*, 2012)

Candidate *Rlm4* genes designated *Brassicica L. maculans* resistance (*BLR*)

Using program called SNAP, 18 NBS-LRR genes predicted Physical map *B. rapa* A7 scaffold 3


NGS reads

Assembly of each individual gene with illumina short reads and the reference genome to check sequence coverage and to design primer pair to amplify the full gene sequence in both cultivars.

Skipton (*Rlm4*) resistant
AgSpectrum (*Rlm4*) susceptible

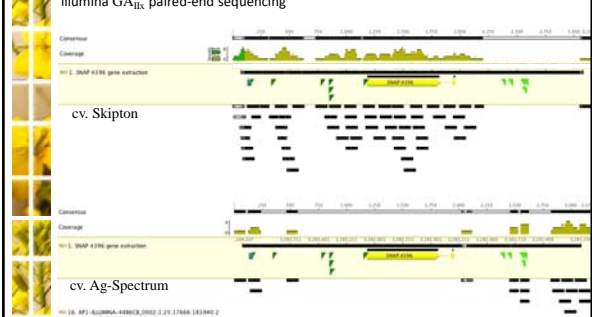
PCR amplification of target gene and sequencing

Rlm4 *B. napus* skipton
Rlm4 *B. napus* Agspectrum




Short sequence read assembly

illumina GA_{IIx} paired-end sequencing



cv. Skipton

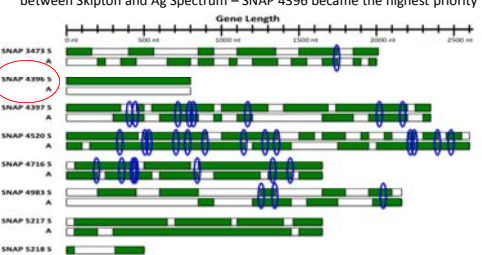

cv. Ag-Spectrum



Methods

Sequencing of Skipton and Ag Spectrum

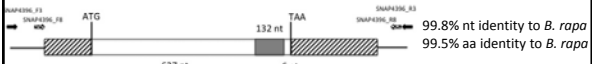
- The genomes of Skipton and Ag Spectrum were partially sequenced using new technologies
- Partial sequence was identified for the 18 candidate genes and compared between Skipton and Ag Spectrum – SNAP 4396 became the highest priority

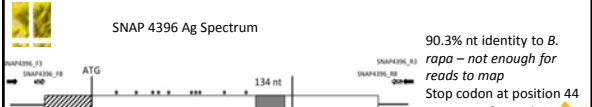
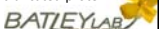
Candidate gene sequence analyses

- Closer inspection of the candidate genes on a nucleotide level
- Confirmed the presence of many mutations in the susceptible Ag-Spectrum for this gene and the 402 nt insertion visible in the gel
- Blasts of the primers to ensure the specificity to the A7 chromosome due to the duplicated nature of Brassicaceae

SNAP 4396 Skipton




SNAP 4396 Ag Spectrum

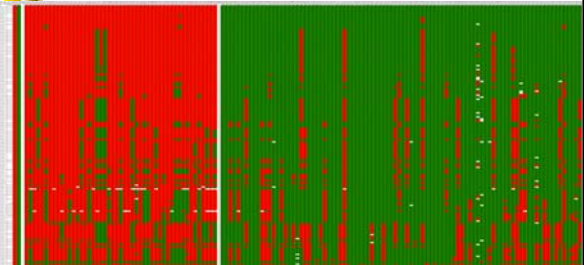
SNP genotyping

Infinium 6K and 60K assays


- 6K Assay was designed predominantly on the Skipton / Ag-Spectrum read analysis
- 60K has 10x greater SNP density across the genome
- SNP design may be less specific – more monomorphs



Infinium 6K SNP genotyping

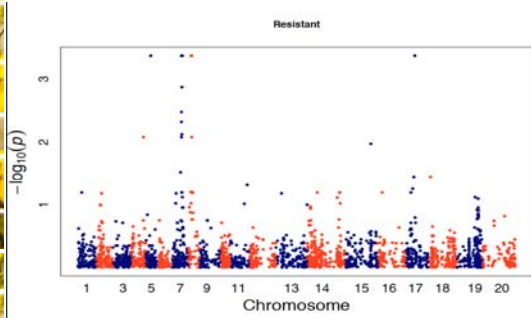


- Attained full genotype and phenotype data for 142 Doubled Haploid lines derived from a Skipton x Ag-Spectrum cross
- Resulted in 2560 informative SNPs across the A and C genome




Manhattan Plot - Genomewise

Resistant



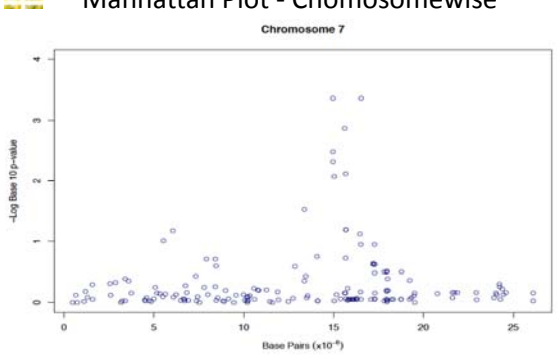
Chromosome

Chromosome A7 and C6 homologs
Chromosomewise plots performed also
SNPs mis-mapped due to duplicate BLAST hit




Manhattan Plot - Chromosomewise

Chromosome 7

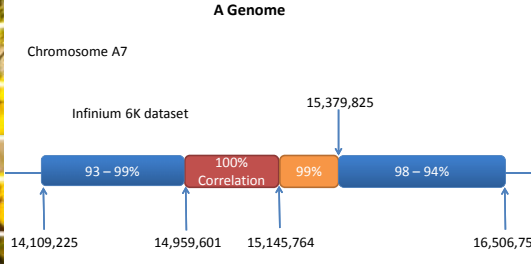


Base Pairs (x10⁶)




A Genome

Chromosome A7



- 100% correlation attained for 6 markers across a 186,164 bp region
- Gradual tapering of correlation for ~ 1 Mbp up and downstream before a sharp correlation decline
- Integration of Infinium 6K assay with the 60K assay to increase resolution




Genotyping by Sequencing

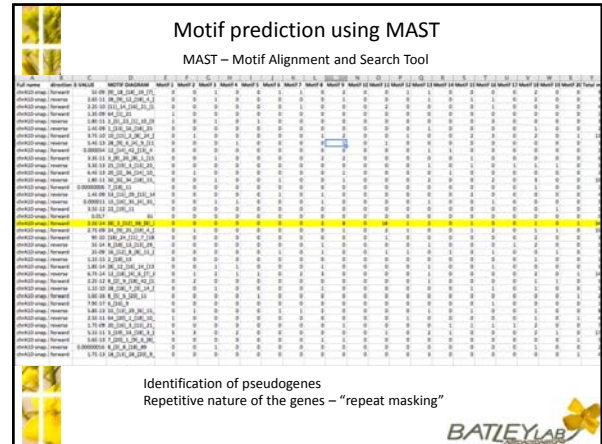
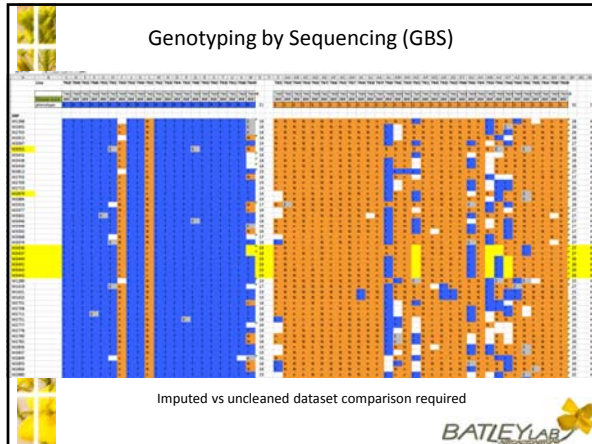
Tapidor and Ningyou population – identification of the *Rlm2* gene

Illumina Hi-Seq 2500 sequencer Melbourne

Currently have approx 59 lines of a TN mapping population sequenced

Combination of paired-end reads and single reads – Paired end reads assist in genome assembly, INDEL identification relative to the reference etc



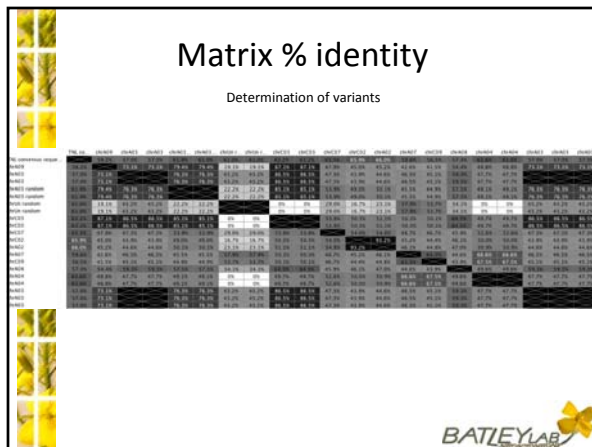
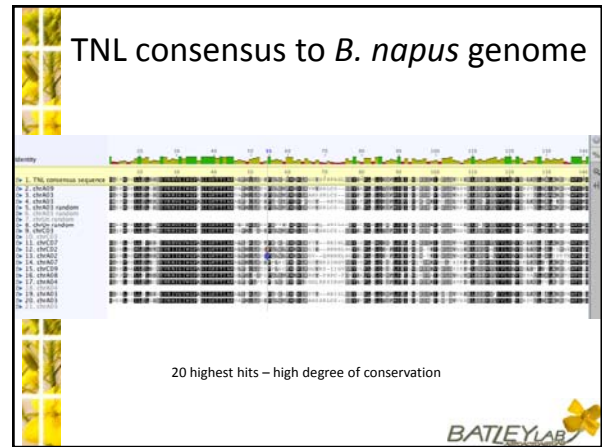


Domain identification

Two approaches:

1. Consensus sequence to source sequence BLAST
 - ↳ TNL and CNL consensus sequences (Ameline-Torregrosa et al., 2008)
1. Source sequence to domain database
 - ↳ Database is in construction and currently have approximately 4,500 NBS-LRR sequences from different genomic sources
 - ↳ Domain database can be utilised to form species specific consensus sequences or pathogen specific NBS-LRR proteins
 - ↳ Domain database will evolve
 - ↳ Cytoplasmic, extracytoplasmic membrane anchored

BATLEYLAB



Summary

How the data is useful for the research

- Further narrowing of the QTLs related to Rlm genes
- Genome-wide marker assisted association
- Breeding line introgression
- Directly transferrable methods and processes to any trait of interest

BATLEYLAB

Acknowledgements

All members of The Batley Lab, past and present



The Edwards Bioinformatics Group

Dr Harsh Raman (NSW Department of Industry and Innovation)

Xiaowu Wang (Chinese Academy of Agricultural Sciences, Beijing, China)

Bayer Crop Sciences, Belgium

CILR lab members

Resources for gene identification


Reference Genomes
 • A and C genome sequences derived from the related *B. rapa* and *B. oleracea*

+

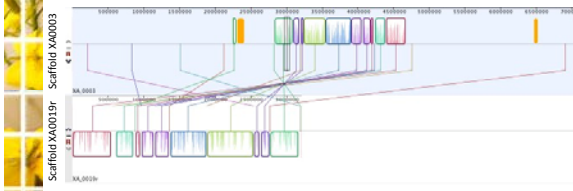
Genetic markers for the *Rlm4* gene
 • (known differences in genetic sequence between the resistant Skipton and susceptible Ag Spectrum)

→ Enabled the identification of a particular region on the A7 chromosome as containing the *Rlm4* gene

Gene prediction




Chromosome 7 duplication

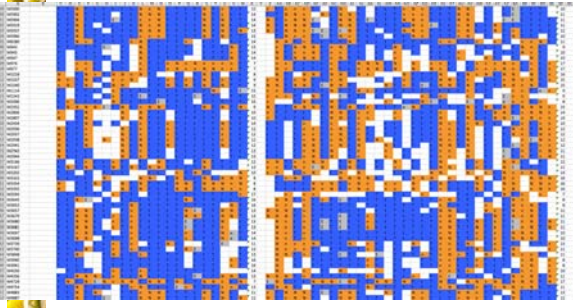


MaVe alignment enables the identification of the collinear blocks present in the duplicated region at the end of chromosome 7


- Consensus sequence differentiation
- *Rlm4* gene relatives – *Rlm7*



Genotyping by Sequencing (GBS)



No Association



SNAP 4396 gene sequence assembly

