**Paper to do**

**BLAST search for names:**

Bcin01g10130

Bcin15g04110

(peptidase dimerization, Bcin01g10130; pectinesterase, Bcin14g00870; protein kinase, Bcin15g04110 ).

**Check this estimate:**

Higher polymorphism rates are reported for the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* (12.3 SNP/kb) (Upadhyaya, Garnica et al. 2014).

**pectinesterase gene (BcT4\_6001, Bcin14g00870)**

To visualize the SNP effects across a single gene and look for evidence of multiple haplotypes, we plotted the effect sizes for all SNPs in this gene and investigated the linkage disequilibrium amongst these SNPs (Figure 6). This showed that the effect of SNPs across this gene vary in effect direction depending on tomato host genotype (Figure 6a), and that there appear to be two different haplotype blocks contributing to the association of this gene to the virulence phenotype (Figure 6b). One block is associated with SNPs in the 5’ untranslated region in SNPs 5-11 and the second block is SNPs that span the entirety of the gene in SNPs 13-26. Interestingly, there are only two SNPs in the open reading frame of the associated gene and this SNP is a silent site polymorphism (Figure 6). This suggests that the major variation surrounding this locus is controlling the regulatory motifs for this pectinesterase.

On figure:

32 SNPs for haplotype blocks

34 SNPs for effect estimates

DRAW LINES ACCORDINGLY:

R script: 14\_Fig8b\_LDplot.R

Select focal gene from

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 start\_codon 826343 826345 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 exon 826235 826345 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 CDS 826235 826345 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 exon 825306 826178 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 CDS 825309 826178 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Chromosome2.2 Botcin\_T4\_vankan\_IMPORTGTF\_3 stop\_codon 825306 825308 0 - 0 gene\_id "BcT4\_6001""; transcript\_id ""BcT4\_6001T1"";"

Lowest: 825306 (stop codon/ exon)

Highest: 8292710 (start codon)

Segments:

Stop 825306 825308

Exon1 825306 826178

CDS 825309 826178

Exon/ CDS 2 826235 826345

Start 826343 826345

For Fig 8a: range is 823323 to 828305

on linux:

bgzip Documents/GitRepos/BcSolGWAS/data/genome/big\_set\_v97iso\_SNPs\_filtered\_qual30\_dp6\_maf20\_recode.vcf

install vcftools

vcftools --gzvcf Documents/GitRepos/BcSolGWAS/data/genome/big\_set\_v97iso\_SNPs\_filtered\_qual30\_dp6\_maf20.recode.vcf.gz --chr Chromosome2 --recode --out Documents/GitRepos/BcSolGWAS/data/genome/chr2\_analysis/chr2\_analysis

this keeps only Chr2 sites

I’ll also make a shorter fragment with only 4kb around the gene of interest:

vcftools --gzvcf Documents/GitRepos/BcSolGWAS/data/genome/big\_set\_v97iso\_SNPs\_filtered\_qual30\_dp6\_maf20.recode.vcf.gz --chr Chromosome2 --from-bp 823323 --to-bp 828305 --recode --out Documents/GitRepos/BcSolGWAS/data/genome/chr2\_analysis/chr2\_analysis\_seg