Development of Genomic Tools for Almond Rootstock Development

Project No.: **12-HORT16-Aradhya/Ledbetter**

Project Leader:
Malli Aradhya
National Clonal Germplasm Repository (NCGR)
UC Davis - USDA-ARS
One Shields Avenue
Davis, CA 95616
530.752.9392
malli.aradhya@ars.usda.gov

Craig A Ledbetter
Crop Diseases, Pests, and Genetics
USDA-ARS
9611 S. Riverbend Ave.
Parlier, CA, 93648
559.596.2817
craig.ledetter@ars.usda.gov

Project Cooperators and Personnel:
Gregory Browne, Crops Pathology and Genetics, USDA-ARS
Daniel Kluepfel, Crops Pathology and Genetics, USDA-ARS
Michael McKenry, UC Riverside & Kearney Ag Center (*Emeritus*)
John Preece, NCGR, USDA-ARS, Davis
Dianne Velasco, doctoral candidate, UC Davis Plant Sciences

Objectives:

1. Develop a set of molecular markers linked to genes conferring resistance to soil borne diseases to improve selection efficiency for disease resistance in rootstock breeding programs.
    a. Identify markers and genomic regions controlling resistance/tolerance to soil borne diseases through association analysis.
    b. Profile root-specific transcripts to identify genes that are differentially expressed in resistant rootstocks and utilize in candidate gene based Single Nucleotide Polymorphisms (SNP) discovery.

2. Molecular characterization of a genetically diverse collection of commercial and newly produced interspecific hybrids.
    a. Genotype commercial and experimental rootstocks in various trials and utilize the disease data for association analysis to identify and validate molecular markers.
    b. Produce and test diverse interspecific hybrids involving wild *Prunus* spp. that are potential donors of resistance to soil borne pests and pathogens.

3. Develop efficient marker assisted selection strategies for rapid development of resistant rootstocks.

a. Develop and validate marker assisted selection strategies using the knowledge gained in objective 1 and 2.
b. Implement marker assisted selection to develop soil borne disease resistant rootstocks.

Interpretive Summary:

The key soil borne diseases of almond, Phytophthora (PHY) rots/Agrobacterium induced crown gall (CG)/lesion and rootknot nematodes (NEM) significantly reduce orchard/nursery productivity and orchard longevity. The widely used fumigant in almond nurseries and orchards, methyl bromide is rapidly being phased out. Finding sustainable and environmentally sound alternatives is a top priority for the Almond Board of California. Rootstocks with resistance to soil borne pathogens will significantly reduce dependence on soil fumigation for optimal productivity. Development and deployment of genetic markers will facilitate rapid selection of resistant rootstock genotypes at juvenile stages without pathogen challenge and grow out. These advantages cannot be overstated when considering selection of commercially viable woody perennial tree rootstocks.

Availability of diverse germplasm, high throughput marker systems for high density genotyping, efficient and foolproof disease testing schemes are the key to the success of this project. Single Nucleotide Polymorphisms (SNPs) are the marker of choice for high density genotyping. This research during the past three years has focused on: (1) screening and identifying reliable sources of durable combined resistance to soil borne diseases; (2) hybridize potential donor species with peach and almond genotypes to produce novel rootstock genotypes; (3) SNP developing, genotyping and disease testing; and (4) identify markers associated with soil borne diseases to develop and validate effective marker assisted selection strategies.

The following are the highlights of the research:

- Identified marker(s) associated with an important soil borne disease, *Agrobacterium* induced crown gall (CG).

- Developed and assembled a wide range of novel interspecific *Prunus* hybrids involving species that are potential donors of disease resistance hitherto not produced.

- A first round of disease testing showed a range of variation in response to soil borne diseases indicating the potential to develop resistant rootstocks.

The approaches taken and the results obtained so far are encouraging and on a long-term basis this research should be able to find significant solutions to soil borne diseases, the major limiting factor for almond production in California. There is no quick fix or substitution to hard work.

Materials and Methods:

<u>Newly developed experimental rootstocks</u>

In 2013 clones of 13 unique rootstock genotypes were distributed to Kluepfel and Brown for crown gall and *Phytophthora* screening, respectively. A final delivery of interspecific hybrids developed in 2012 that underwent embryo culture and multiplication at California Seed and Plant were received in May 2014. These hybrids have been repotted and are currently awaiting distribution (**Figure 1**) to cooperating plant pathologists for disease evaluation. Additional experimental hybrids are undergoing clonal propagation at NCGR for a second round pathology screening.

Two selected rootstock genotypes that showed resistance in first round of disease screening were budded with 'Nonpareil' scion and planted at Wolfskill Experimental Orchards (WEO) for further observations on graft compatibility. Additionally, an area within the almond germplasm collection known to have *Armillaria* infestation has been planted with experimental hybrids for observation.

Commercial and others rootstocks undergoing testing in various trials

Majority of these rootstocks with disease information come from the Greg Browne's replant trial in Parlier, Fowler nursery rootstocks in a Klupfel CG testing trial and a few other sources such Foundation Plant Services. A total of ~190 genotypes including the experimental rootstocks produced in this project and some of the wild *Prunus* spp. used as parents was used in SNP genotyping.

Genomics

Genotyping by sequencing: Leaf tissue samples were collected from the above mentioned 190 rootstock genotypes and DNA isolated with the CTAB extraction buffer (Doyle and Doyle). DNA quality evaluation was performed using agarose gel electrophoresis and *Eco*RI endonuclease digestion of approximately 10% of the 190 samples. DNA samples were quantified with PicoGreen dye and 3 uL of DNA at 50 ng/uL sent to the Institute for Genomic Diversity (IGD) at Cornell University for Genotyping-by-Sequencing (GBS; Elshire et al. 2011) to produce SNP data.

Whole genome resequencing: Accessions of *P. kansuensis*, *P. davidiana*, *P. dulcis*, *P. arabica*, *P. tangutica*, *P. argentea*, Lovell, Tardy Nonpareil, *P. cerasifera*, *P. kuramica*, *P. bucharica*, *P. fenzliana* were resequenced at BGI (formerly Beijing Genome Institute; Shenzen, China) to provide additional support to all genomic and genetic analyses. In addition to these resequenced genomes we are also utilizing whole genome resequencing data of eight *P. dulcis* from another project, four public *P. dulcis*, thirteen public *P. persica*, and three public wild peach species. All data were aligned to the peach reference sequence (Verde et al. 2013) using BWA-MEM (Li 2013, Li and Durbin 2009, Li and Durbin 2010) then variants called with SAMTOOLS and filtered with BCFTOOLS (Li et al. 2009). The variants were called in groups, which provide greater accuracy. The different groups were composed as follows: all samples, all wild and domestic peach and almond samples, only wild and domesticated almond samples, only wild and domesticate peach samples, only domesticated almond samples, and only domesticated peach samples.

Gene expression profiling: Gene expression profiling reveals the patterns of whole genome gene expression and comparing expression profiles between resistant and susceptible rootstock genotypes will permit identification genes expressed in resistant rootstock genomes.

This permits targeting gene spaces specifically responsible for disease resistance. We have selected four commercial rootstocks (widely used Marianna 26-24, Lovell, and Nemaguard) and a fourth one is a promising rootstock in multiple trials (Krymsk 1). The selections have contrasting responses to soil borne pests and pathogens, such as, root knot nematode (RKN), root lesion nematode (RLN), ring nematode (RN), bacterial canker (BC), *Phytophthora* (P), *Armillaria* (A), and crown gall (CG) (**Table 1**).

Association Analyses

Disease screening data: Unpublished categorical screening data for CG was provided by Kluepfel lab at USDA-ARS. Quantitative root knot and root lesion nematode data was gleaned from online sources originating from the McKenry lab at UC Riverside.

SNP data: GBS data for 190 commercial rootstocks, experimental rootstocks, breeder/nursery selections, and diverse species used as parents or of interest was produced in IGD. There were 221,115,528 sequencing reads of which 203,429,249 could be filtered by barcode producing 18,596,235 unique Tags. Resulting SNPs ranged from 164,742 to 909,352, depending on the calling method and filtering parameters. [Resulting SNPs: 483,994 unfiltered HapMap, 164,742 filtered HapMap, and 909,352 VCF].

Data analyses: TASSEL (Bradbury et al., 2007): Mixed Linear Model (MLM) which uses population structure and co-ancestry among genotypes as covariates in computing marker-disease phenotypes associations was employed. Population structure was computed by generating eigen vectors using the principal components analysis (PCA) and coefficient of co-ancestry (kinship matrix) was generated using a module available in the TASSEL software. These two parameters along with SNP genotype data and numerical (quantitative or categorical) phenotype data were used in association analysis.

PLINK analysis (Purcell et al., 2007): The association analysis using a general linear model (GLM), which accounts for only population structure, was used to compute associations. SNP data filtered to keep only biallelic loci. Loci with frequency less 5% and missing >10% data were discarded. The SNPs were then filtered based on pair-wise linkage disequilibrium and subjected to a multidimensional scaling (MDS) analysis (**Figure 2**). The coordinates extracted from the MDS analysis accounting for the population structure was used in the association analysis. A quantile-quantile (QQ) plot of probability distribution of observed associations was generated to assess the goodness of fit with expected probability distributions. The chromosome-wise probability of associations between marker and phenotypes was displayed in a Manhattan plot (**Figure 3**).

Results and Discussion:

Novel rootstocks produced within the project

The first round of disease testing of novel rootstocks generated in the project has yielded encouraging results with wide variation in response to CG and *Phytophthora* screening (**Figures 4 and 5**). The three rootstock genotypes showing resistance to *Phytophthora* were planted in WEO after budding them with ‘Nonpareil’ for further evaluation. These experimental hybrids, which have undergone one round of disease testing along with others from various

disease testing trails (190 genotypes) from Kluepfel and Browne lab were included in the association analysis.

The GBS data from the IGD contained 221,115,528 sequence reads of which 203,429,249 could be filtered by barcode producing 18,596,235 unique Tags. Resulting SNPs ranged from 164,742 to 909,352, depending on the calling method and filtering parameters. The genotype data in combination with disease phenotype data generated by collaborating plant pathologists was utilized for association analysis. The multidimensional scaling (MDS) analysis performed on the SNP data for 190 genotypes indicated simple genetic structure (**Figure 2**), which suggests the least influxes of genetic structure on the genotype-disease phenotype associations.

Association Analyses

The results of association analyses performed using two different methods, mixed linear model as implemented in the TASSEL, which along with genetic structure considers coefficient of co-ancestry of genotypes in computing association is considered more robust than the general linear model implemented in the PLINK analysis, which uses only genetic structure as covariate to compute associations. While these analyses are still preliminary, at this time we have identified significant association of markers with CG, but the analyses failed to come up with any markers for either root knot or lesion nematode infestations. We are further filtering the data set using a number of filtering options to eliminate poor data among the SNP loci identified in the GBS analysis. We expect significant association of SNPs with other diseases (RNK/RLN) when once the data is cleaned up.

The MLM results (**Table 2**) indicate several SNPs across the genome with significant association ($p<0.05$) with CG and R^2 values ranging from 0.09 to 0.11, which is considered significant for traits such as disease resistance with complex inheritance patterns with generally low heritability. It appears that the genetic loci or quantitative loci (QTLs) modulating CG resistance occurs in four different linkage groups. The marker with significant association is found located on chromosome 8 based on the peach reference genome used in this study for SNP discovery and genotyping. The probability values indicate potential association, but the low coefficient of determination (R^2) suggests the markers account for only limited variation in CG disease response.

The analysis using the GLM in PLINK, however, showed a significant association of a SNP with CG located on chromosome 1 (**Table 3**) at the end of an Exon for a putative iron oxidoreductase enzyme gene in *Arabidopsis thaliana*, which appears to be homologous to transcript of a putative gene in *P. persica* (**Figure 6**). Some members of this gene family appear to have a role in pathogen defense response (van Damme et al. 2008).

Genome resequencing and SNP calling

In the genome sequence data a minimally filtered SNP calling discovered from 2,026,306 to 24,206,662 SNPs, depending on species combination used for calling. The lowest number was from the group with only *P. persica* samples whereas the highest number was from the group of all samples across species. With over 10 million SNPs *P. dulcis* had approximately five times the number of SNPs as *P. persica*. This difference is expected when mating systems of the two species, outcrossing versus selfing, are considered.

Transcription profiling

Tissue specific transcription profiling permits identification of genes locally expressed in tissues or organs in response to internal or external stimuli. For example, the soil borne diseases may induce gene expression in roots in the presence of pathogens and may result in a resistant or susceptible reaction. Localizing such candidate genes permit for developing candidate gene based markers, which in most cases are associated with the disease responses. Sampling root tissue and examining gene expression profiles under pathogen challenged and non-challenged conditions may point to genes mediating soil borne disease response. The gene expression profiles will be compared to identify genes that are preferentially expressed to confer resistance to soil borne diseases. This permits targeted discovery of marker with significant association with resistance to soil borne diseases.

References Cited:

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

The International Peach Genome Initiative, Verde I, Abbot AG, Scalabrin S, Jung S, Shu S, Marroni F, et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, 45:487-494.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60. [PMID: 19451168]

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, Epub. [PMID: 20080505]

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25:2078-2079.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker R, Lunter G, Marth G, Sherry ST, McVean G, Durbin R, and 1000 Genomes Project Analysis Group. 2011. The Variant Call Format and VCFtools. *Bioinformatics*, 27:2156-2158.

McKenry MV. Susceptibility of Prunus Rootstocks to Nematodes (Trials by M.V. McKenry). Available at http://ucanr.edu/sites/fruitreport/files/78641.pdf

BGI Cognitive Genomics. Plink 1.9. (https://www.cog-genomics.org/plink2/)

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bder D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

van Damme M, Huibers RP, Elberse J, Van den Ackerveken G. 2008. Arabidopsis RMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *The Plant Journal*. 54:785-793.

Table 1. Rootstocks selected for gene expression profiling.

Rootstock	Parentage	RKN	RLN	RN	BC	P	A	CG
Krymsk 1	*P. tomentosa x cerasifera*	S	R	S	S	U	U	S
Marianna 26-24	*P. cerasifera x munsoniana*	R	S	S	S	R	R	R
Lovell	*P. persica*	S	R	R	R	S	S	S
Nemaguard	*P. persica x davidiana*	R	R	S	S	R	S	R

R/S, Resistance/susceptible; U, unknown; RKN, Root knot nematode; RLN, root lesion nematode; RN, ring nematode; BC, bacterial canker; P, *Phytophthora*; A, *Armillaria*; CG, crown gall
source: UC ANR Fruit Report on rootstocks at http://ucanr.edu/sites/fruitreport/Rootstocks/

Table 2. The Mixed Linear Model (MLM) association analysis of Single Nucleotide Polymorphisms (SNP) and crown gall data using TASSEL.

Marker	Chr	Site	df	F	p	Error df	Marker R^2
S1_212063151	8	15627172	2	5.94159	0.00831	36	0.11168
S1_72042880	2	25165154	2	6.29985	0.00843	31	0.10136
S1_54210819	2	7333093	1	10.30516	0.00513	29	0.09392
S1_136457227	*5*	*10217200*	*2*	*6.0658*	*0.00969*	*31*	*0.09334*
S1_136457228	*5*	*10217201*	*2*	*6.0658*	*0.00969*	*31*	*0.09334*
S1_136457231	*5*	*10217204*	*2*	*6.0658*	*0.00969*	*31*	*0.09334*
S1_7294016	1	7294016	2	5.72594	0.00959	36	0.0907

Table 3. Adjusted general linear model (GLM) association analysis of SNP and crown gall data using PLINK.

CHR	SNP	Bonferroni p-vlaue	Holm_Bonferroni p-value	QQ p-value
1	S1_38007759	0.0447	0.0447	1.43E-05

Figure 1. New embryo rescued interspecific *Prunus* hybrids clonally propagated for disease testing.

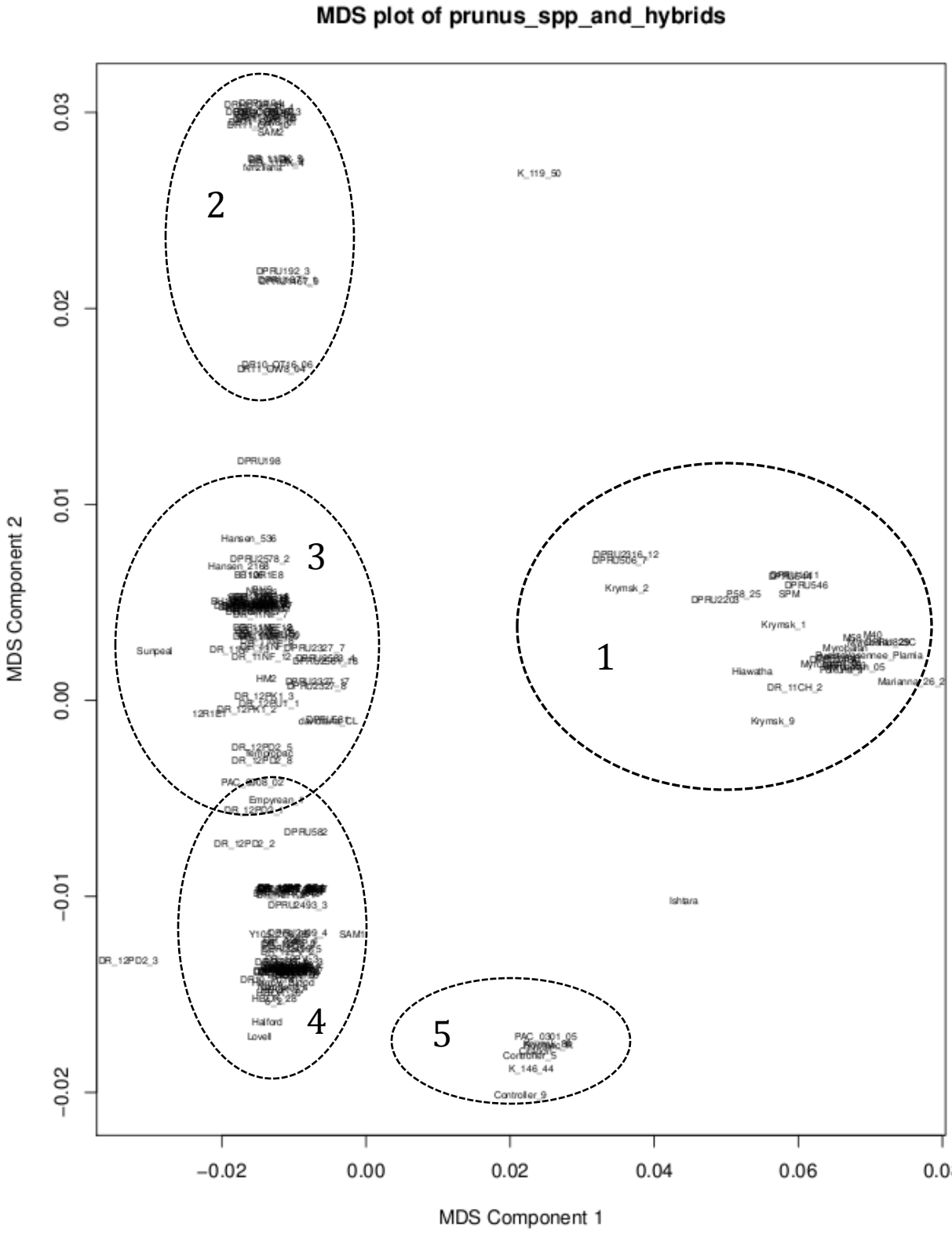

Figure 2. A 2-dimensional plotting of MDS coordinates of SNP data. Gr. 1 - composed of samples with plum background, Gr. 2 - almond and wild almond species, Gr. 3 - peach-almond hybrids, Gr. 4 - peach and wild peach species, and Gr. 5 - hybrids with peach x plum parentage.

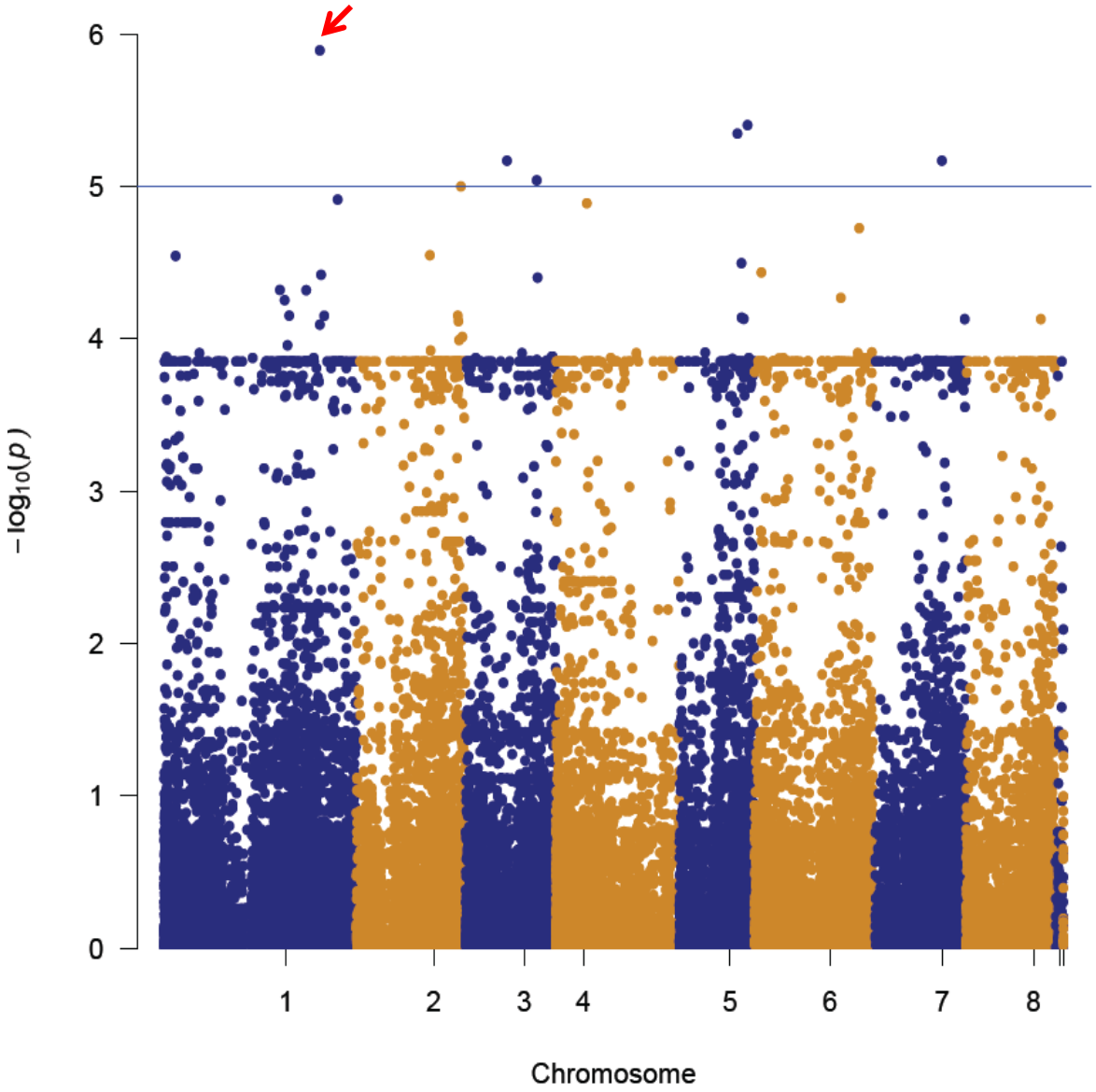

Figure 3. Manhattan plot - distribution of $\log_{10}$(p-values) showing the probability of association between SNP markers and crown gall (CG). Red arrow indicates a SNP with significant association with CG.

Figure 4. Disease testing of interspecific rootstock hybrids for *Phytophthora*.

Figure 5. Disease testing of interspecific rootstock hybrids for crown gall.

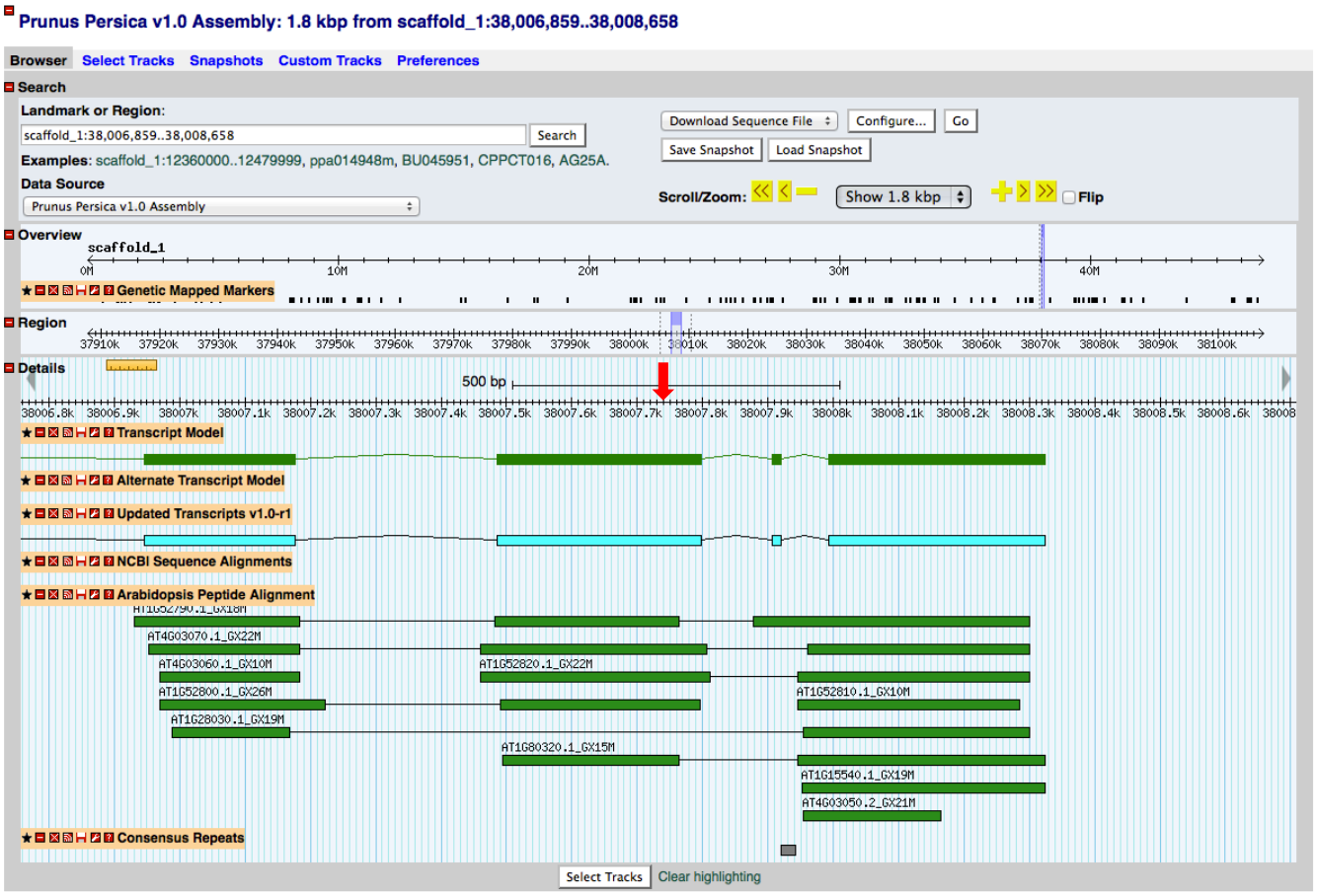

Figure 6. Genome location of SNP significantly associated with crown gall pathogen response. The red arrow indicates the approximate position of the SNP to within a few base pairs, which is putatively located within the exon of a member of an oxidoreductase oxygenase gene family.