Genome wide association study (GWAS) for bacterial wilt resistance in the *Capsicum* core collection

By Lara Brindisi, PhD Candidate

Center for Agricultural Food Ecosystems (RUCAFE)

Rutgers, the State University of New Jersey, USA

Final Report

Submitted to the World Vegetable Center, Taiwan

In Partial Fulfilment of the Requirements

For the Student Internship Program

Supervised by:

Dr. Derek Barchenger (Plant Breeding)

Dr. Ya-ping Lin (Bioinformatics)

Ms. Jaw-Rong Chen (Bacteriology)

Abstract

Bacterial wilt is the most destructive disease for many plants in the Solanaceae family. Bacterial wilt in peppers (*Capsicum* spp.) is caused by *Ralstonia solanacearum.* To identify potential sources of resistance, the G2PSol *Capsicum* core collection and Gene bank accessions from the World Vegetable Center in Taiwan was evaluated using a high throughput laboratory based bacterial wilt screen. The resulting phenotype data was combined with existing genotype data to identify marker trait associations (MTAs) as part of a larger genome wide association study (GWAS). Population structure was estimated to account for the number of ancestral populations using PCA and structure analysis. MTAs were identified on chromosomes 1, 2, 4, 6, 7, 10 and 11 with the most significant being on chromosomes 4, 7 and 11. Future work will focus on screening the rest of the *Capsicum* core accessions, comparing the lab results with greenhouse and field results and re-analyzing the data with high density SNPs.

Introduction

Bacterial wilt is the most destructive disease for many plants in the Solanaceae family. Bacterial wilt in peppers (*Capsicum* spp.) is caused by *Ralstonia solanacearum.* This pathogen is highly complex as it has a broad host range and high levels of diversity within the species with 5 races and 6 biovars (Mamphogoro et al. 2020; EPPO 2022). Symptoms include vascular necrosis, root rot, leaf drooping and dieback (Nelson 2007). Resistance has been identified in several small-fruited type pepper plants, including ‘White Khandari’ and ‘Manganji’ (Matsunaga and Monma 1995; Mimura et al. 2009, Tsuro et al. 2007).

Quantitative trait loci (QTL) mapping has been performed on several crosses and have identified significant QTL on chromosome 1 against strains ‘KP9547’ and ‘WR-1’ (Mimura et al. 2009; Chae et al. 2022), chromosomes 5, 6, and 7 against strain ‘HWA’ (Lee et al. 2022),

chromosomes 7, 8, 9 and 10 against strain ‘HS’ (Lee et al. 2022) and chromosome 10 against ‘Rs-SY1’ (Du et al. 2019). This suggests that bacterial wilt resistance in pepper is likely polygenic and strain specific. To date, no genome-wide association studies (GWAS) have been performed on the bacterial wilt resistance of *Capsicum* spp. This study seeks to be the first to use GWAS to identify marker trait associations (MTAs) for bacterial wilt resistance in peppers using GWAS on the G2PSol *Capsicum* core collection from the World Vegetable Center and additional accessions from the Gene bank in Taiwan.

Materials and Methods

Bacterial wilt seedling preparation

Accessions (n=215) were selected from the *Capsicum* core collection. Seeds (50 per accession) were surface sterilized enclosing those of each accession in mesh tied with rubber bands and treating with a solution of 1% sodium hypochlorite (NaClO; bleach) and 0.05% Tween® 20 detergent, stirring for 5 minutes (Fig. 1). Solution was replaced with sterile water and stirrd for 5 minutes, which was repeated three more times. Seeds were placed in sterile tissue culture boxes on a stack of paper towels saturated with ~100 mL sterile water. PBC1367 was included in every experiment as the susceptible control. Tissue culture boxes were placed in the growth chamber at 28°C in the dark until ~50% of the seeds germinated, and then placed under light (12h photoperiod 4891 Lux; 90 μ-mole/s.m2; the distance between lamp and seedling is 30 cm). Pepper seedlings were ready for screening when the cotyledon fully expanded ~2-4 weeks later.

Bacterial wilt bacterial suspension preparation

Ralstonia pseudosolanacearum strain ‘Pss2074’ (phylotype I, biovar 3, sequevar 34) was maintained via single colony isolation on TTC selective media incubated at 30°C for 2-3 days

(Fig. 2). Pss2074 was incubated for 24 h at 30°C on 523 media. Pss2074 was gently scraped from the media and mixed with 300 mL of water until turbid (OD600 = 3.0-3.2) for screening.

Bacterial wilt laboratory screen

Seedling root tips were cut (30 per accession) and dipped in bacterial suspension for 1 min (Fig 3). Seedlings were transferred to empty plates and exposed to air for 5 minutes. Seedlings were then placed in microcentrifuge tubes filled with 2 mL water in one row of 10 randomly on different racks (n=3) to block for location in the growth chamber. This was repeated without bacterial suspension dip for the controls (6 per accession) and were placed randomly on a separate rack (Fig. 4). Tube racks with one of each accession and the controls were placed on drainage boards in 3 clear plastic bins with 200 mL of water and covered with perforated plastic wrap. Seedlings were incubated in the growth chamber at 28°C in the dark until ~50% of the seeds germinated, and then placed under light (12h photoperiod 4891 Lux; 90 μ-mole/s.m2; the distance between lamp and seedling is 30 cm). Seedlings were scored for severity of bacterial wilt 9-11 days after inoculation (DAI) on a 0-3 scale with 0 representing no damage; 1 representing browning or softening of the stem with potentially slight damage of the leaves; 2 representing moderate browning of the stem and leaves; and 3 representing complete browning. Individuals were considered resistant with a rating less than or equal to 1 and susceptible with a rating greater than 1.

Phenotyping bacterial wilt

Raw scores were recorded as “BW Percent”, “BW Index” or “BW Severity” per block and averaged. BW Percent was calculated as the number of seedlings with a score greater than zero divided by the total number of seedlings. BW Index was calculated as the disease index or the sum of the seedlings with each rating multiplied by the value of that rating, divided by the

total number of seedlings multiplied by the maximum rating. BW Severity was calculated as the average value of the scores. Phenotypes were also transformed into the "Percent Change", which was also log transformed. Percent change was calculated as the value of the control subtracted by the value of the accession divided by the value of the control. A negative value was more resistant than the susceptible control while a positive value was more susceptible than the susceptible control. Calculations are shown in Fig. 5.

GWAS analysis

GWAS was performed to identify marker trait associations. The genotype data provided (pepper_core_431.vcf) was imputed with the LD KNNI method and filtered by a site minimum allele frequency of 0.05 with default settings in TASSEL v5.2.86. Output was saved as Geno1_KNNimp_Filter.hmp.txt. A PCA matrix was constructed with 5 dimensions. Genotype, phenotype and PCA data were incorporated into a generalized linear model (GLM) and the stats output was saved as GLM_geno1_phenoF_PC5_del.txt. A kinship matrix constructed with default settings and added into the mixed linear model (MLM) and the stats output was saved as MLM_geno1_phenoF_PC5_K_del.txt . Accessions screened without the susceptible control on 11/07/22 and 11/28/22 were removed from the analysis for a total of 181 accessions (Cumulative_del_tassel sheet in Lara Ralstonia Evaluations 2022.xlsx).

The analysis was repeated to include the results from the summer student interns. The genotype data provided (pepper_core_431.vcf and BW_all.hmp.txt) was merged, imputed with the LD KNNNI method and filtered by a site minimum allele frequency of 0.05 with default settings in TASSEL. Output was saved as Geno12_KNNimp_Filter.hmp.txt. A PCA matrix was constructed with 5 dimensions. A covariate was added where a value of 1 represented the summer data and 2 represented this fall data (Covariate sheet in Lara Ralstonia Evaluations

2022.xlsx). Genotype, phenotype, PCA and covariate data were incorporated into a generalized linear model (GLM) and the stats output was saved as GLM_geno12_phenoFS_PC5_cov1.txt. A kinship matrix constructed with default settings and added into the mixed linear model (MLM) and the stats output was saved as MLM_geno12_phenoFS_PC5_K_cov1.txt. All accessions screened were included in the analysis for a total of 411 accessions. However, only the BW Percent phenotype was calculated (Summer+fall_tassel sheet in Lara Ralstonia Evaluations 2022.xlsx). Percent change and log transformations could not be applied as the summer work screened the susceptible control separately from the accessions.

Population structure analysis

Principal component analysis was conducted in R/Rstudio v4.2.0. Genetic distances were exported as a matrix from TASSEL and then analyzed in PCA using the cmdscale() command. Structure analysis was performed using R/LEA v3.10.0 on the merged genotype data from summer and fall. Imputed and filtered genotype data from TASSEL was converted to the lfmm structure via the vcf2lfmm() command for structure analysis. Individual ancestry coefficients and ancestral allele frequencies were calculated with the snmf() command and admixture coeffects with the Q() command.

Results

Bacterial wilt screen

Most accessions were susceptible and few were resistant. Representing the phenotype as BW Percent resulted in a heavy skew of the data toward towards susceptibility while the BW Disease Index and BW Severity were more normally distributed. The percent change and log transformations increased the normality for the BW percent data but not for the BW Disease index or the BW Severity data. However, these calculations incorporated a consistent control in

every screen to “normalize” the data. Strain Pss2074 was more virulent than strain Pss71, which was used by the summer student interns.

GWAS analysis

The GWAS on the fall data revealed significant MTAs on chromosomes 4 and 7 for most of the phenotypes in the GLM and MLM models (Figs. 7 and 9) with additional MTAs on chromosomes 1, 2, 10 and 11. The BW Disease Index (Avg) and BW Disease Severity (Avg) did not detect any significant MTAs, which demonstrated the importance of directly comparing the data to the control in this experimental design. The MLM generally performed better than the GLM according to the qq plots (Figs. 8 and 10) but did result in a loss of SNPs from 5,668 SNPs in the GLM to 3,069 SNPs in the MLM. The most significant MTA was on chromosome 11 in the GLM, which was not significant in the MLM. The MTA on chromosome 4 was highly significant in the GLM and the most significant in the MLM.

The GWAS on the merged fall and summer data detected significant MTAs on chromosomes 1, 4, 7 and 11 using the GLM, which were also detected in the fall only data. However, an additional MTA was detected on chromosome 6 in the merged fall and summer GLM (Fig. 11). All MTAs were relatively comparable in significance. No MTAs were identified using the MLM (Fig. 13). The MLM did not perform better than the GLM model as it did with the fall data only (Figs. 12 and 14). Additional models could be run to better incorporate relatedness and cofactors in the MLM.

Population structure analysis

Population structure was observed in the merged fall and summer dataset. About 5 clusters were observed with two groups of *C. annuum,* and one group each of *C. baccatum, C. chinense, C. frutescens* (Fig. 15). Many genotypes were unknown and a few *C. annuum* clustered

with other species, suggesting potential mislabeling of the species name for a few individuals. Cross-entropy criterion analysis suggested 5 ancestral populations (Fig. 16), thus the ancestry matrix was constructed with k=5 (Fig. 17). This also supported using 5 components in the PCA matrix for the GWAS analysis.

Discussion

This data has a lot of potential for incorporation into a scholarly article. To the author's knowledge, no GWAS has yet been performed on the bacterial wilt of peppers (*Capsicum* spp.). QTL studies have been performed on this topic and have identified significant genomic regions on some but not all the chromosomes reported in this paper. QTL have been detected on chromosomes 1 (Mimura et al. 2009; Chae et al. 2022), 6, 7 (Lee et al. 2022) and 10 (Du et al. 2019), but no QTL or MTAs have been detected on chromosomes 2, 4 or 11. More research needs to be done to identify if this work identifies overlapping or novel genomic regions compared to those QTL reported. Other next steps including screening the rest of the *Capsicum* core accessions and repeating the analysis, comparing the lab results with greenhouse and field results and re-analyzing the data with high density SNPs to observe if results are consistent.

Acknowledgements

This work was made possible with funding from the Fruit and Vegetables for Sustainable Healthy Diets (FRESH) by CGIAR and the Council of Agriculture in Taiwan. I would like to extend my deepest gratitude to Dr. Derek Barchenger 德瑞克, Dr. Ya-ping Lin 林亞平 and Ms. Jaw-rong Chen 林亞平 for their mentorship and support. I would like to thank Hung-chia Chang 張紘嘉 and Mei-hong Chuang 莊美紅 for their guidance and teamwork in accomplishing all the bacterial screens. I would also like to thank Hank Lin 林宗翰 and Yulita Dwiningsih 優里塔 for assisting with the screens.

References

Alqudah AM, Sallam A, Baenziger PS and Börner A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. *Journal of Advanced Research, 22*, 119-135. https://doi.org/10.1016/j.jare.2019.10.013

Chae S, Lee K, Do J, Hong S, Lee K, Cho M, Yang E and Yoon J. (2002). QTL mapping of resistance to bacterial wilt in pepper plants (*Capsicum annuum*) using Genotyping-by-Sequencing (GBS). *Horticulturae, 8*(2), 1-21. https://doi.org/10.3390/horticulturae8020115

Du H, Wen C, Zhang X, Xu X, Yang J, Chen B and Geng S. (2019). Identification of a major QTL (*qRRs-10.1*) that confers resistance to *Ralstonia solanacearum* in pepper (*Capsicum annuum*) using SLAF-BSA and QTL mapping. *International Journal of Molecular Sciences, 20*(23), 1-20. https://doi.org/10.3390/ijms20235887

EPPO (2022). *Ralstonia solanacearum* species complex (RALSSO). *EPPO Global Databas.* *https://gd.eppo.int/taxon/RALSSO/distribution*

Lafortune D, Beramis M, Daubeze AM, Boissot N and Palloix A. (2005). Partial resistance of pepper to bacterial wilt is oligogenic and stable under tropical conditions. *Plant Disease, 89*, 501–506.

Lee S, Chakma N, Joung S, Lee JM and Lee J. (2022). QTL mapping for resistance to bacterial wilt caused by two isolates of *Ralstonia solanacearum* in chili pepper (*Capsicum annuum* L.). *Plants*, *11*(12), 1-17. https://doi.org/10.3390/plants11121551

Mamphogoro TP, Babalola OO and Aiyegoro OA. (2020). Sustainable management strategies for bacterial wilt of sweet peppers (*Capsicum annuum*) and other Solanaceous crops. *Journal of Applied Microbiology, 129*(3), 496-508. https://doi.org/10.1111/jam.14653

Matsunaga H and Monma S. (1995). Varietal differences in resistance to bacterial wilt in related species of *Capsicum annuum*. *Capsicum Eggplant Newsletter, 14*, 60-61..

Mimura Y, Kageyama T, Minamiyama Y and Hirai M.(2009). QTL Analysis for Resistance to *Ralstonia solanacearum* in *Capsicum* Accession 'LS2341'. *Journal of the Japanese Society for Horticultural Science, 78*(3), 307-313. https://doi.org/10.2503/jjshs1.78.307

Mimura Y, Yoshikawa M and Hirai M. (2009). Pepper accession LS2341 is highly resistant to *Ralstonia solanacearum* strains from Japan. *HortScience, 44*, 2038–2040.

Nelson SC. (2007). Bacterial Wilt. An Illustrated Glossary of Tropical Plant Pest and Diseases, University of Hawaii, Manoa, *CTAH.* https://www.ctahr.hawaii.edu/nelsons/glossary/Bacterial_wilt.htm

Tsuro M, Minamiyama Y and Hirai M. (2007). QTL analysis for bacterial wilt resistance in Japanese pepper (*Capsicum annuum* L.). *Breeding Research. 9*, 111–115.

Figures

Figure 1. Surface sterilization of selected seeds.

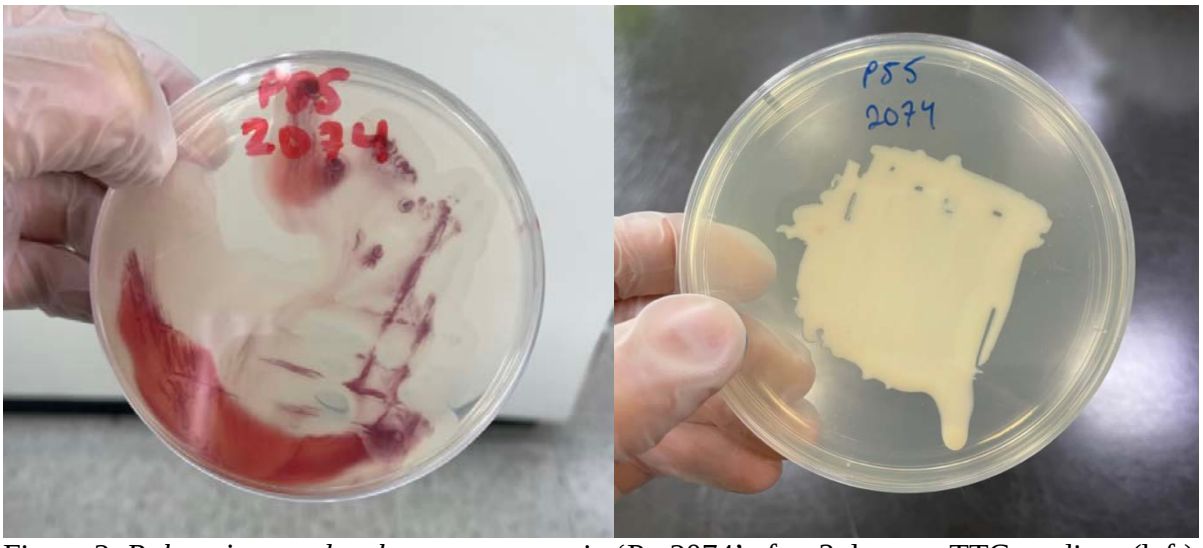

Figure 2. *Ralstonia pseudosolanacearum* strain ‘Pss2074’ after 3 days on TTC medium (left) and 523 medium (right).

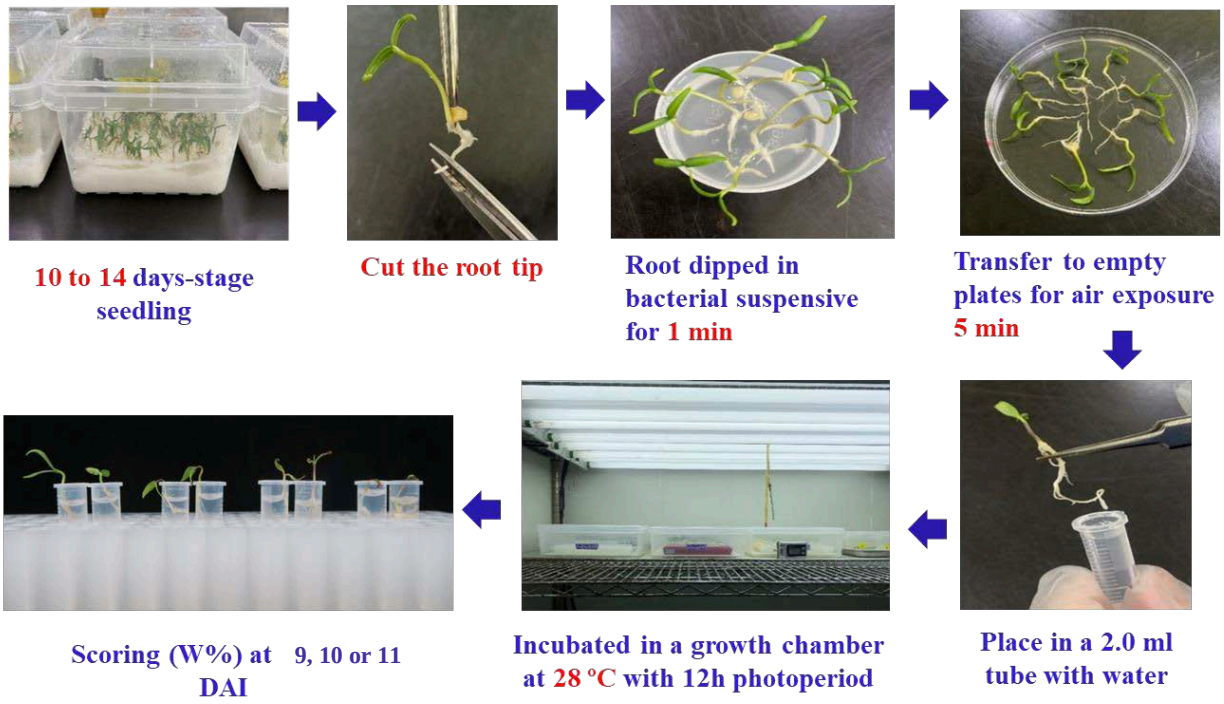

Figure 3. Bacterial wilt screen methodology.

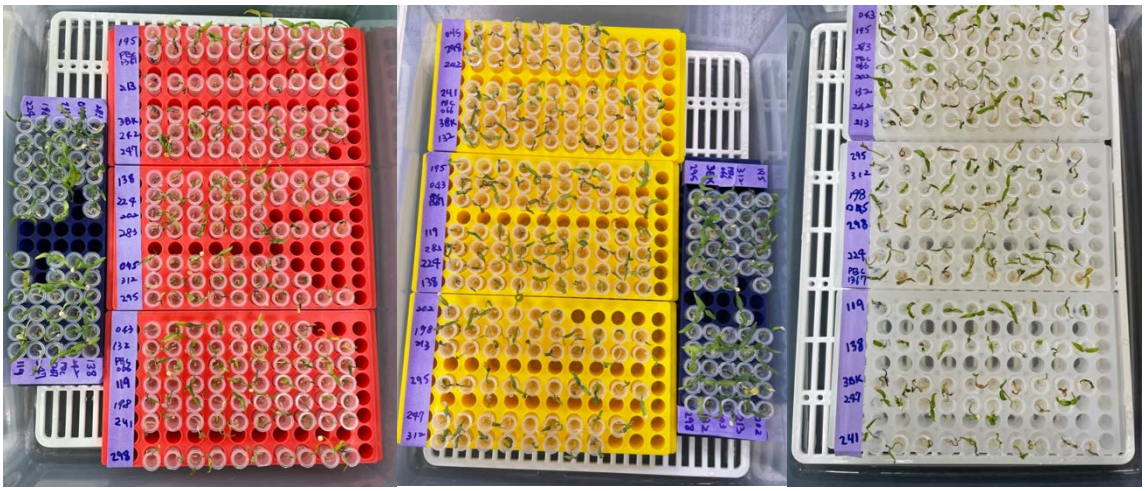

Figure 4. Bacterial wilt screen set up before incubation for 9-1 days.

Visual rating (0-3)

$$\text{BW Percent} = \begin{cases} \frac{\sum N_i}{N} \times 100 & if X_i > 0 \end{cases}$$

$$\text{BW Index} = \frac{\sum_{r=1}^{3} (N_r \times r)}{N \times 3 \text{ (maximum rating)}}$$

$$\text{BW Severity} = \begin{cases} \frac{\sum X_i}{N} \end{cases}$$

$$\text{Percent Change} = \frac{BW\ Phenotype_i - BW\ Phenotype_c}{BW\ Phenotype_c} \times 100$$

$$\text{Log} = log_{10}(Percent\ Change(BW\ Phenotype))$$

N_i = individual plant
N = total number of plants
X_i = rating of individual plant
r = rating (0-3)
N_r = number of plants with r rating
$BW\ phenotype_i$ = BW Percent, BW Index or BW Severity value of individual plant
$BW\ phenotype_c$ = BW Percent, BW Index or BW Severity value of individual plant

Figure 5. Phenotyping mathematical formulas.

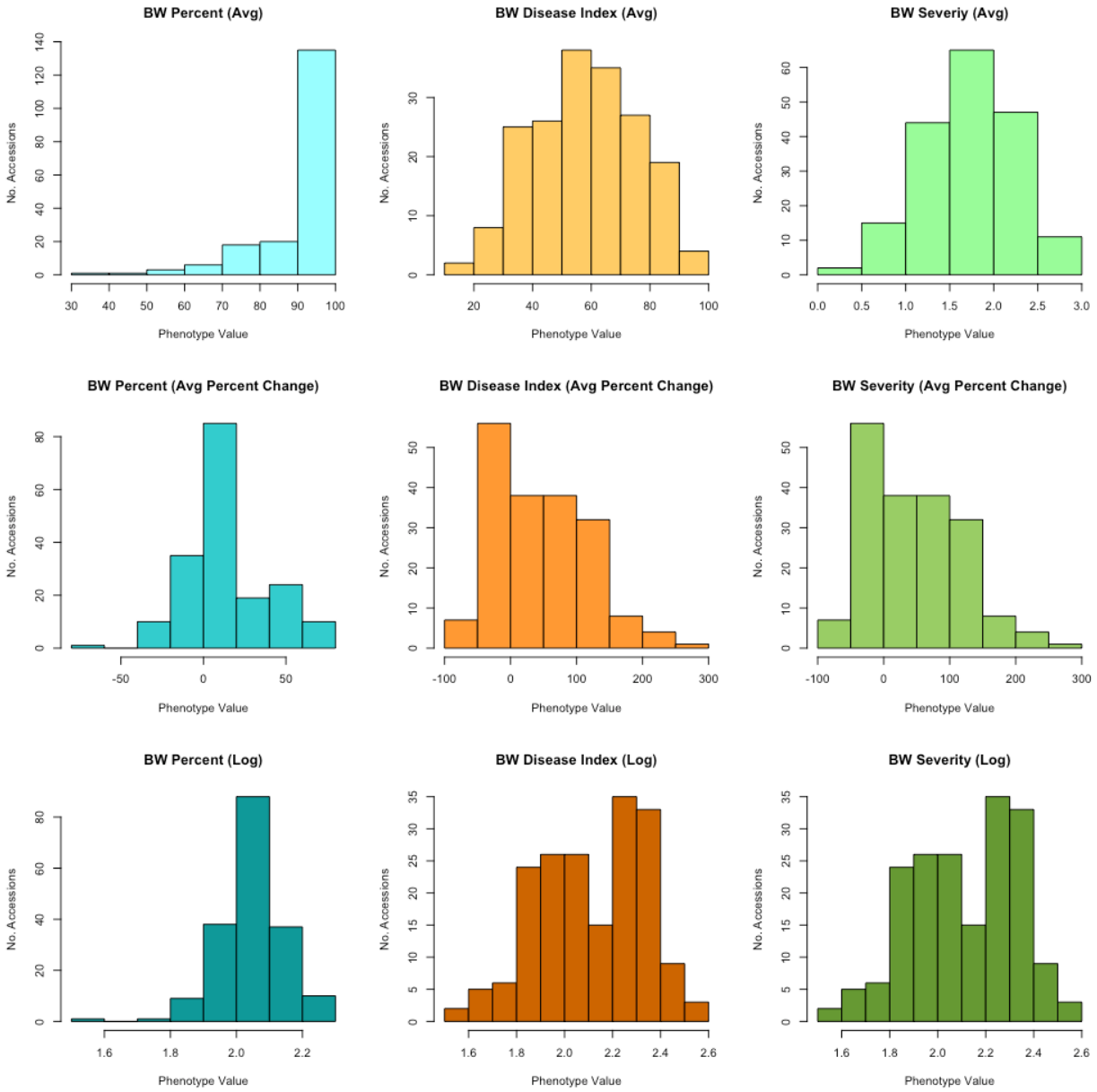

Figure 6. Histograms of each bacterial wilt screen phenotype. Avg represents the average the blocks of the BW Percent, BW Disease Index or the BW Severity scores. The Avg Percent Change represents that percent change values of each Avg score. The Log represents the log transformation of the Avg Percent Change scores.

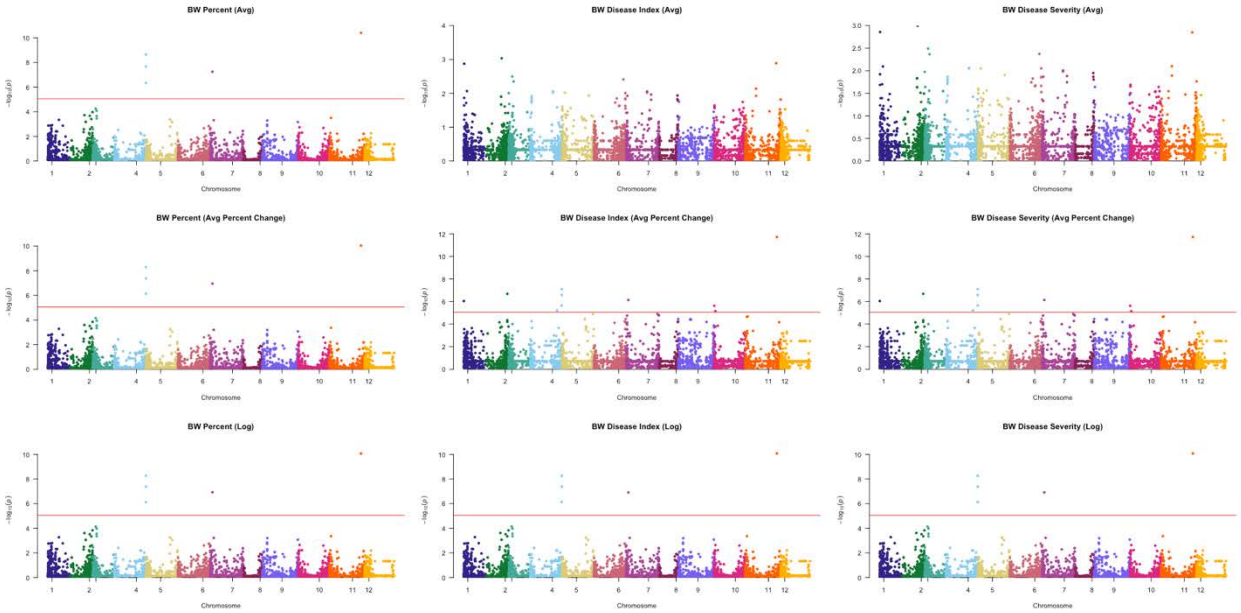

Figure 7. Manhattan plots for GLM of fall data only with 5,668 SNPs and LOD=5.05

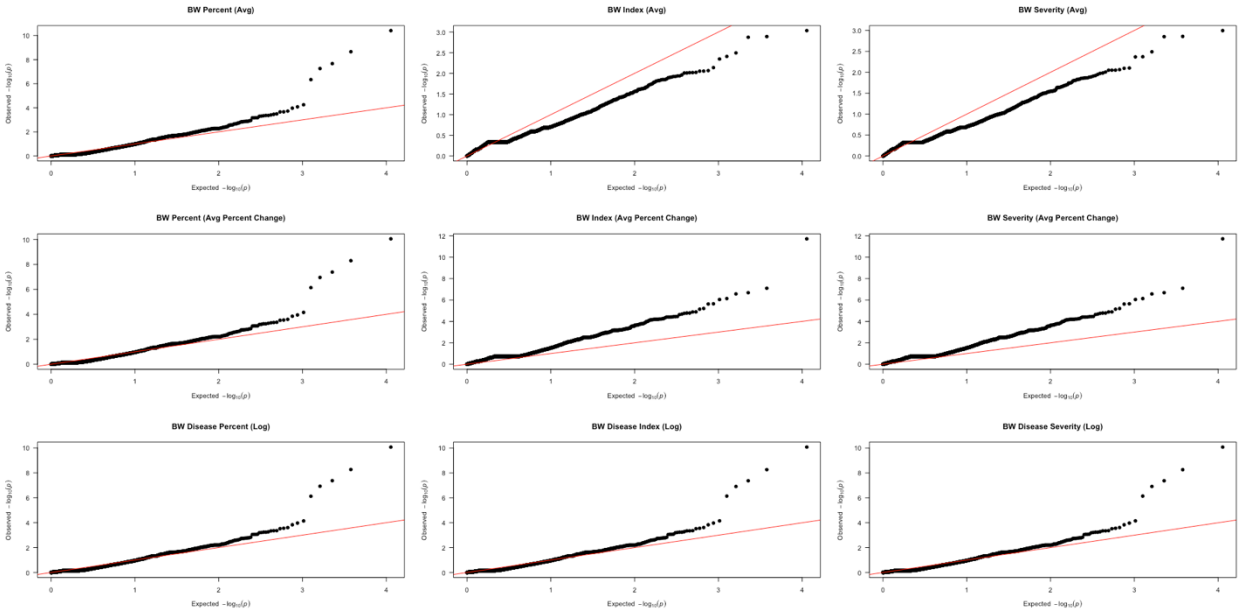

Figure 8. QQ plots for GLM of fall data only.

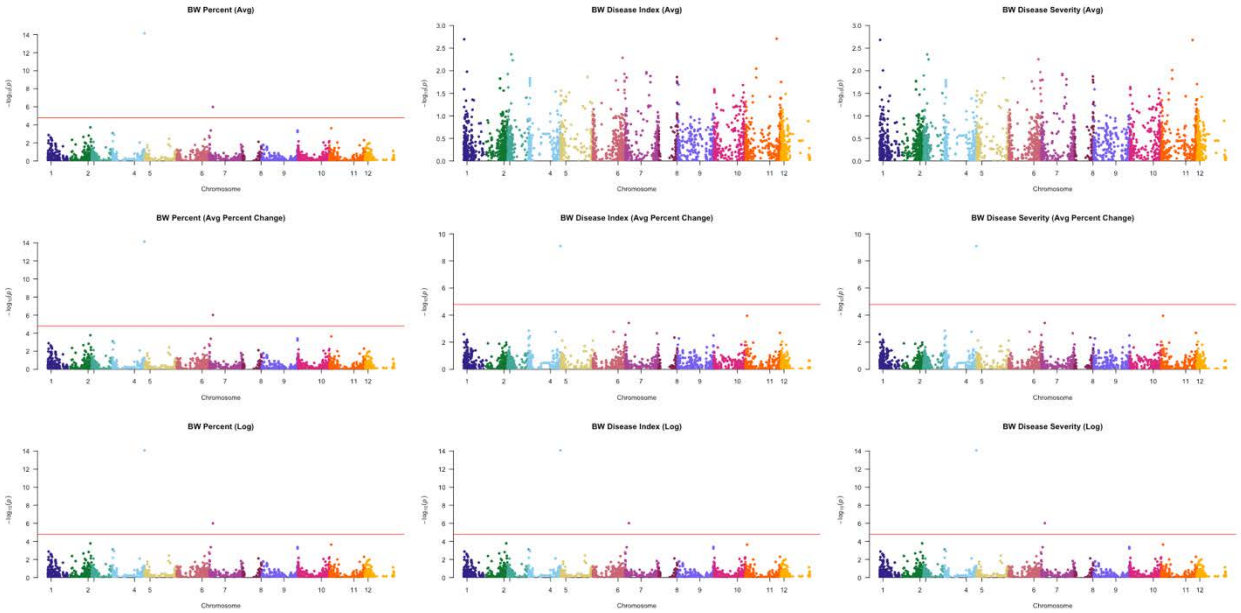

Figure 9. Manhattan plots for MLM of fall data only with 3,069 SNPs and LOD=4.79.

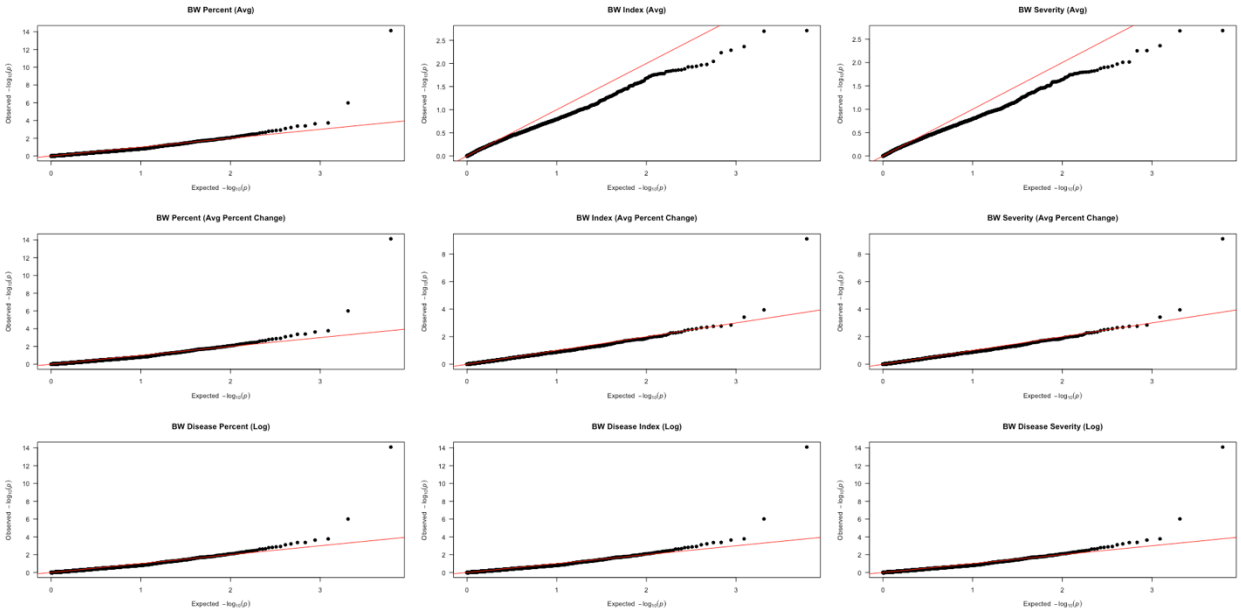

Figure 10. QQ plots for MLM of fall data only.

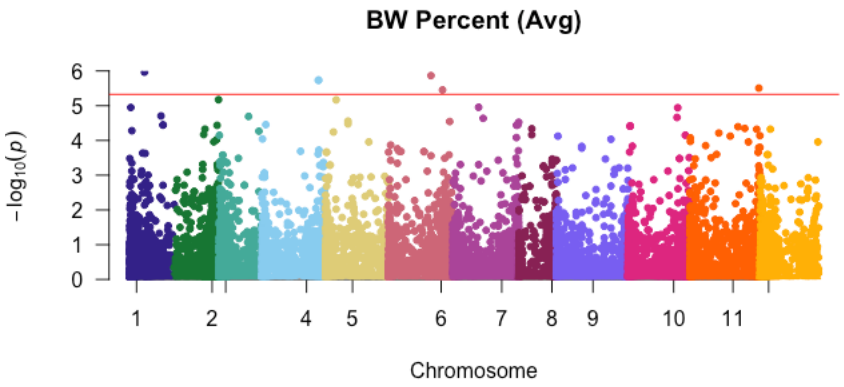

Figure 11. Manhattan plots for GLM of fall and summer data with 10,611 SNPs and LOD=5.33.

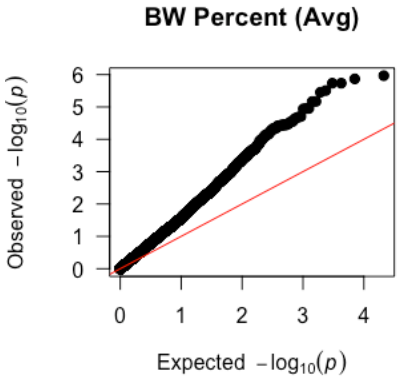

Figure 12. QQ plots for GLM of fall and summer data.

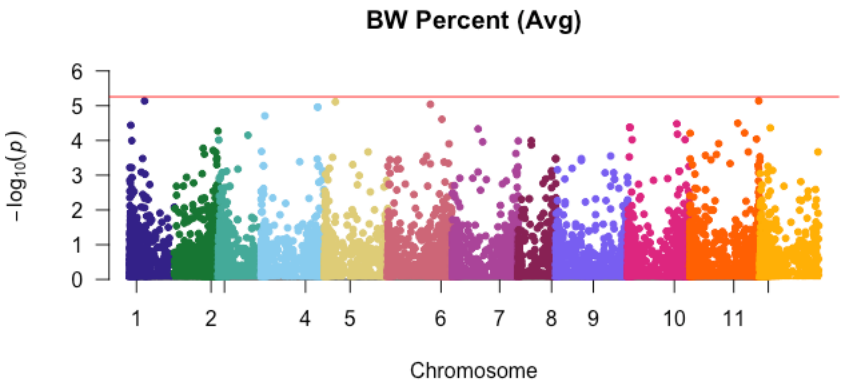

Figure 13. Manhattan plots for MLM of fall and summer data with 9,056 SNPs and LOD=5.26.

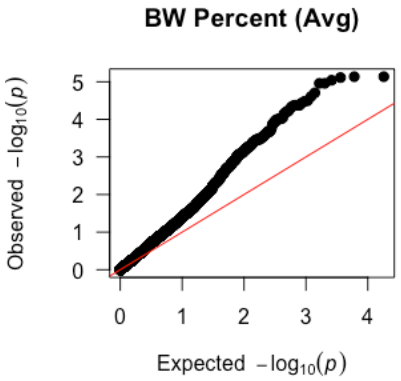

Figure 14. QQ plots for MLM of fall and summer data.

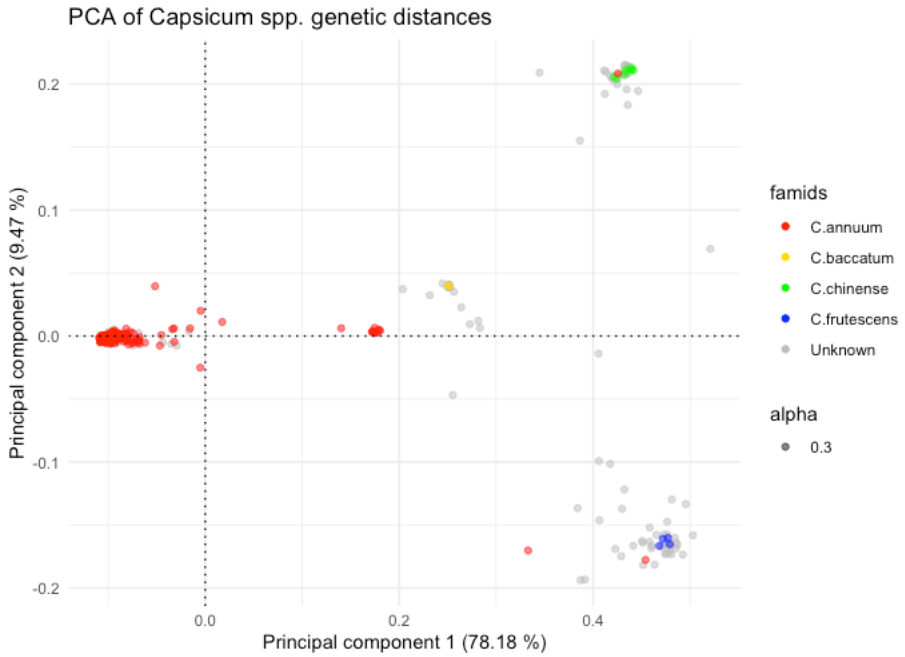

Figure 15. PCA of species incorporated in the fall and summer data.

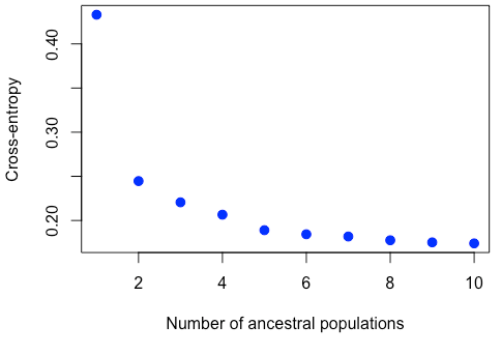

Figure 16. Cross-entropy criterion for all runs between k=1 and k=10.

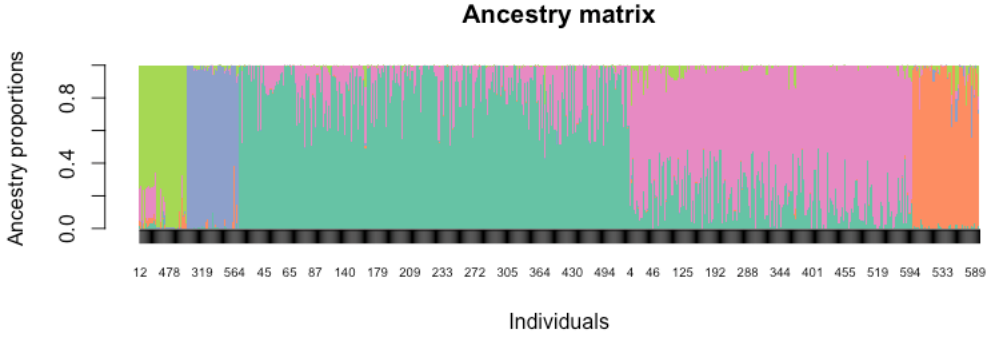

Figure 17. Population structure analysis of individuals at k=5.