PLANT BIOSECURITY
science foundation

Developing molecular ‘fingerprinting’ of myrtle rust disease to facilitate strategies in monitoring and control

Final Report (PBSF023)

DATE February 2021

AUTHOR

Michelle C. Moffitt, Jonathan M. Plett, Robert Park

Project Leader contact details

Name: Michelle C. Moffitt
Address: School of Science, Campbelltown campus, University of Western Sydney, Locked bag 1797, Penrith NSW 2751.
P: 02 4620 3521
M: 0438 553 838
E: m.moffitt@westernsydney.edu.au

Australian Plant Biosecurity Science Foundation
3/11 London Circuit, Canberra, ACT 2601

P: +61 (0)419992914
E: info@apbsf.org.au
www.apbsf.org.au

This document should be cited as:

Moffitt, Plett & Park, 2020, APBSF Project Final Report, Developing molecular 'fingerprinting' of myrtle rust disease to facilitate strategies in monitoring and control

Contents

Contents

1. Executive Summary

Before disease symptoms appear, plant pathogens and their hosts deploy chemical cues, including metabolites and proteins, that impact the speed and degree of pathogenesis. The objective of this study was to identify chemical and genetic cues produced by both *Austropuccinia psidii* and plant, prior to visible disease and investigate how these cues influence pathogenesis and relate to plant susceptibility (resistant – no observed disease; hypersensitive – localised necrosis; and susceptible – significant sign of disease).

Using an untargeted metabolomics approach, we identified a unique molecular fingerprint in *A. psidii*-infected *Melaleuca quinquenervia* leaves during the early stages of infection. Further analysis of the metabolome at 24 hours and 48 hours after infection identified a unique subset of 19 metabolites that are unique to the resistant phenotype which may be important in the resistance mechanism and could be used as a defined metabolite fingerprint to detect the resistant phenotype during early infection. Metabolomics also established that a different metabolite fingerprint also exists according to the plant susceptibility prior to infection. These metabolites may play an important role in the plant's resistance mechanism.

A. psidii expression of small secreted proteins (SSPs) could be detected 48 hours after inoculation in *Eucalyptus grandis* leaves. These encoded SSPs are important in the pathogenesis of *A. psidii.* In particular, three of these SSPs were highly expressed in the susceptible plant leaves, indicating their importance for host colonisation in the susceptible phenotype and showing that they could be used for detecting early infection in susceptible hosts.

Gene expression in the host, *E. grandis*, was analysed prior to- and at 48 hours after *A. psidii* inoculation. The host response to infection differed depending on resistance profile. Differentially expressed genes included genes encoding known disease resistance proteins, as well as other pathways associated with disease resistance including secondary metabolic and phenylpropanoid metabolic pathways and protein phosphorylation. Gene expression markers for plant susceptibility in uninfected cells were also identified.

Together, the results of this study highlight that both metabolomics and transcriptomics reveal chemical and genetic cues that can be used as molecular fingerprints to detect early infection and the corresponding phenotype of the host. In addition, molecular fingerprints were identified that characterise the phenotype prior to infection, which may be an important tool to identify resistant plants in the field or for plant breeding. Future work could further characterise the genes and molecules identified in this study to elucidate the molecular mechanisms of host resistance and *A. psidii* pathogenesis.

2. Introduction

The pathogen *Austropuccinia psidii*, also known as myrtle rust, is a significant threat to Australian ecosystems as it may lead to the decline, and potential loss, of many Myrtaceae species. To date, 382 native Australian species are known to be vulnerable to *A. psidii* infection. Within each of these species, however, there is a range of responses to infection from resistant to highly susceptible due to variations in genotypes between individuals (Makinson et al. 2020). In Australia, 43 species are severely affected and therefore require focused and timely research to identify possible resistant germplasm and/or populations (Makinson et al., 2020). Various attributes have been implicated in pre-formed resistance to *A. psidii* infection in Australian Myrtaceae species. These include cuticular waxes (Santos et al. 2019) and terpene composition in *Eucalyptus* sp. (Yong et al., 2019). Recent transcriptomics experiments have also highlighted differential expression of genes of interest in resistant and susceptible individuals prior to inoculation (Hsieh et al., 2018; S. A. Santos et al., 2020; Tobias et al., 2018). Early gene expression in resistant plants indicates early recognition of the pathogen within 24 hours post infection and activation of defence responses in resistant individuals of *Eucalyptus grandis* and *Syzygium luehmannii* compared with susceptible individuals (Tobias et al., 2018; dos Santos et al., 2019). Early gene expression in the eukaryotic plant pathogens, such as rust, includes genes encoding small secreted proteins (SSPs) that are produced to facilitate pathogenesis.

Previous studies are only just beginning to inform our understanding of the mechanisms that have evolved within the Myrtaceae that enable certain individuals/species to resist *A. psidii.* In addition to the examples given above, the Myrtaceae are known to produce a wealth of novel metabolites (i.e. small bioactive chemicals) that may also be used in plant defenses against pathogens. 'Metabolomics' is the investigation of these small chemicals that are present within a biological sample at a given time. These chemicals can include substrates, intermediates and products of metabolic pathways, as well as signalling molecules, hormones, and secondary metabolites. By adjusting the extraction method for isolating metabolites and the instrumentation used for separation and detection, a structurally different subset of metabolites may be analysed. Targeted metabolomics is the analysis of a set of known metabolites, such as *Eucalyptus* sp. terpenes (Yong et al., 2019), while untargeted metabolomics investigates both known and unknown metabolites. Recently, untargeted metabolomics has shown promise in identifying resistance biomarkers present in a broad range of human and plant diseases, including those found early in *Phytophthora* infection of tomatoes (Garcia et al., 2018), drought resistance in potatoes (Sprenger et al., 2018), and *Plasmopara viticola* defence response in a resistant grape variety (Chitarrini et al., 2017). The broad range of applications of metabolomics in assessing levels and mechanisms of disease resistance through identification of novel biomarkers demonstrates that this methodology could be applied to monitoring resistance to *A. psidii* during the early stages of infection.

3. Aim

The overarching aim in the PBSF023 project is to identify *A. psidii*-specific chemical cues, both metabolites and nucleic acids, produced early in the interaction between pathogen and host, prior to visible disease, and investigate how these cues correlate with pathogenesis.

More specifically, the objectives, listed in the proposal, are:

1. Identification of *A. psidii*-specific chemicals enabling novel, sensitive screening tests
2. Identification of fast-evolving proteins in *A. psidii* for use in future population and evolutionary studies
3. Data on plant traits associated with disease resistance
4. Scoping of new resistance pathways and control options for future research

4. Methods/Process

4.1 Untargeted metabolomics analysis of *Melaleuca quinquenervia* during early intracellular *Austropuccinia psidii* infection

4.1.1 *Austropuccinia psidii* inoculation of *Melaleuca quinquenervia*: A total of nine *Melaleuca quinquenervia* plants of known myrtle rust response phenotype (three each of resistant, hypersensitive and susceptible), were chosen for analysis. Each plant was inoculated with *A. psidii* at the Queensland Department of Agriculture and Fisheries by Dr. Louise Shuey. Samples of leaves were collected at 0 hours, 24 hours, 48 hours and 5 days after inoculation. Six replicates of young leaves from fresh growth on each plant were collected at each time point, weighed and stored at -80 C prior to analysis.

4.1.2 Leaf metabolite extraction: The frozen leaf samples were weighed and immediately ground into powder by bead-bashing in 200 µl extraction solvent (4:4:2 methanol:acetonitrile:deionized water) twice for 30 seconds each. The leaf:solvent mixture was then added with 300 µl additional extraction solvent, and then vortexed to mix for 10 seconds. The leaf:solvent mixture was then sonicated in icy water for 25 min followed by centrifugation at 12,000 rpm for 10 min. The supernatants were collected while avoiding undissolved particles, and filtered through a 0.22 µm syringe filter. The resulting metabolite extracts were stored at -80 °C until metabolite profiling was performed.

4.1.3 Ultra Performance Liquid Chromatography High Definition Mass Spectrometry with Ion Mobility (UPLC HDMS^{E}) analysis: Leaf extracts were analysed on a Waters Acquity I-Class UPLC system and a Waters Synapt G2-Si HDMS with a Waters UniSpray Ionisation source. The metabolites were separated on a Waters ACQUITY UPLC HSS T3 1.8µm 100 x 2.1mm Column at 35° C. The injection volume was 2 µL. The mobile phases were A (Water + 0.1% Formic Acid) and B (Acetonitrile + 0.1% Formic Acid). The chromatographic flow rate was 0.5 mL/min with a 9 min gradient, with mobile phase A held at 99% for 1 min, decreased to 85% over 1 min, decreased to 50% over 2 mins, decreased to 5% over 2 min and increased to 99% over 2 mins. Leucine Enkephalin Lockspray solution (Waters, 1ng/mL) was used as a standard.

Data acquisition was performed with ion mobility separation followed by mass fragmentation and high resolution mass analysis. The mass range of metabolites acquired was 50 - 1200 m/z, the scan time was 0.2 seconds and the elevated energy transfer collision voltage was 20 - 50 eV. For this experiment, the instrument was run in positive ionisation mode with the following settings: Capillary: 0.5 kV, source temperature: 120 °C, sampling cone: 30 V, source offset: 80 V, desolvation temperature: 500 °C, desolvation gas flow: 800 L/Hr, cone gas flow: 20 L/Hr.

4.1.4 Statistical and molecular analysis: The compound measurement table containing the feature spectral information and peak intensity of all samples (including quality-control samples and solvent blanks) was exported from the Progenesis QI for metabolomics (version 2.4) and further analyses were performed using R (version 3.6.3). The data table was scaled by leaf weight and low intensity peaks (intensity < 500) were removed after blank subtraction. The data table was then imported into MetaboAnalystR (version 3.0.3). To improve the power of downstream statistical testing, features that are missing in > 50% samples and near-constant across all samples were filtered based on interquartile range. Additionally, outliers were removed after visual identification using principal component analysis (PCA). The data table was log-transformed before two-way analysis of variance (ANOVA) with multiple comparisons correction (false discovery rate). The significant features with $p < 0.05$ were selected for further clustering analyses. Hierarchical clustering and k-means clustering were performed using R (version 3.6.3). The log-transformed data table was also used for permutational multivariate analysis of variance (PERMANOVA) with the pairwise Adonis wrapper (https://github.com/pmartinezarbizu/pairwiseAdonis) based on the 'adonis' function of the vegan package (version 2.5-6) on R. Progenesis QI was used to structurally classify the significant metabolites identified by the two-way ANOVA and clustering analyses. Based on the identities provided by the Chemical Entities of Biological Interest (ChEBI), metabolite enrichment analysis was performed on the MetaboAnalyst web platform (https://www.metaboanalyst.ca/).

4.2 Transcriptomic and small RNA analysis of *Eucalyptus grandis* during early intracellular *Austropuccinia psidii* infection

4.2.1 *Austropuccinia psidii* inoculation of *Eucalyptus grandis* and phenotype assessment: Forty *Eucalyptus grandis* seeds were germinated and grown at the Western Sydney University (WSU) in controlled growth chambers with 16 h light/8 h dark cycle at 25˚C, 70% relative humidity and 500 µmol m^{-2} s^{-1} light intensity. Cuttings of the plants consisting of 4-6 juvenile leaves were taken and placed in water agar (1%) within 50 ml tubes or 500 ml jars, depending on the size of the cutting. Infection with *A. psidii* spores was performed at the University of Sydney Plant Breeding Institute using the following protocol. Spores were collected from heavily infected *Syzygium jambos* plants by submerging the infected leaves in isopar oil. The inoculum was sprayed onto the leaf cuttings of *E. grandis* and *S. jambos* control while maintained within the tubes or jars. The *A. psidii* spores were allowed to settle for one minute before being capped and transported to WSU. The cuttings were then sprayed with water and left in the dark for 12 hours inside the WSU controlled growth chamber to ensure high relative humidity to assist with spore germination and infection. Infected cuttings were then maintained in a normal light cycle.

For phenotyping of each of the cuttings, leaves were examined for *A. psidii* disease symptoms, 14 days following inoculation and scored as either resistant, susceptible, or hypersensitive.

4.2.2 Leaf RNA extraction and sequencing: Leaf tissue belonging to the resistant and susceptible *E. grandis* myrtle rust phenotypes was collected at 0 hours and 48 hours after *A. psidii* or mock inoculation. For each treatment, three biological replicates were collected and snap frozen in liquid N_2 before storing at -80^{o}C. RNA was then extracted from three replicates for each treatment described above using the Bioline Plant II RNA extraction kit (Bioline, Sydney, Australia) according to the manufacturer's guidelines. Poly-A RNA libraries were then prepared and sequenced by GENEWIZ (Suzhou, China) using an Illumina HiSeq platform and 150bp paired-end configuration.

4.2.3 RNA-seq data analysis of *Austropuccinia psidii* small secreted protein effectors in resistant and susceptible *Eucalyptus grandis:* The RNA-seq data was trimmed to remove adapters and low-quality sequences, and then aligned to the primary transcripts of the *Austropuccinia psidii* reference genome (Tobias et al., 2020). For mapping to the genome of *A. psidii*, we used both a default mapping stringency that included a similarity fraction of 0.8, a minimum length fraction of 0.8, and a lower mapping stringency of 0.7 for both the similarity fraction and minimum length fraction, respectively. The maximum number of hits per read was set to 10. The uniquely mapped RNA reads for each transcript was determined and normalized to reads per kilobase of transcript, per million mapped reads using the RNA-seq analysis function in CLC. Only transcripts with an average normalised count of more than 10 in at least one treatment including the resistant and susceptible inoculated only were considered for further analysis. The Bioconductor package DESeq2 v1.26.0 (Love et al. 2014) was used to normalise raw transcript counts. A principal component analysis (PCA) of Log2 transformed RNA-sequencing data was performed using the mixomics package in R (Rohart et al., 2017). The relative fold change for each transcript was determined against the average expression across all replicate samples of the resistant and susceptible treatments and combined with *A. psidii* small secreted protein (SSP) effectors (Tobias et al., 2020) to profile expression patterns in these two genotypes.

4.2.4 RNA-seq data analysis in *Eucalyptus grandis:* The trimmed RNA-seq data was then aligned to the primary transcripts of the *Eucalyptus grandis* V2 genome using CLC Genomics Workbench 12 (Myburg et al. 2014). For alignments, a similarity fraction of 0.8 and minimum length fraction of 0.8 were used while all other parameters were set to default. The maximum number of hits per read was set to 10 and the uniquely mapped RNA reads for each transcript was determined and normalized to reads per kilobase of transcript, per million mapped reads using the RNA-seq analysis function in CLC. Only transcripts with an average normalised count of more than 10 in at least one treatment and control were considered for further analysis. The Bioconductor package DESeq2 v1.26.0 (Love et al. 2014) was used to normalise raw transcript counts. A principal component analysis (PCA) of Log2 transformed RNA-sequencing data was performed on samples from all treatments and control using the mixomics package in r (Rohart et al., 2017). Statistically significant differentially expressed *E. grandis* genes in the resistant and susceptible genotypes were determined using the Benjamin-Hochberg test for multiple testing with a false discovery rate (FDR) to control for false positives and negatives. Only transcripts with a fold change of $-1 < Log_2FC > 1$ compared to the control and FDR-corrected $p<0.05$ were kept for further analysis. The significantly differentially expressed genes in resistant and susceptible were visualised using R v4.0.0 (R Core Team, 2020).

4.2.5 Identification of genes and pathways driving differences between inoculated resistant and susceptible *E. grandis* genotypes: PCA of the Log_2 transformed RNA-seq data belonging to resistant and susceptible *E. grandis* genotypes inoculated with *A. psidii* at 48 hai was performed using the mixomics package in R (Rohart et al., 2017). The coordinates from PC1 and PC2 were graphed using Excel. Analysis of the ellipse identified the radius at 0.03 thereby establishing the cut-off for gene importance. All genes were ranked by radius and those with a radius of more than 0.03 on the positive side of the x-axis were those with increased abundance in the inoculated resistant treatments, while those on the negative side of the x-axis were those with increased abundance in the inoculated susceptible treatments. The list of genes driving the resistant and susceptible inoculated treatments were combined with the Log_2 transformed data and the relative fold change for each transcript determined against the average expression across all replicate samples of the resistant and susceptible treatments. A heatmap of the relative fold change for each transcript was generated using Morpheus (https://software.broadinstitute.org/morpheus/). Gene Ontology enrichment analysis of the gene lists driving the difference between the resistant and susceptible genotypes was performed using the Plant Transcriptional Regulatory Map (http://plantregmap.gao-lab.org/). A threshold p-value of 0.05 was used to identify enriched GO terms in the categories biological process, molecular function, and cellular component. The top GO terms with an adjusted p-value < 1 x10E-1 were visualised using gprofiler in R v4.0.0 (R Core Team, 2020).

4.2.6 Identification of genes and pathways driving differences between hypersensitive, resistant and susceptible *E. grandis* genotypes in the absence of pathogen challenge: A Partial Least Squares-Discriminant Analysis (PLS-DA) of the Log_2 transformed RNA-seq data belonging to hypersensitive, resistant and susceptible *E. grandis* control was performed using the mixomics package in R (Rohart et al., 2017). We noted that X-variate 1 separated the data of the resistant and hypersensitive from the susceptible and thus used the coordinates from x-variate 1 of the PLS-DA were used to identify genes differentiating the combined group of resistant and hypersensitive from the susceptible. A cut-off of 0.093 was used to obtain the top list of genes differentiating the resistant and hypersensitive group from the susceptible group. Those with a value higher that 0.093 were those with increased abundance in the susceptible control and those lower than -0.093 were those with increased abundance in the resistant and hypersensitive controls. We then noted that X-variate 2 separated the data of the resistant control from that of the hypersensitive control and used the coordinates from x-variate 2of the PLS-DA to identify genes differentiating these two groups. For the coordinates, a cut-off of 0.013 was used to identify the top genes differentiating the resistant group from the hypersensitive group. Those with a value higher that 0.013 were those with increased abundance in the resistant control and those with a value lower than -0.013 were those with increased abundance in the hypersensitive control. The lists of genes driving differences for each group (i.e. resistant and hypersensitive control group, susceptible control, resistant control and hypersensitive control) were combined with the Log_2 transformed count data and the relative fold change for each transcript determined against the average expression across all replicate samples of the hypersensitive, resistant and susceptible controls. A heatmap of the average relative fold change for each transcript in each control genotype was generated using Morpheus (https://software.broadinstitute.org/morpheus/).

5. Achievements, Impacts and Outcomes

5.1 Identification of *A. psidii* infection-specific chemicals enabling novel, sensitive screening tests

The differences in the metabolome between *Melaleuca quinquenervia* phenotypes, resistant, hypersensitive and susceptible, were investigated during the early stages of *A. psidii* infection. Leaves were collected prior to inoculation and 24 hours, 48 hours and 5 days after inoculation. Three plants belonging to each phenotype (resistant, hypersensitive and susceptible) were analysed and six leaves from each plant analysed at each time point (total 216 leaves) for improved statistical analysis. The metabolome of the leaves at each time point was analysed and a total of 11,276 metabolites were identified.

Table 1: The summary of pairwise permanova analysis

Comparison	Factors	Degree of freedom	R2	p-value
hypersensitive_vs_resistant	Phenotype	1	0.04669	0.001***
	Time	3	0.34827	0.001***
	Phenotype x Time	3	0.02854	0.015*
hypersensitive_vs_sensitive	Phenotype	1	0.02803	0.001***
	Time	3	0.37494	0.001***
	Phenotype x Time	3	0.02648	0.032*
resistant_vs_sensitive	Phenotype	1	0.05663	0.001***
	Time	3	0.36828	0.001***
	Phenotype x Time	3	0.03643	0.002**

The leaf metabolomes were analysed using permutational multivariate analysis of variance (PERMANOVA), which suggested that phenotypes of the plants, time post-inoculation and the interaction of these two factors are all significant factors affecting the overall leaf metabolite profiles (Table 1). Amongst these metabolites, we identified 2,384 that changed significantly in their abundance ($p < 0.05$; Figure 1). The abundance of 978 and 2,157 metabolites were influenced by the phenotype of the plants and the time after inoculation, respectively, whereas 909 metabolites were influenced by the interaction of both the phenotype and the time. As time appears to be the most influential factor, this result reveals that *A. psidii*-infection triggered major metabolite responses on *M. quinquenervia* leaves in all phenotypes. We further examined the temporal changes of these significant metabolites identified with analysis of variance (ANOVA). Through hierarchical clustering analysis, we found that the majority of these significant metabolites were enriched shortly after *A. psidii*-infection (24 or 48 hour after inoculation) and gradually reduced at 5 days after inoculation (Figure 2).

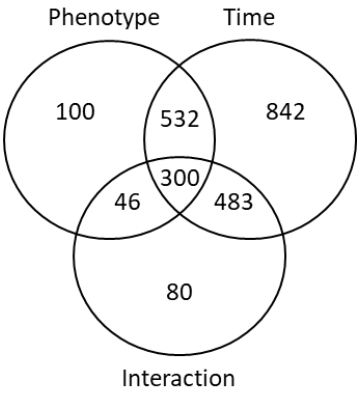


Figure 1: *Austropuccinia psidii* triggered differential metabolite responses on leaves of different phenotypes. This Venn diagram shows the number of metabolites that are significantly changed in terms of abundance under the influence of phenotype, time post-inoculation and the interaction of both phenotype and time. Two-way ANOVA with multiple-comparison adjustment using the false discovery rate approach was performed to identify the significant metabolites (adjusted-p < 0.05) influenced by each factor and their interaction.

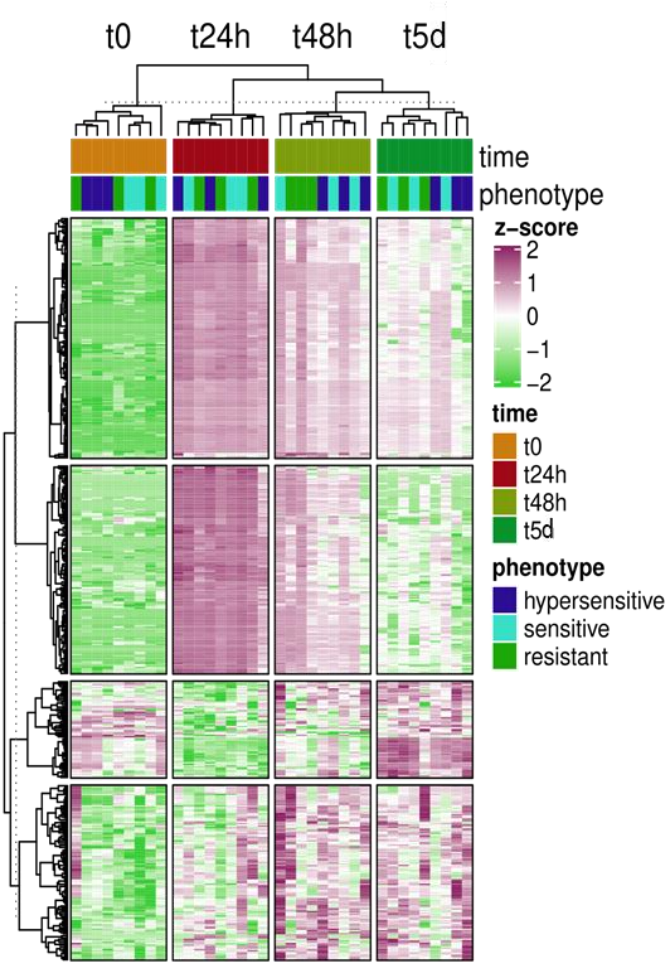
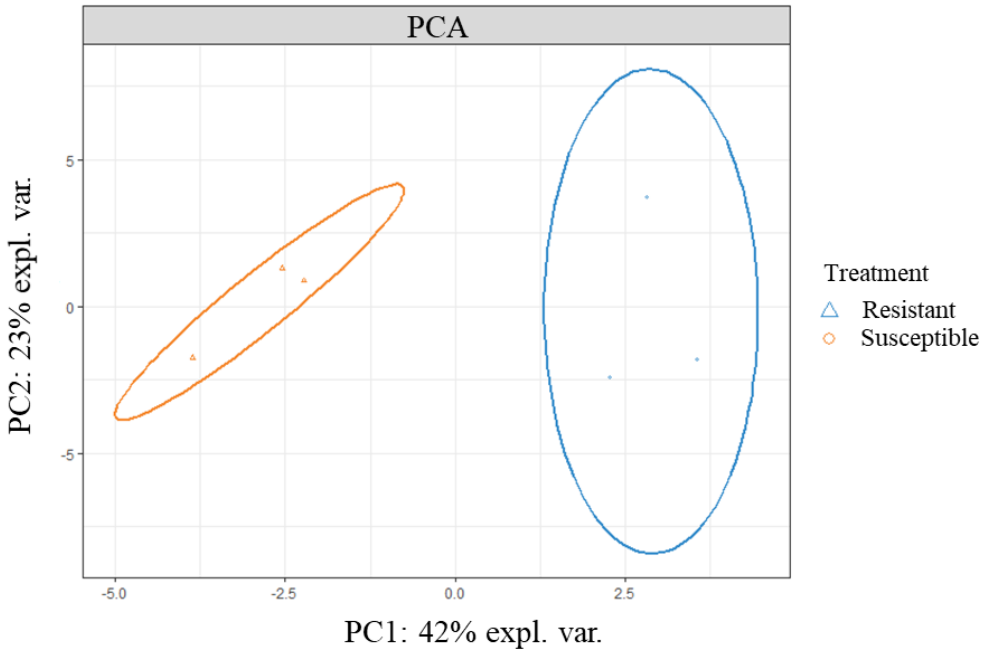


Figure 2: Most metabolites are significantly enriched in the 24 to 48 hours after *A. psidii* inoculation. The heatmap shows the scaled peak intensity of the top-500 significant metabolites (ranked by adjusted p calculated with 2-way ANOVA). The heatmap is clustered by Euclidean distance and clustered with Ward's minimum variance method for both the columns and rows. Purple and green indicate the enrichment and reduction, respectively, of each significant molecular feature (rows) of each sample (columns).

5.2 Identification of fast-evolving proteins in *A. psidii* for use in future population and evolutionary studies

Small secreted proteins ('effectors') are used by pathogens to overcome plant immunity to cause disease. As effectors are fast-evolving proteins, and typically genus- or species- specific, they could be used as novel targets to identify if a particular disease is present in a plant tissue or in a given environment. To determine if expression of genes encoding effector proteins could be used as an early detection platform, we performed RNA-sequencing on the myrtle rust pathogen *A. psidii* during infection of resistant and susceptible *E. grandis* genotypes at 48 hours after inoculation (hai) looking for these small, secreted proteins. Principal component analysis (PCA) of RNA-sequencing samples revealed that 42% of the variation was explained by PC1 separating the samples by treatment of resistant and susceptible (Figure 3A). We were able to detect the expression of four SSP effectors in the tissue of *E. grandis* resistant and susceptible genotypes at 48 hai: *APSI_H017.8250*, *APSI_P004.3021*, *APSI_P008.16995* and *APSI_P029.3884*. Due to the absence of *A. psidii* control RNA-sequencing data, we profiled the relative Log2 transformed expression patterns of the four detectable *A. psidii* SSP effector genes in the resistant and susceptible *E. grandis* tissue and identified that *APSI_H017.8250* displayed similar relative expression in the susceptible and resistant genotypes (Figure 3B). Furthermore, *APSI_P004.3021*, *APSI_P008.16995* and *APSI_P029.3884* displayed high relative expression in the susceptible compared to within the resistant genotype suggesting that these may be important for host colonisation in the susceptible genotype (Figure 3B). As we were able to detect these gene transcripts when *A. psidii* only accounted for of 1% of the RNAseq reads, this would indicate that using gene expression of any of these four effector proteins would be a highly sensitive, highly specific means of detecting this specific pathogen at a very early stage of pathogenesis.

(A)



(B)

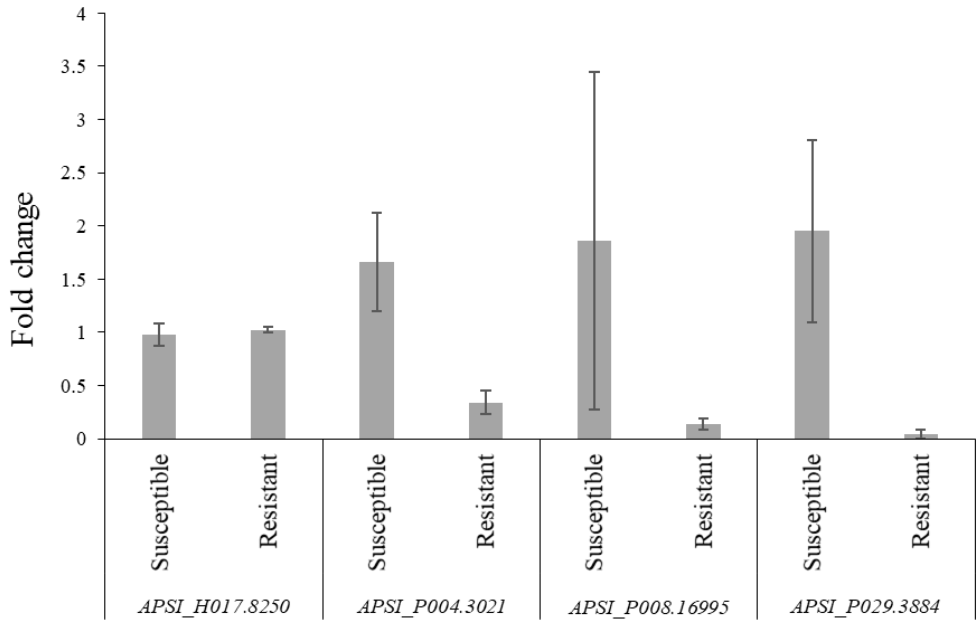


Figure 3: *Austropuccinia psidii* small secreted protein (SSP) effectors are more highly expressed in the susceptible *Eucalyptus grandis* genotypes 48 hours after inoculation (hai). (A) Principal component analysis of RNA-sequencing samples derived from *A. psidii* in resistant and susceptible *E. grandis* genotypes 48 hai. (B) Expression profiles of log2-transformed data of four *A. psidii* small secreted effector genes at 48 hai in *E. grandis* susceptible and resistant genotypes. The y-axis represents the relative fold change and the x-axis represents the treatments, susceptible and resistant for each SSP. The relative fold change is the average of the ratio of transcript abundance for a treatment relative to the average across all treatments. Error bars represent standard error of the mean across replicates for each treatment of resistant and susceptible, respectively.

5.3 Data on plant traits associated with disease resistance

We further analysed metabolomics profile of resistant, hypersensitive and susceptible phenotypes of *M. quinquenervia* during *A. psidii* infection. Although hierarchical clustering did not clearly separate metabolite profiles of leaves from different phenotypes, both PERMANOVA and ANOVA results support that there is significant variation in the leaf metabolite profiles from *M. quinquenervia* of different resistance phenotypes. In an effort to identify the metabolites that distinguish the *M. quinquenervia* plants' susceptibility towards *A. psidii*, enrichment analysis was performed on the metabolite responses that are significantly influenced by phenotype, irrespective of time (ANOVA adjusted-$p < 0.05$). Organoheterocyclic compounds, including terpenoids and flavonoids appeared to be the most enriched metabolite sets that set apart these disease resistance phenotypes (Figure 4).

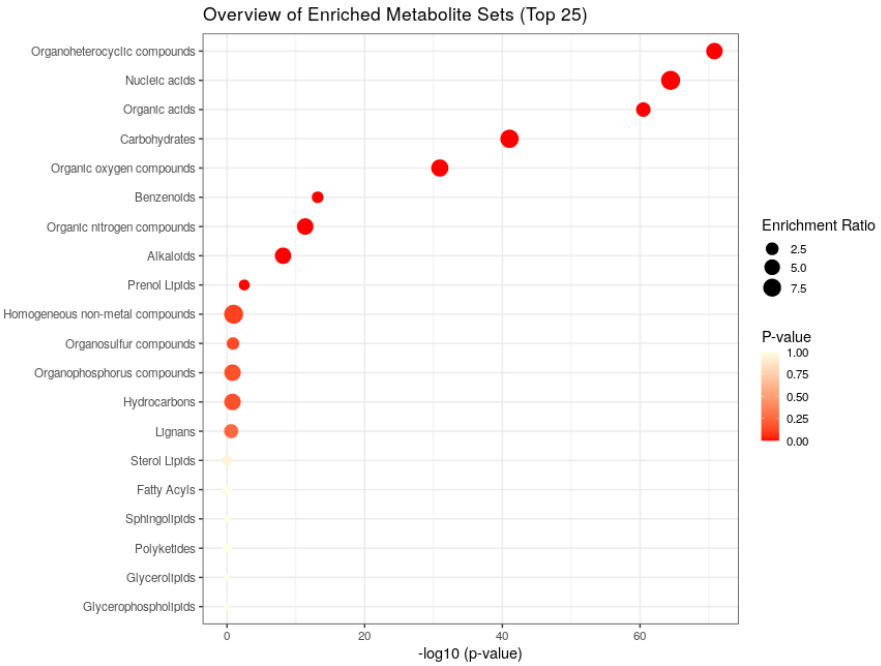
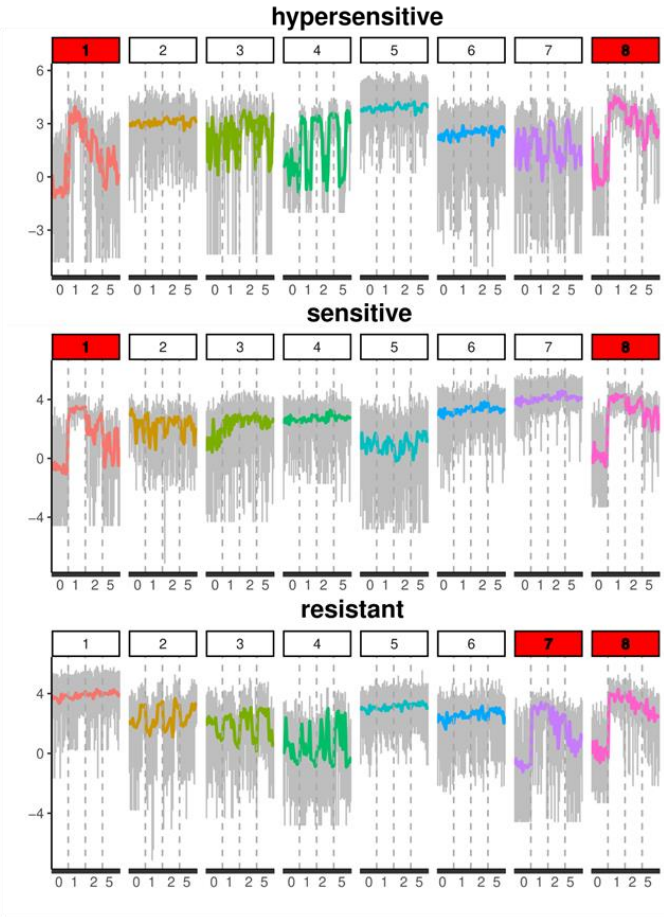


Figure 4: Organoheterocyclic compounds are the predominant group of metabolites that differ between different *M. quinquenervia* disease phenotypes. The graph shows the metabolite classification of the significant metabolites influenced by the phenotypes (adjusted $p < 0.05$, two-way ANOVA with FDR adjustment). The x-axis and colour of the points indicate the significant value of the enrichment set, whereby the size of the points indicate the enrichment ratio generated by metabolite enrichment analysis.

Metabolites that are induced at 24 and 48 hai potentially serve an important role in the defence against *A. psidii* infection. Therefore, metabolites that differ between different phenotypes could have important implications on the outcome of the disease and could, therefore, be used as a biomarker for *A. psidii* diagnostics. Therefore, we further separated metabolic fingerprints of pathogen success by using a *k*-means clustering approach for each phenotype individually (Figure 5A). For each phenotype, we identified two clusters of inducible metabolites. We then compared their identities and found that the majority of these early induced metabolites are similar between different phenotypes (Figure 5B).

(A)



(B)

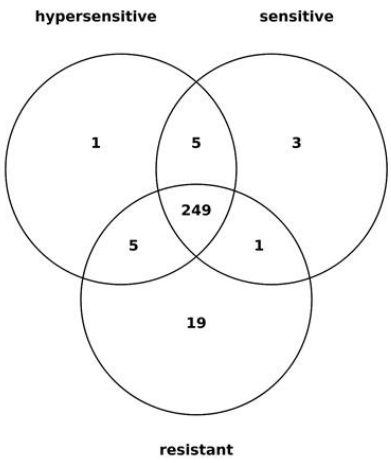


Figure 5: *k*-means clustering separates metabolites into groups with different temporal shifts in abundance. Metabolite responses at 24 to 48 hours after inoculation (hai) are similar between different phenotypes. (A) The plots show the average temporal fluctuation in abundance of metabolites from different *k*-mean clusters. The line in each plot indicates the average log(peak intensity) of the metabolites (y-axis) across the time-series (x-axis; number of days post-inoculation). The metabolites in the key clusters highlighted in red have an increased abundance at the 24 hour and 48 hour after infection time points. **(B)** The Venn diagram shows the number of metabolites that are enriched at 24 and 48 hai (red clusters) in each phenotype.

Metabolite enrichment analysis was performed on 24 and 48 hai samples to establish the structural identity of metabolites that are unique to each phenotype. The majority of these inducible metabolites are terpenoids, fatty alcohols, and dipeptides (Figure 6). Nineteen metabolites were found to be uniquely enriched amongst the resistant plants, and the identities of 9 metabolites were predicted based on mass spectrometry fragmentation patterns (Table 2). While many of these predicted structures have unknown roles in disease resistance, (2R,6x)-7-Methyl-3-methylene-1,2,6,7-octanetetrol 2-glucoside (367.1960321 m/z) is a fatty acyl glycoside and related molecules are reported to have antimicrobial activity, enhance disease protection, and enhance plant growth.

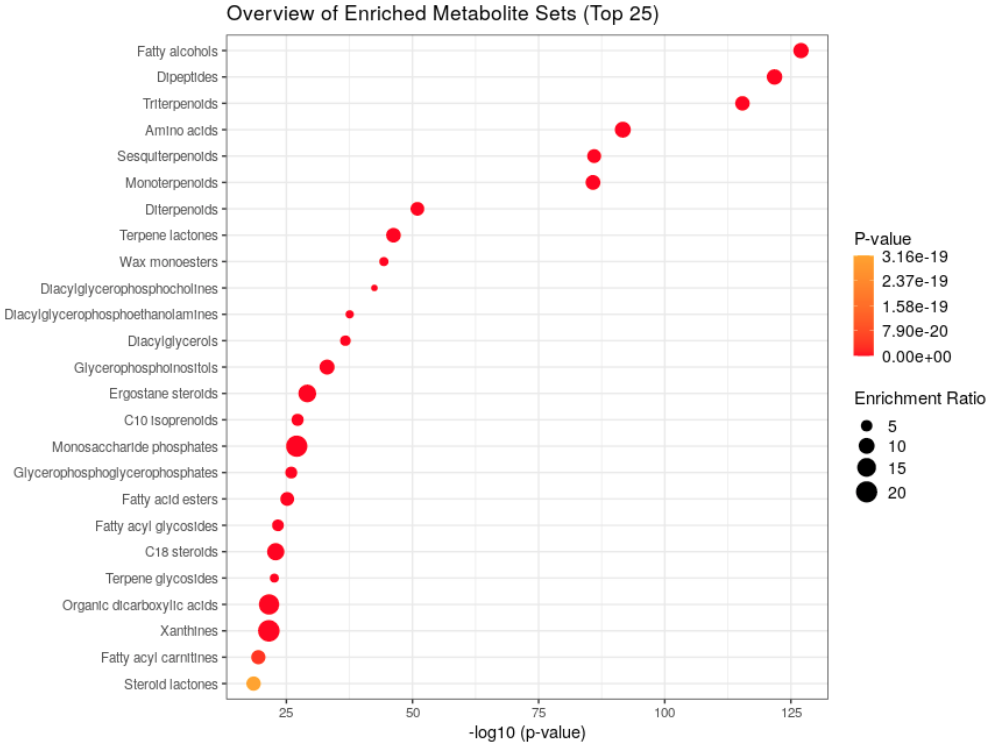


Figure 6: Enrichment of fatty alcohols, dipeptides and triterpenoids are the major metabolite that distinguish between phenotypes in *Melaleuca quinquinervia* leaves 24 and 48 hours after *Austropuccinia psidii* inoculation. The graph shows the classification of the metabolites that are enriched at 24 to 48 hours after inoculation and are specific to the phenotype (the key clusters identified by *k*-means clustering). The x-axis and colour of the points indicate the significant value of the enrichment set, whereby the size of the points indicate the enrichment ratio generated by metabolite enrichment analysis.

Table 2: Structural prediction of the inducible metabolites that are unique to resistant plants 24 to 48 hours after inoculation with *Austropuccinia psidii*

Compound	Identification score	Description	m/z	Retention time (min)
2.86_366.1888n	38.7	(2R,6x)-7-Methyl-3-methylene-1,2,6,7-octanetetrol 2-glucoside	367.1960321	2.863466667
5.03_624.5280m/z	35.8	N-icosanoyl-15-methylhexadecasphing-4-enine	624.5280082	5.0305
6.18_791.4806m/z	37.4	1-oleoyl-2-[(3E)-hexadecenoyl]-sn-glycero-3-phosphoglycerol	791.4805652	6.1779
7.13_593.2426m/z	33.5	zafirlukast	593.2425551	7.125733333
7.13_593.3270m/z	35.8	amitriptyline	593.3269639	7.125733333
7.16_636.5160m/z	38.4	methyl 3-hydroxypalmitate	636.5160247	7.16145
7.30_656.2849n	37.7	coproporphyrinogen III(4-)	657.2921888	7.297133333
7.57_623.3873m/z	32.4	lutein 5,6-epoxide	623.3873252	7.567783333
7.85_609.8435m/z	23.8	cis-3-(2,2-dibromovinyl)-2,2-dimethylcyclopropanecarboxylic acid	609.8434753	7.85275

As a complement to the work in *M. quinquenervia* we analysed RNA-sequencing of resistant and susceptible *E. grandis* genotypes during challenge with *A. psidii.* Similar to the metabolomic work, we used the *E. grandis* system also at 48 hours post leaf inoculation versus mock-inoculated leaves to identify plant traits associated with disease resistance at the early stage of plant-pathogen interaction. We also included a hypersensitive *E. grandis* genotype that was mock-inoculated with spore carrier suspension for comparative analysis with resistant and susceptible control plantlets with the aim of identifying genotypic differences before pathogen challenge that may be driving differences in resistance. PCA of RNA-sequencing data revealed that 43% of the variation was explained in PC1 separating the samples into control and *A. psidii* inoculated for the resistant and susceptible *E. grandis* genotypes (Figure 7A). PC1 also explained the largest variation for RNA-sequencing data differentiating the control hypersensitive and control resistant genotypes from the control susceptible genotype, and the inoculated resistant from the inoculated susceptible (Figure 7A). We then performed differential expression analysis between resistant and susceptible *E. grandis* during *A. psidii* challenge at 48 hai relative to the mock-inoculation to identify the number of significantly up-regulated and down-regulated gene set. In the resistant genotype inoculation with the pathogen led to the transcription of 1,766 genes being induced, and 2,503 genes being repressed whereas lower differential regulation was observed in the susceptible lines of *E. grandis* with 785 genes up-regulated and 524 genes down-

regulated (Figure 7B). Therefore, susceptible *A. psidii* hosts are less transcriptomically responsive to the presence of the pathogen.

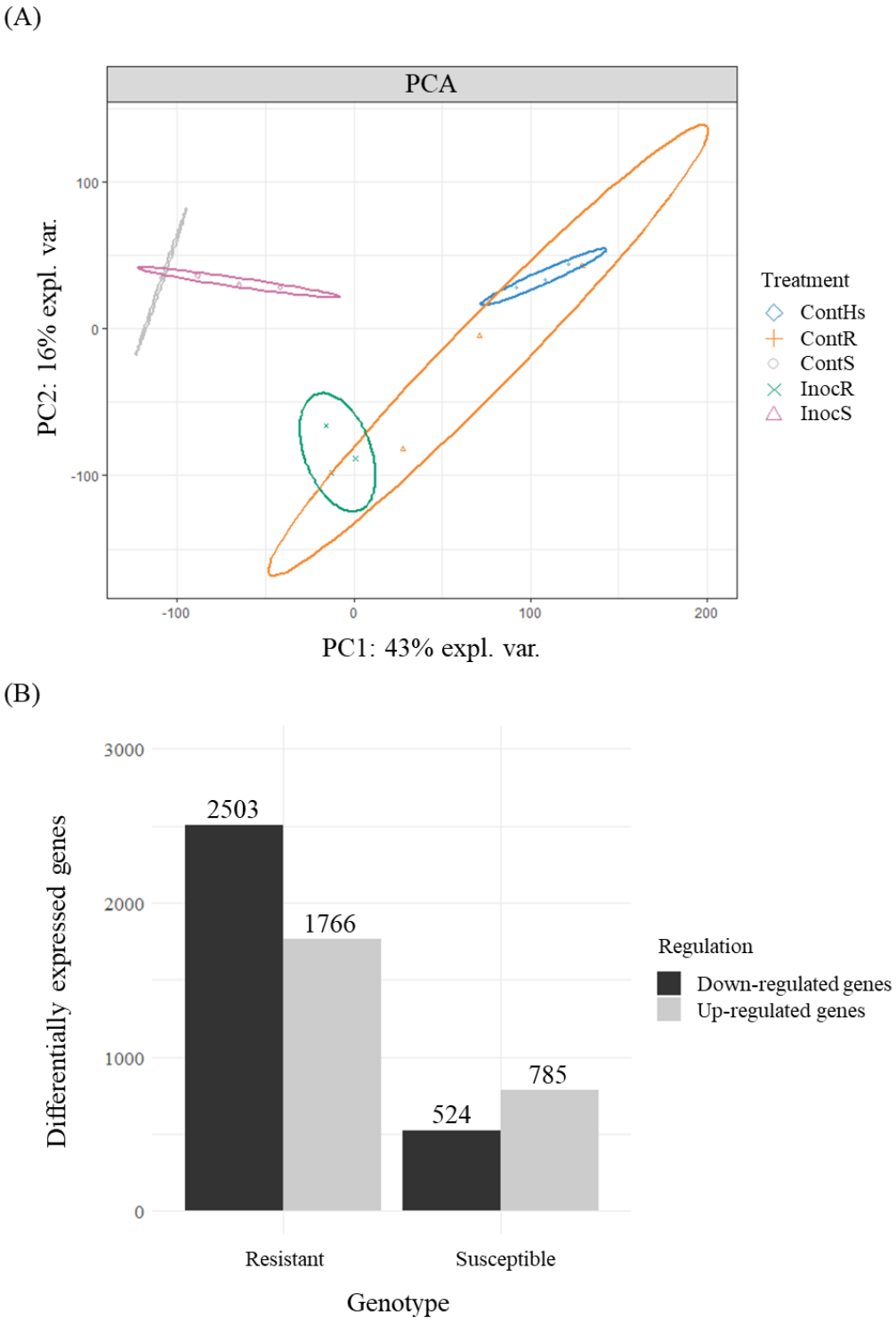
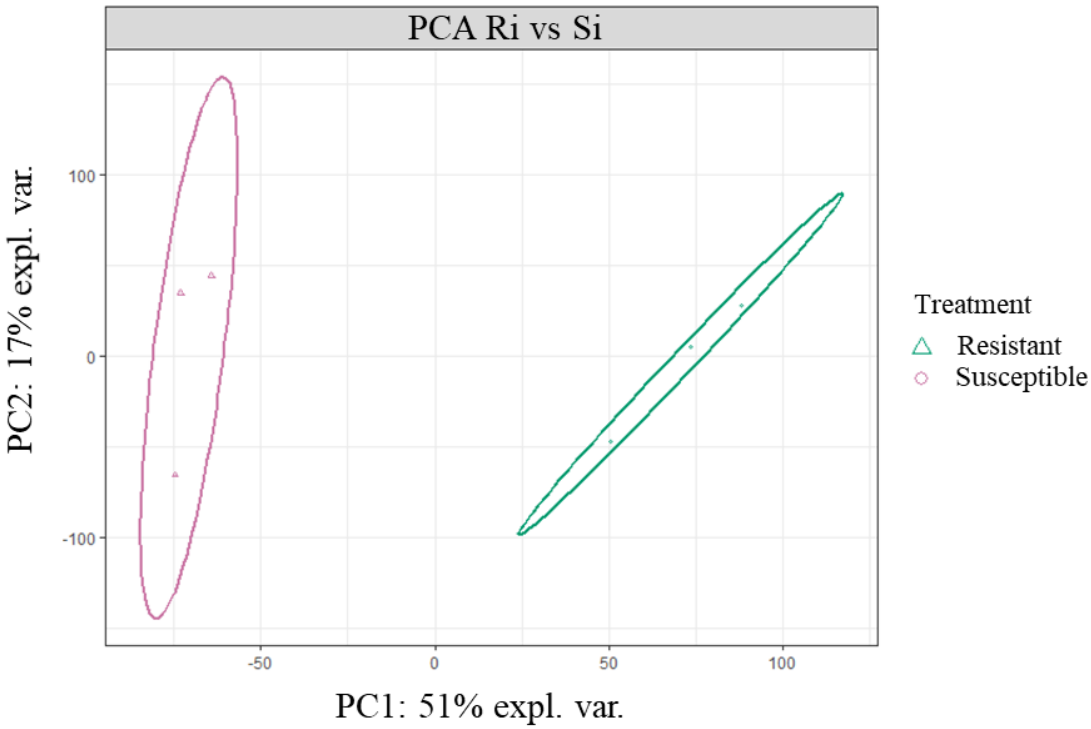


Figure 7: Genome-wide evaluation of RNA-sequencing data in resistant, susceptible, and hypersensitive *Eucalyptus grandis* genotypes infected with *Austropuccinia psidii* 48 hours after inoculation (hai). (A) Principal component analysis of RNA-sequencing samples derived from control (Cont) and inoculated (Inoc) resistant (R), susceptible (S) and hypersensitive *Eucalyptus grandis* genotypes inoculated with *A. psidii* at 48 hai. (B) Significantly differentially expressed genes between inoculated and control resistant and susceptible *Eucalyptus grandis* genotypes (-1 < Log2FC > 1; p-value 0.05).

We then sought to compare only the transcriptomic response of resistant and susceptible *E. grandis* genotypes inoculated with *A. psidii* to identify genes and pathways associated with disease resistance during early pathogen challenge. PCA of RNA-sequencing samples revealed that 51% of the variation was explained by PC1 separating the samples into resistant and susceptible at 48 hours post *A. psidii* inoculation (Figure 8A). From the PCA loadings, we identified 158 genes in the resistant genotype and 142 genes in the susceptible genotype that were driving difference between the two during *A. psidii* challenge (Figure 8B). Expression profiling of the resistance gene drivers revealed that of the 158 genes, 30 genes displayed very high expression in the resistant compared to the susceptible of which 11 genes have been previously linked to disease resistance including *FH interacting protein 1* (*Eucgr.C01931.1*), *20S proteosome beta subunit G1* (*Eucgr.H02300.1*) , *cysteine-rich receptor-like kinase* (*Eucgr.E03311.1*), *receptor-like protein* (*Eucgr.F00315.1*), *ankyrin-repeat family protein* (*Eucgr.C01885.1*), *terpene synthase* (*Eucgr.E03311.1*) and *histone deacetylase 5* (*Eucgr.L02064.1*; Figure 4A-i). Furthermore, 77 genes displayed a 2-3-fold higher expression in the resistant compared to the susceptible of which 16 have been previously linked to disease resistance including disease resistance proteins, cysteine-rich receptor-like kinase, Leucine-rich repeat (LRR) protein kinase and nitrate transporter 2.5 (Figure 9A-ii). Finally, 51 genes displayed similar expression patterns in the resistant and susceptible with these genes displaying marginally higher expression in the resistant and of these genes, 20 are linked to disease resistance functions (Figure 9A-iii). Profiling the expression of the 142 susceptible gene drivers revealed 9 genes that displayed very high expression in the susceptible compared to the resistant and of these, 3 genes were linked to disease resistance including one *LRR protein kinase* (*Eucgr.G01270.1*) and two *NB_ARC domain containing disease resistance proteins* (*Eucgr.F01653.1* and *Eucgr.J02621.1*; Figure 9B-i). Interestingly, a *Trimeric LpxA-like enzyme* (*Eucgr.L03047.1*) displayed opposite expression being lowly expressed in the susceptible and highly expressed in the resistant (Figure 9B-ii). In addition, 68 genes displayed 2-3-fold higher expression in the susceptible compared to the resistant and of these 10 genes have been linked to disease resistance such as disease resistance proteins, pathogenesis-related 1 protein and heat shock protein II (Figure 9B-iii). Finally, 64 genes displayed similar or marginally higher expression in the susceptible *E. grandis* genotype compared to the resistant genotype (Figure 9B-iv). Of these, 16 were linked to disease resistance functions.

(A)



(B)

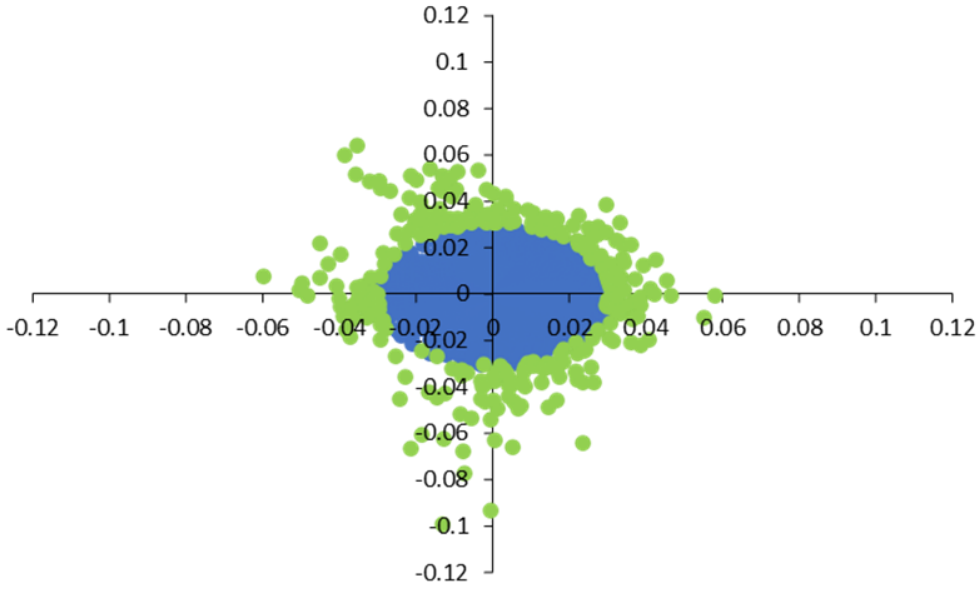


Figure 8: Identification of genes driving differences in resistant and susceptible *Eucalyptus grandis* genotypes 48 hours after inoculation (hai) with *Austropuccinia psidii* (A) Principal component analysis (PCA) of RNA-sequencing samples derived from inoculated resistant and susceptible *Eucalyptus grandis* genotypes inoculated with *A. psidii* at 48 hai. **(B)** The loadings of the genes identified from the PCA analysis. Green points are genes with the highest radius among all genes and are the most important in driving the separation between the resistant and susceptible. The remaining genes (blue) are the least important in driving the separation.

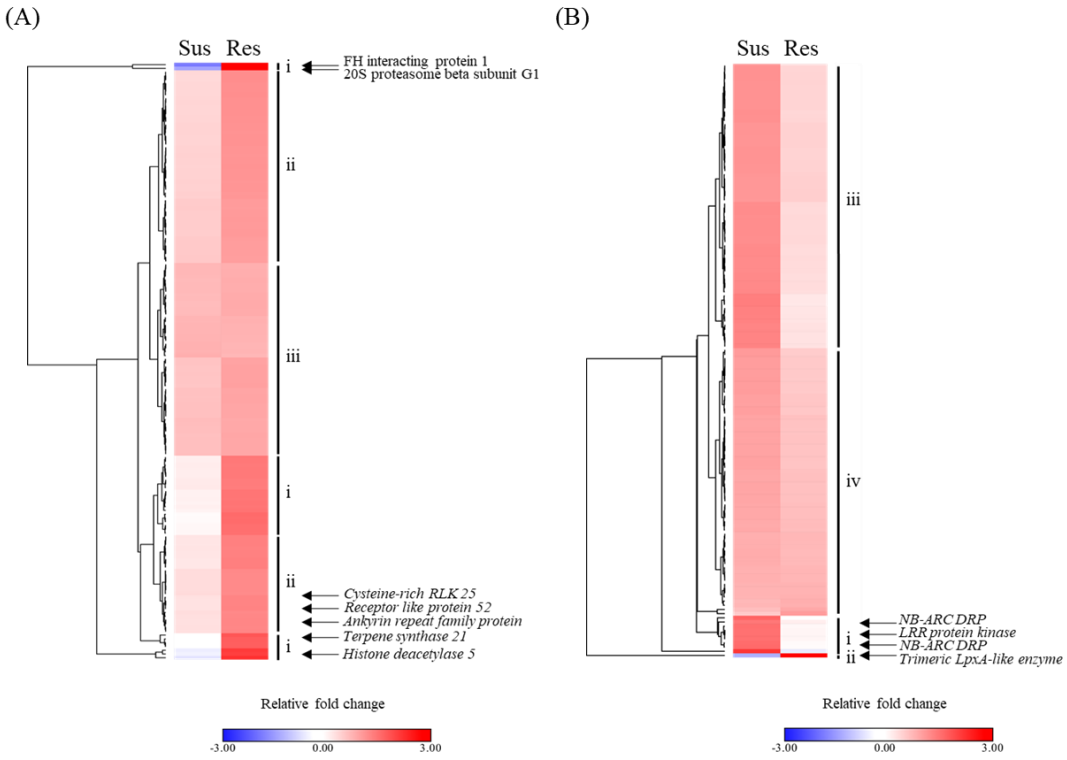


Figure 9: Expression patterns of genes driving differences between susceptible and resistant *Eucalyptus grandis* genotypes inoculated with *Austropuccinia psidii* at 48 hours after inoculation (hai). (A) Hierarchical clustering of the log2-transformed data of 158 genes driving the resistant *E. grandis* genotype during *A. psidii* challenge at 48 hai compared to the susceptible *E. grandis* genotype. All the data points shown are the ratio of transcript abundance in each treatment relative to the average across all treatments where up-regulated genes are presented as blue (low) and down-regulated genes are presented as red (high), respectively. The heatmap is annotated on the right-hand side with hierarchical clustering groups denoted i-iii.. i: susceptible genes displaying very high expression in the resistant compared to the susceptible; ii: genes displaying 2-3-fold higher expression in the resistant compared to susceptible; iii: genes displaying similar expression patterns between resistant and. (B) Hierarchical clustering of the log2-transformed data of 142 genes driving the susceptible *E. grandis* genotype during *A. psidii* challenge at 48 hai compared to the resistant *E. grandis* genotype. All the data points shown are the ratio of transcript abundance in each treatment relative to the average across all treatments where up-regulated genes are presented as red (high) and down-regulated genes are presented as blue (low), respectively. i: genes displaying very high expression in the susceptible compared to the resistant; ii: genes displaying very high expression in the resistant compared to the susceptible; iii: genes displaying 2-3-fold higher expression in the susceptible; iv: genes displaying similar expression patterns between resistant and susceptible. RLK: Receptor-like protein kinase, DRP: Disease resistance protein, LRR: Leucine-rich repeat.

Gene Ontology (GO) enrichment of the resistance gene drivers revealed that of the 158 genes, 127 had a GO annotation and of these 54 were enriched (p-value < 0.05). Of the enriched GO terms within the category biological process (BP), protein phosphorylation, secondary metabolic process, phenylpropanoid metabolic process and response to gibberellin were among the pathways observed that have been previously linked to plant disease resistance (Figure 10). Within the gene ontology category cellular compartment (CC), extracellular region, apoplast and proteosome core complex were among the enriched terms (Figure 10). The category molecular function (MF) contained the largest number of enriched terms among the gene ontology categories of which catalytic activity, small

molecule binding, ATP binding, oxidoreductase activity, transferase activity, protein serine/threonine kinase activity and aminoacyl-tRNA ligase activity were examples of the enriched terms in this category (Figure 10). GO enrichment of the 142 susceptible gene drivers revealed that 108 genes had GO annotations of which 39 were enriched at (p-value < 0.05). Of the terms enriched in the GO category BP, response to stress, defence response, systemic acquired resistance and hydrogen peroxide catabolic process were among the processes that have been reported for plant disease resistance (Figure 11). Within the GO category CC, the major compartments that were enriched included the extracellular region and cytosol of ribosome (Figure 11). For the MF category, enriched terms included for example ADP binding, lipid binding, oxidoreductase activity and carboxy-lyase activity (Figure 11).

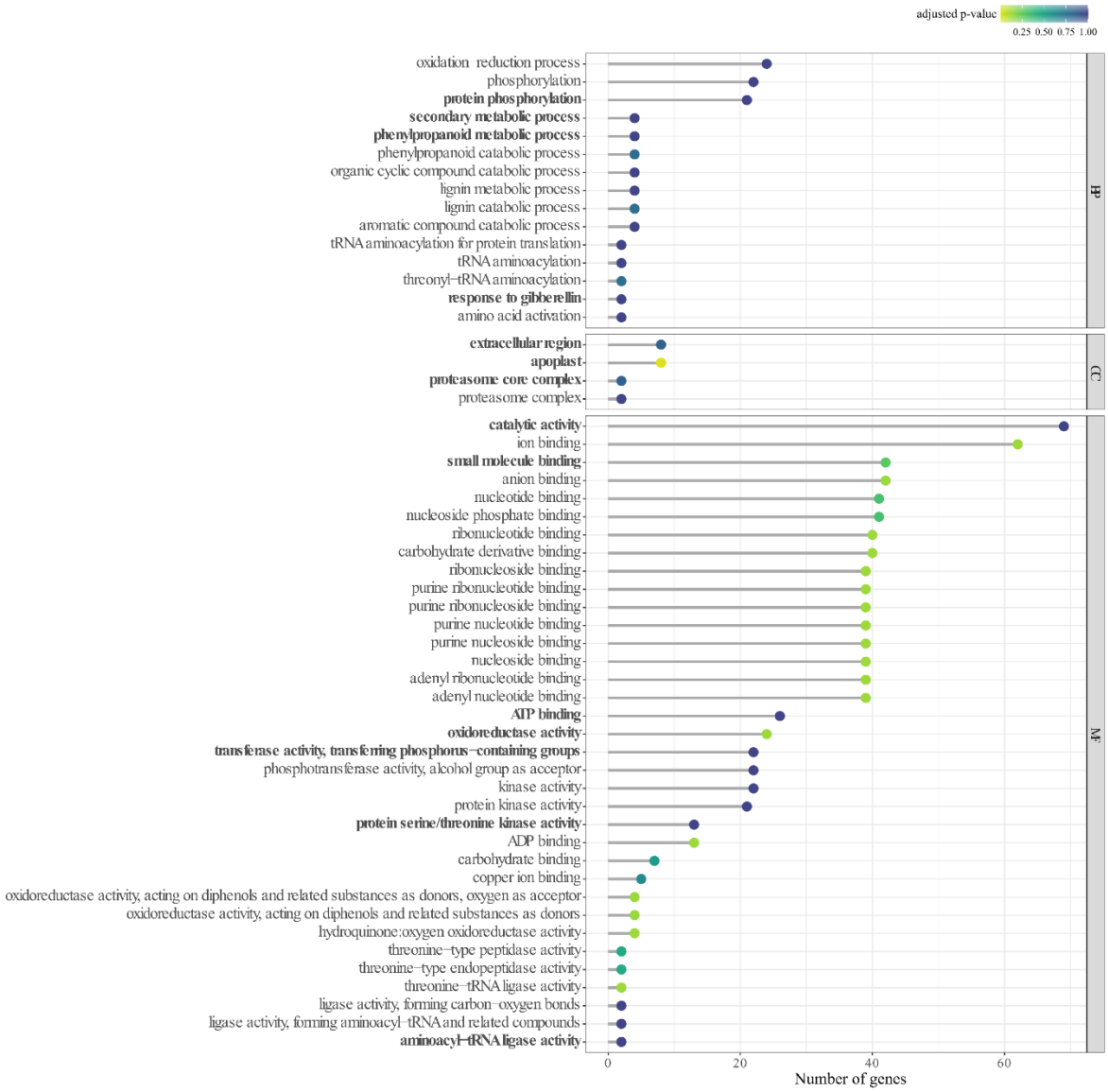


Figure 10: Disease resistance pathways are enriched in the resistant *E. grandis* genotype at 2 days after *A. psidii* challenge. Enriched Gene Ontology terms in the categories biological process (BP), cellular compartment (CC) and molecular function (MF) for the 158 genes identified as drivers of the resistance genotype during early pathogen challenge at 48 hours after inoculation (hai). The GO terms are shown on the y-axes with interesting terms bolded. The number of genes associated with each GO term are shown on the x-axes. The Fisher's exact test adjusted p-value statistic for each GO term is shown based on a blue-yellow gradient (adjusted p-value < 1.00).

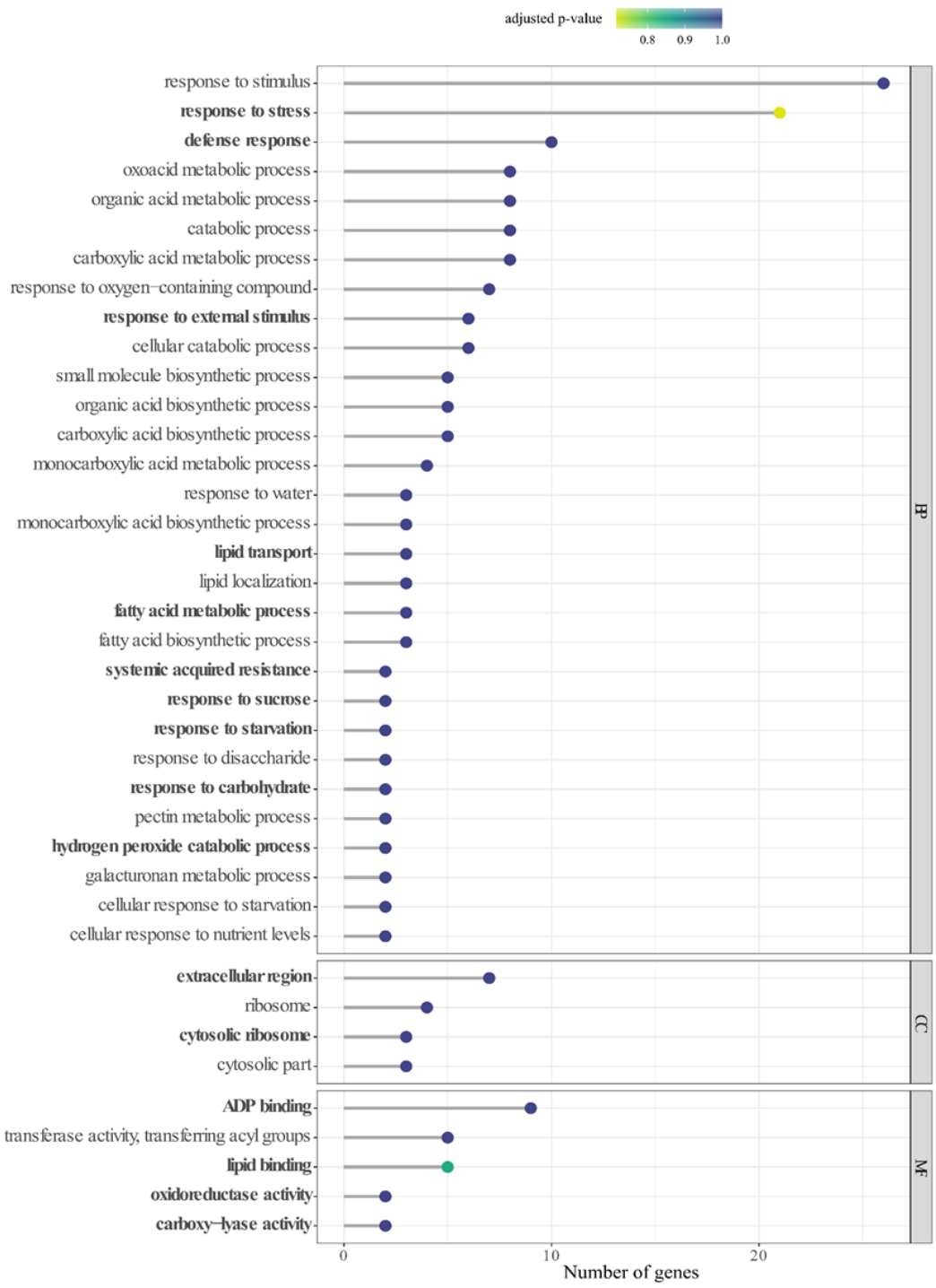


Figure 11: Disease resistance pathways are enriched in the susceptible *E. grandis* genotype at 2 days after *A. psidii* challenge. Enriched Gene Ontology terms in the categories biological process (BP), cellular compartment (CC) and molecular function (MF) for the 142 genes identified as drivers of the susceptible genotype during early pathogen challenge at 48 hours after inoculation (hai). The GO terms are shown on the y-axes with interesting terms bolded. The number of genes associated with each GO term are shown on the x-axes. The Fisher's exact test adjusted p-value statistic for each GO term is shown based on a blue-yellow gradient (adjusted p-value < 1.00).

5.4 Scoping of new resistance pathways and control options for future research

We then focused our efforts on the transcriptomes of mock-inoculated hypersensitive, resistant, and susceptible *E. grandis* genotypes to determine if there were genotypic differences that may be promoting resistance prior to challenge with the myrtle rust pathogen *A. psidii*. If we can identify baseline genetics that are responsible for pre-adapted immunity, this could be used in future to screen for germplasm that is likely to survive myrtle rust challenge.

Partial Least Squares-Discriminant Analysis (PLS-DA) revealed that 43% of the variation between RNA-sequencing sample types was explained by X-variate 1 separating the samples of the susceptible genotype from those of the resistant and hypersensitive (Figure 12A). X-variate 2 explained 12% of the variation separating the samples of the hypersensitive genotype from the resistant genotype and the susceptible genotype (Figure 12A). We then focused on the loadings from X-variate 1 to identify the genes that most differentiated the susceptible genotype from the combined grouping of the hypersensitive and resistant genotype. From the loadings we identified 695 genes for the combined hypersensitive and resistant group and 924 genes for the susceptible genotype that were driving the difference between the two groups in the absence of the pathogen (loadings of X-variate 1 cut-off: +/- 0.093, Figure 12B)). To identify the genes differentiating the hypersensitive genotype from the resistant genotype, we focused on the loadings from X-variate 2 of the PLS-DA. We identified 569 genes for the hypersensitive genotype and 372 genes for the resistant genotype that were most differentiating the two (loadings of X-variate 2 cut-off: +/- 0.03, Figure 12C).

Expression profiling of the log2 transformed data from the combined group of the hypersensitive and resistant genotypes revealed that of the 695 genes, 128 genes that displayed high expression in the resistant and hypersensitive group compared to the susceptible group, 15 were linked to disease resistance including disease resistance protein, glycosyl hydrolase, protein kinases and terpene synthases (Figure 13A-i). Furthermore, 215 genes displayed a 1.5-3-fold higher expression in the hypersensitive group compared to the susceptible group, of which 32 were linked to disease resistance such as *TIR-NBS-LRR class disease resistance protein* (*Eucgr.E03556.1*), *LRR protein kinase* (*Eucgr.J02919.1*), *terpene synthase-like* (*Eucgr.K00827.1*) and *glycoside hydrolase* (*Eucgr.D01668.1*; Figure 13A-ii). Finally, 352 genes displayed similar expression patterns between the hypersensitive, resistant, and susceptible genotypes with only marginally higher expression in the combined resistant and hypersensitive group compared to the susceptible group and of these 24 genes were identified with functions linked to disease resistance including protein kinases (Figure 13A-iii). Of the 924 susceptible gene drivers, 182 genes displayed high expression in the susceptible compared to the resistant and hypersensitive genotypes of which 16 were linked to disease resistance such as *basic chitinase* (*Eucgr.I02267.1*), *pathogenesis-related protein 4* (*Eucgr.L03258.1*) and a mitogen activated protein kinase kinas kinase (*Eucgr.A01110.1*; Figure 13B-i). Two genes, a *plant invertase* (*Eucgr.E01454.1*) and *ABC-2 type transporter* (*Eucgr.G02776.1*) displayed very high expression in the susceptible compared to the resistant and hypersensitive suggesting a potential role in promoting susceptibility (Figure 13B-ii). Moreover, 137 genes were identified that displayed a 1.5-2-fold higher expression in the susceptible compared to the resistant and hypersensitive genotypes of which 11 displayed disease resistance functions including cysteine-rich receptor like protein kinase (RLK), disease resistance protein and wound-response protein (Figure 13B-ii). Lastly, 605 genes displayed similar expression between the hypersensitive and resistant group versus the susceptible genotype with slightly higher expression in the susceptible genotype and of these 29 were linked to disease

resistance mechanisms with some examples including LRR-kinase, mitogen-activated protein kinase (MAPK), glycoside hydrolase, enhanced disease resistance protein and heat shock protein (Figure 13B-iii).

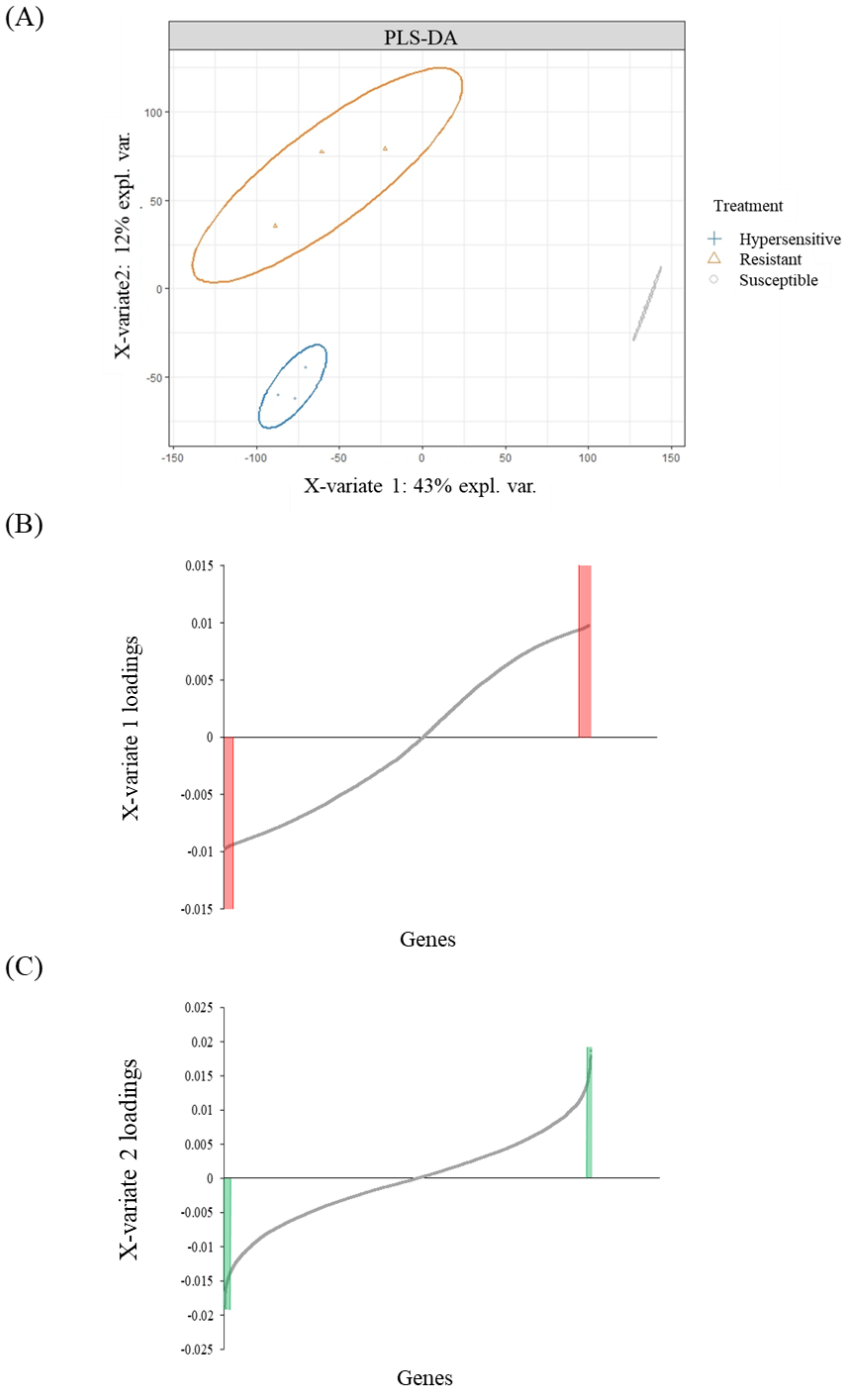


Figure 12: Identification of genes driving differences in hypersensitive, resistant and susceptible *Eucalyptus grandis* genotypes in the absence of *Austropuccinia psidii* challenge (A) Partial Least Squares-Discriminant Analysis (PLS-DA) of RNA-sequencing samples derived from hypersensitive, resistant and susceptible *Eucalyptus grandis* plants in the absence of *A. psidii* challenge. (B) Loadings of genes driving separation of the combined resistant and hypersensitive genotypes, and the susceptible genotype. The red shaded areas show the genes driving the separation between the combined resistant and hypersensitive genotypes, and susceptible genotype at a cut-off of +/- 0.093 on X-variate 1, respectively. (C) Loadings of genes driving separation of the resistant and

hypersensitive genotypes. The green shaded areas show the genes driving the separation between the resistant and hypersensitive genotypes at a cut-off of +/- 0.013 on X-variate 2, respectively.

Of the 569 genes differentiating the hypersensitive genotype, 149 were highly expressed in the hypersensitive genotype compared to the resistant and the susceptible genotypes (Figure 13C-i). Of these, a *Terpene synthase* (*Eucgr.E03610.1*), *LRR kinase* (*Eucgr.I00392.1*) and *Eucgr.F00487.1* with unknown function displayed the highest relative expression in the hypersensitive genotype compared to the susceptible and resistant genotypes. Furthermore, a total of 49 genes were annotated with disease resistance functions. In addition, 10 displayed opposite relative expression, being very lowly expressed in the hypersensitive while highly expressed in the resistant and susceptible and of these three genes have prescribed disease resistance functions including terpene synthase (*Eucgr.L02082.1*), *LRR kinase* (*Eucgr.I00453.1*) and *Pathogenesis-Related 1* (*Eucgr.L01705.1*; Figure 13C-ii). In addition, a *Basic-leucine zipper transcription factor* (*Eucgr.J02482.1*)and *UDP-glucosyl transferase* (*Eucgr.L01660.1*) were identified as the most lowly expressed genes in the hypersensitive relative to the resistant and susceptible (Figure 13C-ii). Furthermore, 229 genes displayed 1-3-fold higher expression in the hypersensitive genotype compared to the resistant whereas the majority of these genes were similarly expressed in the susceptible genotype and of these 52 were prescribed resistance functions including disease resistance proteins, cysteine-rich RLKs and LRR kinase proteins (Figure 13C-iii). Finally, 181 genes displayed similar expression patterns between the hypersensitive, resistant, and susceptible genotypes and within this set, 17 were linked to disease resistance such as disease resistance proteins, cysteine-rich RLKs and ethylene forming enzyme (Figure 13C-iv). Within the 371 genes most differentiating the resistant genotype from the hypersensitive genotype, 60 were most abundant in the resistant and nine have prescribed disease resistance functions including disease resistance protein, glutathione-S-transferase and RLK (Figure 13D-i). Of these, a *disease resistance protein* (*Eucgr.H03831.1*), *cysteine-rich RLK* (*Eucgr.E03041.1*), *cytochrome P450* (*Eucgr.D00205.1*) and *Eucgr.E03377.1* of unknown function that displayed the highest relative expression in the resistant compared to the hypersensitive and susceptible genotypes (Figure 13D-i). We also identified a gene, Crinkly related 4 (*Eucgr.H00418.1*) that was down-regulated in the resistant relative to the hypersensitive and susceptible genotypes (Figure 13D-ii). In addition, 72 genes displayed marginally higher relative expression in the resistant compared to the hypersensitive and susceptible (Figure 8D-iii). Lastly, 239 genes displayed similar expression patterns between the resistant, hypersensitive, and susceptible (Figure 13D-iv).

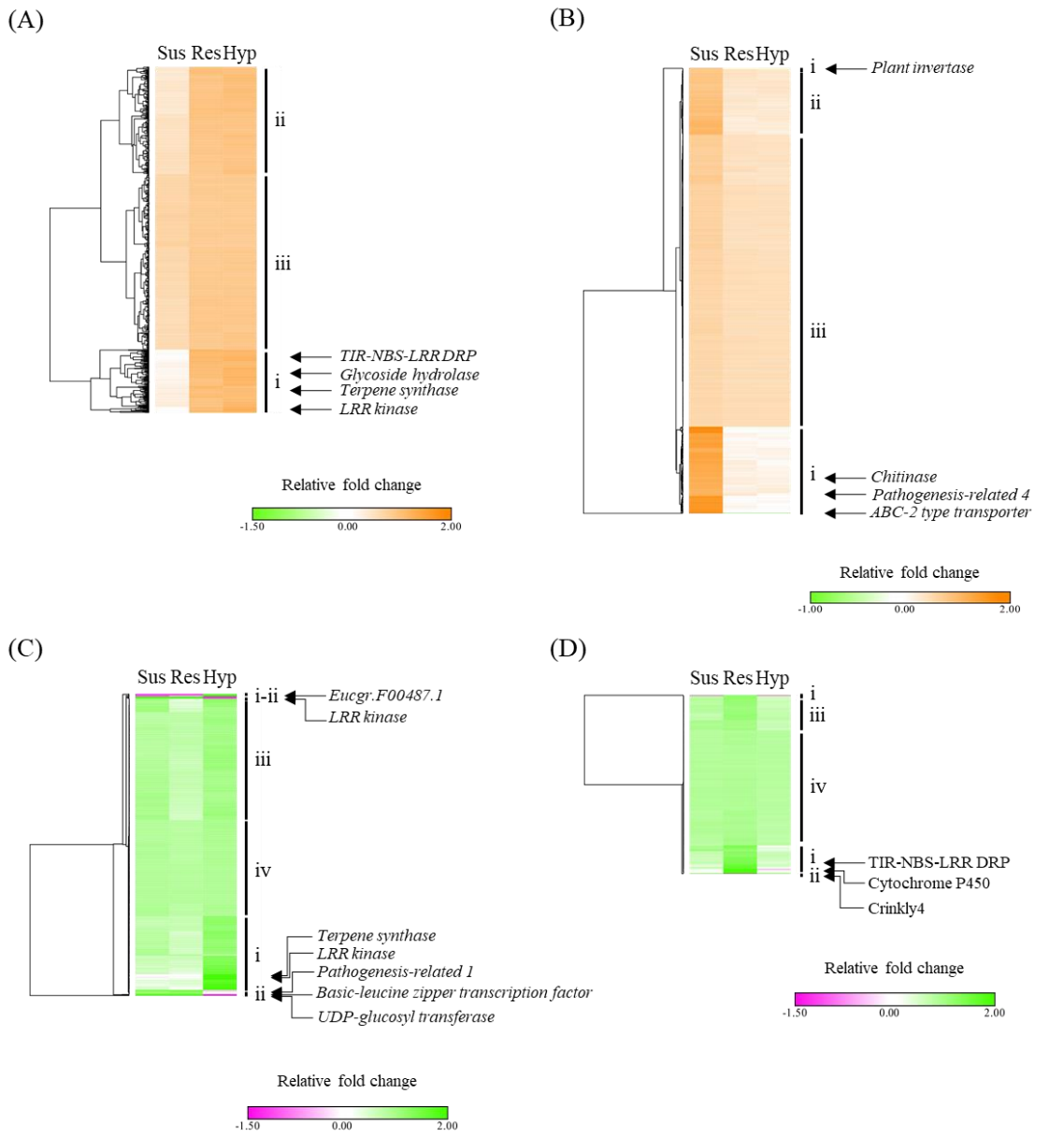


Figure 13: Expression patterns of genes driving differences between *Austropuccinia psidii* hypersensitive, resistant and susceptible *E. grandis* genotypes in the absence of the pathogen. (A and B) Hierarchical clustering of Log2-transformed data of 695 and 924 genes driving the combined group of the resistant and hypersensitive genotypes and susceptible genotype, respectively based on X-variate 1 of PLS-DA separating these two groups. All the data points shown are the ratio of transcript abundance in each treatment relative to the average across all treatments for each heatmap, respectively, where up-regulated genes are presented as green (low) and down-regulated genes are presented as orange (high), respectively. The heatmaps are annotated on the right-hand side with hierarchical clustering groups denoted i-iii where i: genes displaying very high expression in the resistant and hypersensitive combined group compared to the susceptible (A) and vice versa (B); ii: genes displaying 2-3-fold higher expression in the resistant and hypersensitive combined group compared to susceptible (A) and vice versa (B); iii: genes displaying similar expression patterns between resistant, hypersensitive and susceptible in A and B, respectively. (C and D) Hierarchical clustering of Log2-transformed data of 569 and 372 genes driving the hypersensitive and resistant, respectively based on X-variate 2 of PLS-DA separating these two groups. All the data points shown are the ratio of transcript abundance in each treatment relative to the average across all treatments for each heatmap, respectively, where up-regulated genes are presented as purple (low) and down-regulated genes are presented as green (high), respectively.

Outcomes:

The data generated has contributed towards the Myrtle Rust Action plan. We discovered a subset of chemical and genetic fingerprints that can be used for screening for the early stages of infection by *A. psidii.* **We report that both metabolomic and transcriptomic techniques are applicable in identifying plant phenotypes and that they could be used to informing breeding strategies (Action 4.3.1).**

Both metabolomics and transcriptomics has identified a number of **resistance pathways that should be targets for new areas of research**. Additionally, pathways associated with the susceptible phenotype have been identified and **should be the target of research for novel control methods (Actions 4.3.3; 4.3.5).**

We have identified four small secreted proteins expressed by *A. psidii* that are essential to early infection, three of which are highly expressed within the susceptible phenotype. **These genes should be particularly monitored for changes in Australian populations of *A. psidii* (Action 5.3.1).** Evolution of these proteins over time could impact early plant recognition of the parasite by resistant germplasm.

Reporting:

- **Dissemination of activities using social media:** Twitter feeds have been used to disseminate information about myrtle rust. At the stage of publication of our research data, we will further disseminate findings of our research, with reference to APBSF.

- **Engagement early career researchers through integration of research with special undergraduate student labs in the 'Plant Health and Biosecurity' unit at Western Sydney University:** Laboratory research was cancelled throughout the year due to the SARS-CoV2 pandemic. Despite this, undergraduate student Fatima Karagully was provided with an online project to write a literature review on the impact of myrtle rust. This was also disseminated to other undergraduate students via two video presentations on the topic.

We also employed two casual early career researchers for the processing of samples and data analysis performed for this study. Dr Johanna Wong performed *A. psidii* infections, in collaboration with Dr Peri Tobias at the University of Sydney, phenotype scoring, metabolomic sample processing and data analysis. Dr Wong completed her PhD in 2020 and is now employed full time at the Department of Primary Industries. Mr Donovin Coles performed data analysis of the transcriptome of *E. grandis*. He is due to complete his PhD in 2021 at Western Sydney University.

- **Contribution to Masterclasses on plant health:** This event was cancelled due to the SARS-CoV2 pandemic, however we will be contributing to the 2021 event with a presentation.

- **Publication of results in high impact journals:** We are preparing two manuscripts based on the research presented above for publication in high impact journals. The first manuscript will describe the transcriptomic analysis of early *A. psidii* infection in susceptible vs. resistant *E. grandis* phenotypes.

The second manuscript will describe the metabolomics analysis of early *A. psidii* infection in susceptible vs. resistant *M. quinquenervia* phenotypes. Both manuscripts will be published in the preprint server BioRxiv or similar to ensure rapid dissemination of the results to the myrtle rust community and beyond.

- **Presentation of results at the ASM conference 2020:** The ASM conference in 2020 was cancelled due to the SARS-CoV2 pandemic. A presentation was given at the Western Sydney University School of Science research seminars which was advertised on twitter in order to reach a larger audience. We also hope to present our results to the Myrtle Rust Action Plan Symposium and will continue to look for opportunities to present this and future myrtle rust research to a wider audience in 2021 and beyond.

6. Discussion and Conclusion

This transcriptomics and metabolomics study identified unique chemical and genetic cues that appear during the early interaction between *A. psidii* and a host plant. Further, a sub-set of these chemical and genetic signatures appear to predict if the outcome of this interaction will be resistant, hypersensitive, or susceptible, before any visible signs of infection.

Our project employed untargeted metabolomics, which enabled us to identify global changes to a broad spectrum of known and unknown metabolites. Our results show that in *M. quinquenervia*, *A. psidii* had a significant influence on the metabolome at 24 to 48 hai, with the plant metabolic response decreasing by 5 days after inoculation. This result is consistent with other transcriptomic studies, which have indicated that plant response is greatest at 24 hai. A subset of metabolites were identified that are influenced by phenotype, including the organoheterocyclic compounds terpenoids and flavonoids. Further analyses will be performed to identify metabolites specific to resistant, hypersensitive and susceptible phenotypes. During the early time points (24 to 48 hai), 19 metabolites could be identified that are enriched in the resistant phenotype, including a molecule with a predicted structure known to enhance disease protection (Grellet Bournonville et al., 2020). This is the first reported untargeted metabolomics study of *A. psidii* infection in an Australian Myrtaceaeous species. A major advantage of metabolomics compared with transcriptomics is that some pathways may be additionally regulated at either the transcriptional or post-transcriptional level. Currently, the computational tools used to predict structures of metabolites based on mass spectrometry profile are still under development and there are limited studies of metabolomics during rust infection of any plant. As a result, many of the metabolite structures that were identified in this study cannot be predicted and further work is required to understand the molecular mechanisms associated with resistance provided by these metabolites.

We subsequently performed RNA-sequencing on the myrtle rust pathogen *A. psidii* during infection of resistant and susceptible *E. grandis* genotypes at 48 hours after inoculation. Transcripts encoding for three *A. psidii* small secreted proteins were highly expressed in the susceptible compared to the resistant host genotypes. These can be used as markers to distinguish between resistant and susceptible at early stage of infection with very high sensitivity and species-specific accuracy. Analysis of the *E. grandis* transcripts 48 hours after *A. psidii* challenge, identified that 11 genes can be used as

markers for resistance, while 9 genes can be used as markers of susceptibility. Here we also identified that *E. grandis* displays genotypic differences prior to *A. psidii* challenge that may promote resistance of this species to infection by *A. psidii.* We identified 32 genes that are markers for resistance (resistant and hypersensitive) vs susceptible, while 16 genes were identified as markers for susceptibility.

Previous studies have investigated the transcriptomics of *M. quinquenervia* and other Australian Myrtaceae in response to *A. psidii* (Hsieh et al., 2018; S. A. Santos et al., 2020; Tobias et al., 2018). These studies have also investigated the differences between resistant and susceptible plants at the gene expression level. Hsieh et al. (2018) were unable to find any consistent resistance gene expression in *M. quinquenervia* plants with resistance to *A. psidii*. It was suggested that this may be due to more than one stress response being involved in resistance to infection (Hsieh et al., 2018).

7. Recommendations

Our results support the aim of this study, that a curated set of both chemical and genetic cues can be used to for novel, sensitive screening tests. Based on the four main objectives of this project, we provide the following recommendations:

1. Identification of *A. psidii*-specific chemicals enabling novel, sensitive screening tests

The results of this study have identified a curated set of metabolites that can be used to detect the early stages of *A. psidii* before infection is visible.

Recommendations:

- Metabolic fingerprinting can be developed as a sensitive method for detecting *A. psidii* infection in the first 48 hours, regardless of the resulting phenotype.
- Metabolic fingerprinting is not recommended for mid- to late-stage *A. psidii* infection
- Additional experimentation is required to streamline a defined metabolic fingerprint, which can be used for other Myrtaceaeous species.

2. Identification of fast-evolving proteins in *A. psidii* for use in future population and evolutionary studies

Small secreted protein (SSP) gene expression by *A. psidii* has been detected and can be considered a highly sensitive and specific way of detecting the presence of the pathogen. Relative expression of the SSP genes is different in resistant and susceptible hosts.

Recommendations:

- Screening tools for detection of early infection could be developed based on the SSP genes detected in this study.
- Differential expression of these SSP genes in resistant and susceptible hosts across a larger range of Myrtaceaeous species should be tested to confirm that this pattern of expression is consistent.

3. Data on plant traits associated with disease resistance

Metabolomics and transcriptomics were effective in identifying a set of metabolites and transcripts that were specific to the resistant and susceptible phenotypes of *Melaleuca quinquenervia* and *Eucalyptus grandis*, respectively.

Recommendations:

- Confirmation of phenotype-specific metabolic fingerprints is required across a range of Myrtaceaeous species to determine its applicability to screening for plant breeding purposes.
- Isolation and structural characterisation of metabolites associated with resistance during early infection to better understand their role in the resistance mechanism.
- Screening for orthologues of the resistance and susceptible genes in new model systems to see if they are broadly applicable biomarkers of phenotype.
- Use a refined gene list generated in the previous point to then investigate the mechanism of these genes contributing to disease resistance for future germplasm screening.

4. Scoping of new resistance pathways and control options for future research

A set of resistance genes were identified to be upregulated in *Eucalyptus grandis* prior to *A. psidii* inoculation that can differentiate between resistant, hypersensitive and susceptible phenotypes.

Recommendations:

- Develop a panel of markers that are linked to phenotype, by testing a wider range of Myrtaceae species. This could be developed using the ThermoFisher AgriSeq™ HTS Library Kit as a basis.

9. Appendices, References, Publications

Chitarrini, G., Soini, E., Riccadonna, S., Franceschi, P., Zulini, L., Masuero, D., Vecchione, A., Stefanini, M., Di Gaspero, G., Mattivi, F., & Vrhovsek, U. (2017). Identification of Biomarkers for Defense Response to *Plasmopara viticola* in a Resistant Grape Variety. In Frontiers in Plant Science (Vol. 8). https://doi.org/10.3389/fpls.2017.01524

dos Santos, I. B., da Silva Lopes, M., Bini, A. P., Tschoeke, B. A. P., Verssani, B. A. W., Figueredo, E. F., Cataldi, T. R., Marques, J. P. R., Silva, L. D., Labate, C. A., & Quecine, M. C. (2019). The *Eucalyptus* Cuticular Waxes Contribute in Preformed Defense Against *Austropuccinia psidii*. In Frontiers in Plant Science (Vol. 9). https://doi.org/10.3389/fpls.2018.01978

Garcia, P. G., dos Santos, F. N., Zanotta, S., Eberlin, M., & Carazzone, C. (2018). Metabolomics of *Solanum lycopersicum* Infected with *Phytophthora infestans* Leads to Early Detection of Late Blight in Asymptomatic Plants. In Molecules (Vol. 23, Issue 12, p. 3330). https://doi.org/10.3390/molecules23123330

Grellet Bournonville, C., Filippone, M. P., Di Peto, P., Trejo, M. F., Couto, A. S., Mamaní de Marchese, A., Díaz Ricci, J. C., Welin, B., & Castagnaro, A. P. (2020). Strawberry fatty acyl glycosides enhance disease protection, have antibiotic activity and stimulate plant growth. Scientific reports, 10(1), 8196. https://doi.org/10.1038/s41598-020-65125-7

Hsieh, J.-F., Chuah, A., Patel, H. R., Sandhu, K. S., Foley, W. J., & Külheim, C. (2018). Transcriptome Profiling of *Melaleuca quinquenervia* Challenged by Myrtle Rust Reveals Differences in Defense Responses Among Resistant Individuals. Phytopathology, 108(4), 495–509.

Makinson, R., Pegg, G., Carnegie, A. (2020). Myrtle Rust in Australia: A National Action Plan. Australian Plant Biosecurity Science Foundation, Canberra, Australia.

R Core Team 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. *bioRxiv*, 108597.

Santos, S. A., Vidigal, P. M. P., Guimarães, L. M. S., Mafia, R. G., Templeton, M. D., & Alfenas, A. C. (2020). Transcriptome analysis of *Eucalyptus grandis* genotypes reveals constitutive overexpression of genes related to rust (*Austropuccinia psidii*) resistance. Plant Molecular Biology. https://doi.org/10.1007/s11103-020-01030-x

Silva, E., da Graça, J. P., Porto, C., Martin do Prado, R., Hoffmann-Campo, C. B., Meyer, M. C., de Oliveira Nunes, E., & Pilau, E. J. (2020). Unraveling Asian Soybean Rust metabolomics using mass spectrometry and Molecular Networking approach. Scientific Reports, 10(1), 138.

Sprenger, H., Erban, A., Seddig, S., Rudack, K., Thalhammer, A., Le, M. Q., Walther, D., Zuther, E., Köhl, K. I., Kopka, J., & Hincha, D. K. (2018). Metabolite and transcript markers for the prediction of potato drought tolerance. Plant Biotechnology Journal, 16(4), 939–950.

Tobias, P. A., Guest, D. I., Külheim, C., & Park, R. F. (2018). De Novo Transcriptome Study Identifies Candidate Genes Involved in Resistance to *Austropuccinia psidii* (Myrtle Rust) in *Syzygium luehmannii* (Riberry). Phytopathology, 108(5), 627–640.

Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., Jones, A., Lou, Z., Zhang, P., Sandhu, K., Smith, G. R., Tibbits, J., Chagné, D. & Park, R. F. 2020. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. ***bioRxiv*****,** 2020.03.18.996108.

Yong, W. T. L., Ades, P. K., Goodger, J. Q. D., Bossinger, G., Runa, F. A., Sandhu, K. S., & Tibbits, J. F. G. (2019). Using essential oil composition to discriminate between myrtle rust phenotypes in *Eucalyptus globulus* and *Eucalyptus obliqua*. In Industrial Crops and Products (Vol. 140, p. 111595). https://doi.org/10.1016/j.indcrop.2019.111595


PLANT BIOSECURITY
science foundation