TRANSCRIPTOME ANALYSIS AND METABOLITE PROFILING FOR DISSECTION OF MECHANISMS OF RESISTANCE TO PHYTOPHTHORA ROOT ROT (PHYTOPHTHORA RUBI) IN RED RASPBERRY (RUBUS IDAEUS L.)

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Judson Arthur Ward

January 2012

TRANSCRIPTOME ANALYSIS AND METABOLITE PROFILING FOR THE DISSECTION OF MECHANISMS OF RESISTANCE TO PHYTOPHTHORA ROOT ROT (PHYTOPHTHORA RUBI) IN RED RASPBERRY (RUBUS IDAEUS L.)

Judson Arthur Ward, Ph.D.

Cornell University 2012

ABSTRACT

In red raspberry (*Rubus idaeus L.)* resistance to *Phytophthora* Root Rot (PRR), caused by *P. rubi,* is a critical factor in disease management. Therefore, identifying loci linked to resistance is a priority. Previous studies suggest that resistance in the cultivar 'Latham' is quantitative and induced. RNAseq was performed to investigate induced resistance in 'Latham' and susceptibility in the cultivar 'Titan'. Four sequencing libraries ('Latham' control, 'Latham' treatment, 'Titan' control, and 'Titan' treatment) were constructed from mRNA pooled across 3 time points (2 days, 5 days, and 20 days post inoculation). The libraries were sequenced independently, resulting in 15,213,629 reads from the 'Latham' control, 18,923,742 reads from the 'Latham' treatment, 18,363,149 reads from the 'Titan' control, and 18,145,135 reads from the 'Titan' treatment. Reads were aligned to the previously sequenced woodland strawberry (*Fragaria vesca)* genome with Bowtie/TopHat. Gene expression was quantified and differential expression was calculated with Cufflinks. This detected 15,164 genes in the 'Latham' control, 15,716 genes in the 'Latham' treatment, 14,896 genes in the 'Titan' control, and 13,991 genes in the 'Titan' treatment. The union of the four data sets included 16,956 unique genes. De novo assembly of all

reads with Trinity produced over 36,000 transcripts with BLASTx hits (maximum e-value threshold of $1.0x10^{-3}$) to vascular plants and over 500 with hits to *Phytophthora.* Significantly differentially expressed genes between the 'Latham' control and 'Latham' treatment included genes directly involved in defense responses such as Pathogenesis Related 10 proteins (mal-d), WRKY transcription factors, and NPR1. These results suggest activation of the Salicylic Acid (SA) resistance pathway. Other defense responses commonly observed in SA mediated resistance include upregulation of key genes in the lignin biosynthetic pathway, such as caffeoyl-CoA O-methyltransferase and key genes in the tricarboxylic acid cycle (TCA) such as ATP citrate synthase. Metabolite profiling with GC/MS revealed few differences between treatments and controls, but significantly higher concentrations of citrate were found in the 'Latham' control compared to the 'Titan' control (p adjusted=0.0008) and in the 'Latham' treatment compared to the 'Titan' treatment (p adjusted=0.0164), suggesting that 'Latham' has increased flux through the TCA.

BIOGRAPHICAL SKETCH

Judson Arthur Ward was born June 22, 1979 and raised in Pierre, South Dakota. He and his sister, Carrie Ward, are the only two children of Don and Mary Ward. Both siblings enjoyed working with their parents at family businesses until graduation from high school. Judson attended the University of Arizona and in June of 2004 he graduated Magna cum Laude with a Bachelor's degree in Plant Science. He managed tissue culture laboratories for Driscoll's Strawberry Associates in Watsonville CA while his wife, Dr. Rhiannon Crain, worked toward her PhD in informal science education at the University of California, Santa Cruz. In the fall of 2008 he began working with Dr. Courtney Weber at Cornell. He completed his work in the fall of 2011.

ACKNOWLEDGMENTS

I would like to acknowledge my parents for instilling in me a solid and admirable work ethic. Their enthusiasm for art and science helped me achieve this PhD. I am also very thankful for the support of my wife, Rhiannon. She is always willing to entertain my scientific rants and was pivotal in directing me toward scholarly goals as an undergraduate. I am also thankful for my bright and loving son, Obe, who was tolerant of my attention to the computer in our living room throughout the process. I will have fond memories of him playing Legos behind me or sitting on my lap listening to Prokofiev's *Peter and the Wolf* with headphones while I wrote. I hope that he will share warm feelings toward academics and that I have not induced some aversion!

I am also grateful for the guidance provided by Dr. Courtney Weber and my other committee members, Dr Chris Smart and Dr. Lailiang Cheng who were all supportive throughout the process. I am especially grateful to Dr. Weber for his willingness to commit resources to this project. I would also like to thank William Boone for his help in the lab and field and for the many great conversations in the office.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

LITERATURE REVIEW

1.1. Red Raspberry Genetics and Genomics

Red raspberry is a member Rosaceae family and the subfamily Rosoideae. It belongs to the genus *Rubus,* which is not included in any Tribe of the Rosoideae subfamily (Potter et al. 2007). The number of species in *Rubus* is estimated to be between 600 and 800 (Thompson 1995). *Rubus ideaus* L. (red raspberry) is the most economically important species and is comprised of a European subspecies, *R. ideaus vulgatus* Arrhen., and a North American subspecies, *R. ideaus strigosus* Michx. Both subspecies are diploid (2n = 2x = 14) with a basic set of seven chromosomes (Thompson 1995). The nuclear haploid genome of *R. idaeus* is approximately 280 MB in size (Arumuganathan & Earle 1991).

R. idaeus is highly heterozygous, particularly in wild populations, due to the existence of dominant self-incompatibility alleles that function to limit pollen tube formation in the style (Keep 1968). Germplasm used in breeding programs has been selected for self-compatibility, but inbreeding depression can be severe (Keep 1968; Keep 1969). This constrains population development in breeding and molecular studies. Breeding programs use recurrent mass selection with only limited self-pollination and backcrossing for specific traits.

In molecular studies, map construction has been performed by building maternal and paternal maps separately with the pseudo-test cross method (Grattapaglia & Bertolucci 1995; Grattapaglia & Sederoff 1994) or by constructing a singular map with the use of allelic bridges

(Ritter et al. 1990; Maliepaard & Jansen 1997). These methods take advantage of the various marker configurations in populations derived from heterozygous individuals but complicate the map generation process.

1.2. Identification of the causal agent of Phytophthora Root Rot (PRR) of Red Raspberry

In the 1930's cases of a "raspberry dieback" were reported throughout Scotland and England in association with poorly drained soils and standing water (Waterston 1936). Initially, *Pythium*, *Podospora*, *Fusarium*, and *Botrytis* species were isolated from the affected plants, but none of these were capable of reproducing disease symptoms upon inoculation of healthy plants (Waterston 1936). The first documented evidence that a *Phytophthora* species was involved in this "raspberry dieback" was given by Waterston when he isolated a *Phytophthora* species and suggested that it belonged to the *cactorum–omnivora* group (Waterston 1936). Later, Converse and Swartz (1965) published a short abstract suggesting that the *Phytophthora* species, which caused disease on red raspberry, resembled *Phytophthora fragariae*. They rejected this species designation when they found their isolates were only pathogenic on raspberry and not on strawberry (Converse & Schwartze 1968; Converse 1965). Three years later they decided that these isolates were "*Phytophthora erythroseptica* Pethyb (sensu lato)" and declared that this species was the primary cause of "wet soil root rot" in the Pacific Northwest of the United States (Converse & Schwartze 1968). Later, isolates that were most pathogenic on raspberry in Scotland were identified as a form of *Phytophthora megasperma*, but it was noted that these isolates resembled *P. fragariae* (Duncan et al. 1987).

Wilcox began to resolve some of these conflicting reports when he performed virulence tests and culture comparisons of isolates from New York and demonstrated that the most

common virulent species attacking raspberry were a variant of *P. fragariae* that was not pathogenic on strawberry (Wilcox 1989b). He also reviewed Waterston's 1937 work and found that the sporangia and oospores drawn and measured by Waterston were "similar to those of *P. citricola* Swada*",* pathogenic on raspberry but not commonly detected (Wilcox 1989b). This work led to an international collaboration that involved side-by-side comparisons of cultures, electrophoretic banding patterns, and pathogenicity of isolates from around the world. This definitive work clearly demonstrated that the common *Phytophthora* isolates from around the world which were highly pathogenic in raspberry were a closely related variant *P. fragariae* and thus these isolates were designated as *P. fragariae* var. *rubi* (Wilcox et al. 1993).

In 1997 Cooke and Duncan took this a step further and analyzed phylogenetic relationships of several *Phytophthora* species based on Internal Transcribed Spacer (ITS) sequence (ITS1 and ITS2 sequences) and confirmed that *P. fragariae* var. *rubi* was very closely related to *P. fragariae* var. *fragariae* (Cooke & Duncan 1997). Cooke and Duncan also found that other close relatives included *P. cambivora* and *P. cinnamomi*. These results were later confirmed in analyses of many *Phytophthora* species (Forster 2000; Blair et al. 2008). The close relationship of *P. fragariae* var. *rubi* to *Phytophthora cinnamomi* raises interesting questions regarding host specificity and host resistance since the host range of *P. fragariae* var. *rubi* so far includes only raspberry, while *P. cinnamomi* is known to have a host range greater than 3,000 species (Hardham 2005). This also reinforces the importance of further research on the mechanisms of durable resistance in red raspberry because understanding this resistance may contribute to the development of resistance to *P. cinnamomi* in other species. Based on gene flow studies, in 2007, it was demonstrated that *P. fragariae* var. rubi is a distinct species and was thus designated *P. rubi* (Man in 't Veld 2007).

1.3. Biology and Diagnosis of *P. fragariae* var. *rubi*

P. fragariae var. *rubi* is homothallic and produces non-papillate sporangia (Wilcox et al. 1993). Unlike aerially dispersed *Phytophthora* species such as *P. infestans*, the life cycle of *P. fragariae* var. *rubi* is almost exclusively subterranean (Wilcox 1992). *P. fragariae* var. *rubi* zoospores are attracted by chemotaxis to the zone of elongation in fine raspberry root tissue where they encyst, germinate, and penetrate host tissue (Brunner-Keinath & Seemueller 1992) [as cited by Laun in 1997 (Laun & Zinkernagel 1997): the original paper is only available in German]. After penetration of fine root tissues the mycelia grow through the root and the disease progresses upward in the plant sometimes producing lesions on the cane (Wilcox 1989a).

Diagnosis of the disease by isolation can be difficult since roots can be difficult to surface sterilize and other competing non-pathogenic fungi or bacteria frequently grow despite the use of selective media. The most commonly used approach to isolating this pathogen is use of a selective cornmeal agar prepared with pimaricin, ampicillin, rifampicin, and hymexazol in concentrations given by Jeffers and Martin (Jeffers & S. Martin 1986). Once isolated, a pathologist can identify the pathogen by morphology and can fulfill Koch's Postulates to confirm that the isolate is virulent and the cause of the observed symptoms. Since isolation is a tedious process with a high failure rate, numerous molecular approaches have been developed for detection. These include Enzyme Linked Immunosorbent Assay (ELISA) based methods (Olsson & Heiberg 1997) and a number of polymerase chain reaction (PCR) based detection assays (Stammler et al. 1993; Bonants et al. 1997).

1.4. Control of PRR in Red Raspberry

PRR caused by *P. fragariae* var. *rubi* is best managed by growing the most resistant variety available on a raised bed (Wilcox et al. 1999; Maloney et al. 1993; Maloney et al. 2005; Heiberg 1995). The first comprehensive study of management examined the effect of bed height, straw mulch, preplant application of *Trichoderma virens*, and seasonal application of metataxyl. This study found that "strong differences in cultivar susceptibility overwhelmed the effects of all other factors" under the conditions in the study (Wilcox et al. 1999). Secondly, raised beds were found to significantly increase yield in moderately susceptible varieties and straw mulch increased PRR, which was consistent with previous work (Maloney et al. 1993; Heiberg 1995).

1.5. Host resistance

The ineffectiveness of management practices alone in controlling PRR further highlights the importance of genetic resistance. Durable resistance to PRR in red raspberry has been identified in the cultivar 'Latham', which developed from a cross of 'King' x 'Loudon' in 1908 at the University of Minnesota and was distributed to growers in 1914 (Shoemaker 1934). Crandall (1997) documented very good resistance to PRR in 'Latham' in his development of new resistant varieties for the Pacific Northwest of the United States (Crandall 1977).

Now, over 30 years after this resistance was first identified and over 100 years since the cross was first made, there are still no reports of a resistance-breaking strain of *P. fragariae* var. *rubi*. The durability of resistance may be due to a number of factors, including any one or any combination of the following:

1. A unique mechanism of resistance relying predominantly upon two major dominant genes (supported by Pattison et al. 2004; 2007);

2. Lack of selection pressure because of unknown alternative hosts or survival of the pathogen on susceptible varieties;
3. Low genetic recombination in the pathogen due to its homothallic nature.

The answer to the question of why this resistance has proven to be so durable can only be through a more thorough and careful investigation of host resistance and biology of the pathogen.

Because of its well-documented and durable resistance ‘Latham’ has been the subject of many studies of PRR resistance in red raspberry. One notable exception was a study performed by Laun and Zinkernagel (1997) in which they made microscopic observations of the highly susceptible red raspberry cultivar 'Schonemann', a moderately resistant red raspberry cultivar 'Winklers Samling' (= cv. ‘Asker’), and the highly resistant hybrid-berry cultivar ‘Tayberry’ (raspberry x blackberry). No differences were found in encystment between the cultivars, but significantly reduced sporulation, limited pathogen spread, and healthy root regeneration was observed in ‘Winklers Samling.’ In ‘Tayberry’ little root damage was observed aside from some necrotic root tips. In 'Schonemann' *Phytophthora* colonized the entire root system, spread to the base of the stem, and had increased formation of oospores throughout.

The descriptions of disease progression in ‘Winklers Samling’ and ‘Latham’ are similar, but it is unknown whether the resistance is mechanistically the same. In the case of ‘Tayberry’ the resistance mechanism may be distinct possibly because its resistance was derived from blackberry. The work of Laun and Zinkernagel is the only documentation of PRR in red raspberry at the microscopic level. Additional studies at the microscopic level to elucidate the interaction between the pathogen and roots would be helpful in understanding the resistance derived from ‘Latham.’

Documentation of ‘Latham’ resistance is, however, growing at the genetic level. Pattison (2004) developed a hydroponic system to screen raspberry plants for resistance to PRR (Pattison et al. 2004). This hydroponic system was subsequently used to study the inheritance of resistance of PRR and to screen plants for resistance in mapping studies. Segregation ratios were examined in several populations derived from NY00-34 (‘Titan’ x ’Latham’) and a dominant two-gene model was proposed that explains major components of the resistance in ‘Latham’ (Pattison et al. 2007). Under this model ‘Latham’ is thought to be homozygous for a dominant resistance gene at one locus and heterozygous at another locus.

Because many cases of resistance are linked to dominant resistance genes (R genes) a logical next step for Pattison was to search for a polymorphism in an R-gene that segregated with resistance to PRR. R genes are typically NB-LRR [Nucleotide Binding (NB) and Luecine Rich Repeat (LRR)] resistance genes and are the basis of the gene-for-gene hypothesis (Flor 1971) wherein the plant contains a resistance gene corresponding to a given pathogen effector. Now that the mechanisms of gene-for-gene resistance is known this type resistance is commonly referred to as effector triggered immunity (ETI) (Jones & Dangl 2006; Boller & He 2009).

In mapping, Pattison (2007) identified polymorphisms in R genes (resistance gene analog polymorphisms or RGAPs) with previously developed degenerate primers (X. Chen et al. 1998) that were based on conserved R gene motifs. In addition to these specific markers, numerous anonymous markers such as randomly amplified polymorphic DNA (RAPD) markers and amplified fragment length polymorphisms (AFLPs) were used for linkage mapping and QTL analysis in a segregating BC_1 population ((‘Titan’ x ‘Latham’) x ‘Titan’). In the final map, few RGAPs were in close proximity to the two important loci identified by QTL analysis (Pattison et al. 2007), but many additional R genes may have been missed due to sequence divergence at the

degenerate primer annealing sites or because segregating polymorphisms were not found in the sequences. In total, Pattison and Samuelian examined only 75 R genes (Pattison et al. 2007; Samuelian et al. 2008), which was an exhaustive effort for the method. However, there are likely to be many more R-genes in raspberry (perhaps as many as 400) given that the genome size of *R. idaeus* is over 280MB (Arumuganathan & Earle 1991) and the small ~ 145MB genome of *A. thaliana* (Arumuganathan & Earle 1991) is known to contain over 200 NBS-LRR genes (Meyers et al. 2003).

Another observation made by Pattison was that roots of 'Latham' appear to be initially somewhat susceptible to PRR. However, after a period of susceptibility (where roots die), new roots emerge and remain healthy despite the continued presence of the pathogen (Pattison et al. 2004; Pattison 2004). This suggests that there is a change in gene expression following exposure to *P. rubi* that limits infection and spread in these new roots. It also fits well with observations made by Laun et al. (1997) that roots of the moderately resistant 'Winklers Samling' were characterized by strong root regeneration. Laun et al. (1997) noted that following initial response to the pathogen an "intense regeneration of unaffected roots was visible" in later growth of 'Winklers Samling.' This suggests that 'Latham' and 'Winklers Samling' may share some common mechanisms of resistance and that both may be linked to changes in gene expression that allow for more vigorous root growth with limited spread of the pathogen.

This change in gene expression could be due to recognition of pathogen effectors by R-genes or could be due to recognition of essential components of the pathogen itself. Recognition of pathogen effectors, which is typically indirect (Arabidopsis Interactome Mapping Consortium 2011) usually results from changes in gene expression that lead to the hypersensitive response (HR) (Jones & Dangl 2006). HR is thought to be an effective defense against biotrophic

pathogens that require living cells for suppression of plant immune responses. However, *Phytophthora* is known to occasionally escape the HR and establish the needed biotrophic interaction (Kamoun et al. 1999; Y. Chen & Halterman 2011; Vleeshouwers et al. 2000).

Plants have evolved other mechanisms of defense that work in tandem with HR. One example is activation of the salicylic acid defense pathway. Synthesis of salicylic acid is often seen following effector recognition and HR. Increased production of salicylic acid leads to a change in redox status that breaks an NPR1 dimer in the cytoplasm and allows NPR1 monomers to enter the nucleus (Mou et al. 2003). Following entry into the nucleus these NPR1 monomers interact with TGA transcription factors (Zhou et al. 2000). This results in activation of defense responses such as increased expression of pathogenesis related genes (PR genes).

PR genes are a diverse group, and can be divided into 17 families based on function (van Loon et al. 2006). In the Rosaceae family the PR-10 genes have been extensively studied because they are associated with plant defense and because of their role in food allergies (Ebner et al. 1991; Gao et al. 2005). The induction of PR-10 genes is possible through either the jasmonic acid (JA) pathway or the salicylic acid (SA) pathway (Park et al. 2004; McGee & Hamer 2001). In apple leaves it has been shown that several classes of PR-10 are inducible with salicylic acid (Ziadi et al. 2001). Again, in apple, PR-10 was induced after inoculation with the hemibiotrophic pathogen *Venturia inequalis* (apple scab) in a resistant apple cultivar after an "intense necrotic resistance reaction of leaf cells" (Chevalier et al. 2008).

The function of PR-10 proteins is not completely understood, but several studies suggest ribonuclease activity, which defends against viruses, and antimicrobial activity with an unknown mode of action against bacteria, fungi, and oomycetes. Recombinant *Capsicum annuum* (hot pepper) PR-10 was effective against both Tobacco Mosaic Virus (TMV) and against

Phytophthora capsici in culture (Park et al. 2004). In this study, PR-10's ribonuclease activity was tested. It was suggested that PR-10 functioned to cleave invading viral RNA, although a mechanism by which the recombinant PR-10 was inhibiting *P. capsici* in culture was not proposed. In another study, a PR-10 from *Oxalis tuberosa* Mol. was tested against various plant pathogens and was found to inhibit the growth of *P. cinnamomi* in a dose-dependent fashion, but it failed to inhibit growth of hyphae in six other *Phytophthora* species at the concentrations tested (Flores et al. 2002). Interestingly, *P. cinnamomi* and *P. rubi* are closely related, both being members of clade 7 of a recent phylogenetic study of *Phytophthoras* (Blair et al. 2008) while the other species tested by Flores et al. (2002) belong to other clades.

An R-gene independent response could also be important in defense against *P. rubi* and could be triggered by pathogen-associated molecular patterns (PAMPs) (Medzhitov & Janeway 1997). Janeway (1989) proposed that conserved microbial patterns could be detected by induced immune responses (Medzhitov 2009; Janeway 1989). These PAMPs are considered to be conserved, surface exposed, and indispensible parts of microbes (Medzhitov & Janeway 1997). Examples of PAMPs include bacterial flagellin and chitin in fungi. There are also a number of PAMPs that are characterized in oomycetes including beta-glucans (Sharp & Valent 1984) which are essential components of oomycete cell walls and also calcium-dependent cell wall transglutaminase (Brunner et al. 2002).

In plants, PAMPs are typically detected by transmembrane pattern recognition receptors (TPRRs) that are successful at transmitting extracellular detection of the pathogen into cells, resulting in PAMP triggered immunity (PTI) (Boller 2009; Jones & Dangl 2006). PTI is thought to have similar, but slower outcomes compared to ETI. However, PTI is not linked to HR.

While many TPRRs and PAMP pairs have been identified for bacteria and fungi there are only limited cases of TPRRs that are specific to *Phytophthora* PAMPS. The receptor for detection of beta-glucan in soybean is one example, but it defies the typical structure of other TPRRs identified to date. Rather than having a transmembrane domain it may instead indirectly associate with other TPRRs on the cell membrane (Boller & He 2009). The lack of characterized TPRRs for *Phytophthora*-specific PAMPs would make it difficult to identify polymorphisms in these genes that may be contributing to resistance in raspberry. These extracellular TPRRs target highly conserved portions of microbes and therefore it is unlikely that there would be much functional diversity in these genes within raspberry. Instead, changes in how individual genotypes respond following recognition of PAMPs are likely to be important. Some ellicitors trigger mild defense responses when applied to a plant while others, such as the small cystein-rich elicitors, can trigger necrosis and other defense responses (Kamoun 2006), that is thought to be distinct from HR.

Whether arising from ETI or PTI it has been suggested that responding to pathogen attack is potentially metabolically expensive (Tian et al. 2003; Bolton 2009; Zangerl et al. 1997; Smedegaard-Petersen & Tolstrup 1985; Heil et al. 2000) at least under field conditions (Heidel et al. 2004). This is particularly the case in situations where increased components of the cell wall may play a role in disease resistance (Niemann et al. 1992; Heidel et al. 2004). Examples of components added to the cell wall in response to pathogens include callose deposition and lignification.

1.6. Review of Methods for Gene Expression Analysis

There are numerous potential technologies available for differential expression analysis, these range from the many old technologies that are still considered effective to the computationally intensive new technologies that rely on sequencing. One early approach was Differential Display Reverse Transcriptase Polymerase Chain Reaction (DDRT-PCR) (Liang and Pardee 1992). In DDRT-PCR, as originally described, one selective primer composed of 5'T_{11} plus two extra bases and one random primer of 6-7bp were used to amplify sub pools of mRNA that could be differentially visualized on sequencing gels. However, DDRT-PCR commonly yielded results that were not reproducible (Malhotra et al. 1998). Other widely used technologies capable of analyzing the expression of multiple genes include the cDNA microarray (Schena et al. 1995) and Serial Analysis of Gene Expression (SAGE) (Velculescu et al. 1995).

The cDNA microarray generates a continuous data structure because it works by hybridization of fluorescently labeled cDNA followed by fluorescent detection and thus generating a relative abundance estimate. In SAGE actual counts of transcripts are generated from sequencing libraries created by restriction digestion of cDNA followed by a series of steps that results in a concatenated string of fragments that are sequenced with the Sanger method and subsequently, informatically separated and mapped to a genome or set of expressed sequence tags (ESTs). For non-model organisms the method of cDNA AFLP was pioneered about the same time as the microarray (Bachem et al. 1996); (Vuylsteke et al. 2007) and as used along with DDRT-PCR, because these methods require the least amount of a prior knowledge about the genome.

Another technology that was is similar to SAGE is Massively Parallel Signature Sequencing (MPSS) (Brenner et al. 2000) and like SAGE, MPSS relies on very short read counts (initially 32mers) that could be aligned to a reference genome or ESTs. The advantage of MPSS

over other SAGE methods was that the dramatic increase in sequencing depth (Brenner et al. 2000). MPSS library preparation, however, required a complex cloning stage and the technology required for sequencing was not made widely available. Together, these count based methods set the stage for current methods in transcriptomics by influencing ideas for new molecular methods and computational advances that are now revolutionizing the analysis of transcriptomes.

There are three dominant forms of sequencing in use currently for transcriptiome analysis, including Illumina's (Illumina, San Diego, CA) method of highly parallel sequencing by synthesis (Illumina/Solexa sequencing), Applied Biosystems' (Applied Biosystems, Foster City, CA) SOLiD sequencing by ligation, and Roche's 454 pyrosequencing (Roche/454 Life Sciences, Branford, CT). Illumina and SOLiD sequencing produce short, but high depth data sets typically ranging in length from 36nt to 150nt in length. Roche's 454 sequencing reads have the advantage that they are typically much longer, but are more expensive per base and produce data sets that are typically low depth compared to short-read technologies because of cost constraints.

At the most basic level there are now two paradigms for transcriptome analysis from next generation sequencing technologies. These are the alignment-first method and the assembly-first method (Haas and Zody 2010). Both are relevant for consideration in non-model organism transcriptome analysis. The alignment-first method is completely dependent upon the existence of a reference genome while the assemble-first method is dependent on a reference genome only if identification of intron-exon junctions is required for the experimental goals. For non-model organisms both approaches are in the early days of development and there are significant computational challenges ahead. Despite this these methods are extremely powerful and allow researchers in non-model organisms to generate specific and testable molecular-level hypotheses

about complex problems such as the dissection of metabolic pathways or plant-disease interactions.

1.6.1. Alignment-first method

At the moment no method exists that was specifically designed for an alignment-first approach when a genome other than the true reference genome is used. The use of this method is still considered experimental, but this approach presents a convenient statistical framework when the goals of an experiment are hypothesis generation in a complex system. An alignment-first method takes advantage of predefined gene-space (i.e. the gene models of the related organism), which allows for direct comparison between treatments and controls and between replicates. The alignment-first method utilizes software packages including Scripture (Guttman et al. 2010), Cufflinks (Trapnell et al. 2010), and several R packages such as DEGseq (Wang et al. 2009), edgeR (Robinson et al. 2010), and a Two-Stage Poison Model (Auer and Doerge 2011). Each of these alignment-first algorithms depends upon on tools like Bowtie (Langmead et al. 2009) and/or Tophat (Trapnell et al. 2009) for initial mapping of reads to the reference genome.

1.6.2. Assemble-first Method

Transcriptome assemblers are often built on the backbone of a genome assembler or are simply the genome assembler itself with altered parameters. In the case of the later, the use of the assembler for transcriptome assembly was secondary which can result in overly complex analysis pipelines. Examples of genome assemblers adapted for the assembly of transcriptomes include miraEST (Chevreux et al. 2004), Trans-ABySS (Robertson et al. 2010), Oases (unpublished: http://www.ebi.ac.uk/~zerbino/oases/), and Rnnotator (J. Martin et al. 2010). Much progress has been made recently with regards to the de novo transcriptome assembly

problem. For instance, the Trinity assembler (Grabherr et al. 2011) represents an advance for non-model organisms because it was designed specifically to deal with the problem of true de novo transcriptome assembly directly from RNA-seq data.

1.7. References

Arabidopsis Interactome Mapping Consortium, 2011. Evidence for network evolution in an Arabidopsis interactome map. Science, 333(6042), pp.601–607.

Arumuganathan, K. and Earle, E., 1991. Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter, 9(3), pp.208–218.

Auer, P.L. and Doerge, R.W., 2011. A Two-Stage Poisson Model for Testing RNA-Seq Data. Statistical Applications in Genetics and *Molecular Biology*, 10(1).

Bachem, C.W. et al., 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant Journal*, 9(5), pp.745–753.

Blair, J.E. et al., 2008. A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. *Fungal genetics and biology*, 45(3), pp.266–277.

Boller, T., 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology*, 60, pp.379–406.

Boller, T. and He, S.Y., 2009. Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. *Science*, 324(5928), pp.742–744.

Bolton, M.D., 2009. Primary Metabolism and Plant Defense—Fuel for the Fire. *Molecular Plant-Microbe Interactions*, 22(5), pp.487–497.

Bonants, P. et al., 1997. Detection and identification of Phytophthora fragariae Hickman by the polymerase chain reaction. *European Journal of Plant Pathology*, 103(4), pp.345–355.

Brenner, S. et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 18(6), pp.630–634.

Brunner, F. et al., 2002. Pep-13, a plant defense-inducing pathogen-associated pattern from Phytophthora transglutaminases. *The EMBO journal*, 21(24), pp.6681–6688.

Brunner-Keinath, S. and Seemueller, E., 1992. On the diagnosis of Phytophthora root rot of raspberry. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, 44(8), pp.179–182.

Chen, X., Line, R. and Leung, H., 1998. Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. *Theoretical And Applied Genetics*, 97(3), pp.345–355.

Chen, Y. and Halterman, D.A., 2011. Phenotypic characterization of potato late blight resistance mediated by the broad-spectrum resistance gene RB. *Phytopathology*, 101(2), pp.263–270.

Chevalier, M. et al., 2008. Specific activation of PR-10 pathogenesis-related genes in apple by an incompatible race of Venturia inaequalis. *Biologia Plantarum*, 52(4), pp.718–722.

Chevreux, B. et al., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research*, 14(6), pp.1147–1159.

Converse, R. and Schwartze, C., 1968. A root rot of red raspberry caused by Phytophthora erythroseptica. *Phytopathology*, 58(1), pp.56–59.

Converse, R.A.S.C., 1965. Phytophthora sp. from Washington pathogenic on roots of red raspberry. Report and Abstracts of the 1965 Annual Meeting of the Potomac Division of the American Phytopathological Society, p.1.

Cooke, D. and Duncan, J., 1997. Phylogenetic analysis of Phytophthora species based on ITS1 and ITS2 sequences of the ribosomal RNA gene repeat. *Mycological Research*, 101, pp.667–

677.

Crandall, P., 1977. Red raspberry root rot research [Phytophthora erithroseptica]. *Annual Report - Proceedings of the Oregon Horticultural Society*, 68, p.57.

Duncan, J., Kennedy, D. and Seemueller, E., 1987. Identities and pathogenicities of Phytophthora spp. causing root rot of red raspberry. *Plant Pathology*, 36(3), pp.276–289.

Ebner, C. et al., 1991. Common epitopes of birch pollen and apples--studies by western and northern blot. *The Journal of allergy and clinical immunology*, 88(4), pp.588–594.

Flor, H., 1971. Current status of the gene-for-gene concept. *Annual Review of Phytopathology*.

Flores, T. et al., 2002. Ocatin. A novel tuber storage protein from the andean tuber crop oca with antibacterial and antifungal activities. *Plant Physiology*, 128(4), pp.1291–1302.

Forster, H., 2000. Phylogenetic relationships of Phytophthora species based on ribosomal ITS I DNA sequence analysis with emphasis on Waterhouse groups V and VI. *Mycological Research*, 104(9), pp.1055–1061.

Gao, Z., de Weg, W. and Schaart, J., 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (Malus domestica). *Theoretical and Applied Genetics*. 111, pp. 171–183

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652.

Grattapaglia, D. and Bertolucci, F., 1995. Genetic mapping of QTLs controlling vegetative propagation in Eucalyptus grandis and E. urophylla using a pseudo-testcross strategy and RAPD markers. *Theoretical and Applied Genetics*. 90, pp. 933-947**.**

Grattapaglia, D. and Sederoff, R., 1994. Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers.

Genetics, 137(4), pp.1121–1137.

Guttman, M. et al., 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnology*, 28(5), pp.503–510.

Haas, B.J. and Zody, M.C., 2010. Advancing RNA-Seq analysis. *Nature Biotechnology*, 28(5), pp.421–423.

Hardham, A., 2005. Phytophthora cinnamomi. *Molecular Plant Pathology*, 6(6), pp.589–604.

Heiberg, N., 1995. Control of Root-Rot of Red Raspberries Caused by Phytophthora-Fragariae Var Rubi. *Plant Pathology*, 44(1), pp.153–159.

Heidel, A.J. et al., 2004. Fitness costs of mutations affecting the systemic acquired resistance pathway in Arabidopsis thaliana. *Genetics*, 168(4), pp.2197–2206.

Heil, M., Hilpert, A. and Kaiser, W., 2000. Reduced growth and seed set following chemical induction of pathogen defence: does systemic acquired resistance (SAR) incur allocation costs? *Journal of Ecology*, 88, pp.645–654.

Janeway, C.A., 1989. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor symposia on quantitative biology*, 54 Pt 1, pp.1–13.

Jeffers, S.N. and Martin, S., 1986. Comparison of 2 Media Selective for Phytophthora and Pythium Species. *Plant Disease*, 70(11), pp.1038–1043.

Jones, J.D.G. and Dangl, J.L., 2006. The plant immune system. *Nature*, 444, pp.323–329.

Kamoun, S., 2006. A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. Annual Review of Phytopathology, 44, pp.41–60.

Kamoun, S., Huitema, E. and Vleeshouwers, V., 1999. Resistance to oomycetes: a general role for the hypersensitive response? *Trends in Plant Science*, 4(5), pp.196–200.

Keep, E., 1969. Dwarfing in the raspberry, Rubus idaeus L. *Euphytica*, 18, pp.256–276.

Keep, E., 1968. Incompatibility in Rubus with Special Reference to R. Idaeus L. *Can. J. Genet. Cytol.*10, pp.253–262.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), p.R25.

Laun, N. and Zinkernagel, V., 1997. A comparison of the resistance against Phytophthora fragariae var. rubi, the causal agent of a raspberry root rot. *Journal of Phytopathology-Phytopathologische Zeitschrift*, 145(5-6), pp.197–204.

Liang, P. and Pardee, A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257(5072), pp.967–971.

Malhotra, K. et al., 1998. Interaction and effect of annealing temperature on primers used in differential display RT-PCR. *Nucleic Acids Research*, 26(3), pp.854–856.

Maliepaard, C. and Jansen, J., 1997. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genetical Research*. 70, *pp.* 237–250.

Maloney, K. et al., 2005. Suppression of phytophthora root rot in red raspberries with cultural practices and soil amendments. *Hortscience*, 40(6), pp.1790–1795.

Maloney, K., Wilcox, W. and Sanford, J., 1993. Raised Beds and Metalaxyl for Controlling Phytophthora Root-Rot of Raspberry. *Hortscience*, 28(11), pp.1106–1108.

Martin, J. et al., 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics*, 11, p.663.

McGee, J. and Hamer, J., 2001. Characterization of a PR-10 pathogenesis-related gene family induced in rice during infection with Magnaporthe grisea. *Molecular Plant-Microbe Interactions*. 14(7) pp. 877–886.

Medzhitov, R., 2009. Approaching the asymptote: 20 years later. *Immunity*, 30(6), pp.766–775.

Medzhitov, R. and Janeway, C.A., 1997. Innate immunity: impact on the adaptive immune response. *Current Opinion in Immunology*, 9(1), pp.4–9.

Meyers, B.C. et al., 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*, 15(4), pp.809–834.

Mou, Z., Fan, W. and Dong, X., 2003. Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. *Cell*, 113(7), pp.935–944.

Niemann, G.J. et al., 1992. Differences in relative growth rate in 11 grasses correlate with differences in chemical composition as determined by pyrolysis mass spectrometry. *Oecologia*, 89(4), pp.567–573.

Olsson, C. and Heiberg, N., 1997. Sensitivity of the ELISA test to detect Phytophthora fragariae var. rubi in raspberry roots. *Journal of Phytopathology-Phytopathologische Zeitschrift*, 145(7), pp.285–288.

Park, C.-J. et al., 2004. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant Journal*, 37(2), pp.186–198.

Pattison, J.A., 2004. Investigating the Inheritance of Phytophthora Root Rot (Phytophthora fragariae var. rubi) Resistance in Red Raspberry (Rubus Idaeus L.) Using Molecular Markers. Unpublished dissertation. Cornell University.

Pattison, J.A., Samuelian, S.K. and Weber, C.A., 2007. Inheritance of Phytophthora root rot resistance in red raspberry determined by generation means and molecular linkage analysis. *Theoretical And Applied Genetics*, 115(2), pp.225–236.

Pattison, J.A., Wilcox, W. and Weber, C.A., 2004. Assessing the resistance of red raspberry (Rubus idaeus L.) genotypes to Phytophthora fragariae var. rubi in hydroponic culture.

Hortscience, 39(7), pp.1553–1556.

Potter, D. et al., 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution*. 266: pp.5–43.

Ritter, E., Gebhardt, C. and Salamini, F., 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, 125(3), pp.645–654.

Robertson, G. et al., 2010. De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), pp.909–912.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.

Samuelian, S.K. et al., 2008. Isolation and linkage mapping of NBS-LRR resistance gene analogs in red raspberry (Rubus idaeus L.) *Tree Genetics and Genomes*. 4, pp. 881–896.

Schena, M. et al., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235), pp.467–470.

Sharp, J. and Valent, B., 1984. Purification and partial characterization of a beta-glucan fragment that elicits phytoalexin accumulation in soybean. *Journal of Biological Chemistry*. 259(18), pp**.** 11312-11320.

Shoemaker, J. 1934. Small-fruit culture; a text for instruction and reference work and a guide for field practice. P. Blakiston's son & co., inc. p 434.

Smedegaard-Petersen, V. and Tolstrup, K., 1985. The Limiting Effect of Disease Resistance on Yield. *Annual Review of Phytopathology*, 23(1), pp.475–490.

Stammler, G., Seemueller, E. and Duncan, J., 1993. Analysis of Rflps in Nuclear and

Mitochondrial-DNA and the Taxonomy of Phytophthora-Fragariae. *Mycological Research*, 97, pp.150–156.

t Veld, W., 2007. Gene flow analysis demonstrates that Phytophthora fragariae var. rubi constitutes a distinct species, Phytophthora rubi comb. nov. *Mycologia*, 99(2), pp.222–226.

Thompson, E.M., 1995. Chromosome Numbers of *Rubus* Species at the National Clonal Germplasm Repository. *Hortscience*, 30(7), pp.1447–1452.

Tian, D. et al., 2003. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature*, 423(6935), pp.74–77.

Trapnell, C. et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), pp.511–515.

Trapnell, C., Pachter, L. and Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–1111.

van Loon, L.C., Rep, M. and Pieterse, C., 2006. Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology*, 44, pp.135–162.

Velculescu, V., Zhang, L. and Vogelstein, B., 1995. Serial analysis of gene expression. *Science*, 270(5235), pp.484–487.

Vleeshouwers, V. et al., 2000. The hypersensitive response is associated with host and nonhost resistance to Phytophthora infestans. *Planta*, 210, pp.853–864.

Vuylsteke, M., Peleman, J. and van Eijk, M.J.T., 2007. AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. *Nature Protocols*, 2(6), pp.1399–1413.

Wang, L. et al., 2009. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26(1), pp.136–138.

Waterston, J., 1936. A Note on the Association of a species of Phytophthora with a "Die-Back" Disease of the Raspberry. *Transactions of the Botanical Society of Edinburgh*, XXXII(II), p.7.

Wilcox, W., 1989a. Identity, Virulence, and Isolation Frequency of 7 Phytophthora Spp Causing Root-Rot of Raspberry in New-York. *Phytopathology*, 79(1), pp.93–101.

Wilcox, W., 1989b. Identity, virulence, and isolation frequency of seven Phytophthora spp. causing root rot of raspberry in New York. *Phytopathology*, 79(1), pp.93–101.

Wilcox, W., 1992. Phytophthora Root and Crown Rots. *Disease Identification Sheet No. 7*.

Wilcox, W. et al., 1993. Identity of a Phytophthora species attacking raspberry in Europe and North America. *Mycological Research*, 97, pp.817–831.

Wilcox, W., Pritts, M. and Kelly, M., 1999. Integrated control of Phytophthora root rot of red raspberry. *Plant Disease*, 83(12), pp.1149–1154.

Zangerl, A.R., Arntz, A.M. and Berenbaum, M.R., 1997. Physiological price of an induced chemical defense: photosynthesis, respiration, biosynthesis, and growth. *Oecologia*, 109(3), pp.433–441.

Zhou, J.M. et al., 2000. NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid. *Molecular Plant-Microbe Interactions*, 13(2), pp.191–202.

Ziadi, S. et al., 2001. Characterization in apple leaves of two subclasses of PR-10 transcripts inducible by acibenzolar-S-methyl, a functional analogue of salicylic acid. *Physiological and Molecular Plant Pathology*. 59, pp. 33-43.

Chapter 2

STRATEGIES FOR TRANSCRIPTOME ANALYSIS IN NON-MODEL PLANTS

2.1. Abstract

Even with recent reductions in sequencing costs most plants lack the genomic resources required for successful short-read transcriptome analyses as performed routinely in model species. Several approaches for the analysis of short read transcriptome data are reviewed for non-model species for which the genome of a close relative is utilized in place of a true reference genome, including multiple methods for de novo assembly and RNA-Seq. Two approaches using a data set from *Phytophthora*-challenged *Rubus idaeus* L. (red raspberry) are compared. RNA-Seq was performed with over 70,000,000 86nt Illumina reads derived from *R. idaeus* L. roots using publicly available informatics tools (Bowtie/Tophat and Cufflinks). RNA-Seq identified 16,956 putatively expressed genes. De novo assembly was performed with the same data set and a publicly available transcriptome assembler (Trinity). A BLAST search with a maximum e-value threshold of 1.0×10^{-3} revealed that over 36,000 transcripts had matches to plants and over 500 to *Phytophthora.* Gene expression estimates from RNA-Seq and de novo assembly were compared for raspberry (Pearson's correlation = 0.730). Together, RNA-Seq and de novo assembly constitute a powerful method of transcriptome analysis in non-model organisms. RNA-Seq provides a framework for differential expression testing if alignments are made to the predefined gene-space of a close relative and de novo assembly provides a more robust method of identifying unique sequences and sequences from other organisms in a system. The use of

these methods is considered experimental in non-model systems, but can be used to generate resources and specific testable hypotheses.

2.2. Introduction

A transcriptome is a snapshot of the gene expression in a given cell or tissue at a given moment provided by capturing the total mRNA within that tissue. This view into the gene action within a cell or tissue at a particular moment in time represents not only the expression of active genes, but also the combination of all isoform sequences (produced through alternative splicing and variant alleles) within the cells. Utilizing appropriate approaches for producing transcript libraries and analysis tools, the variation in gene expression induced by changing environmental conditions (i.e. before and after biotic or abiotic stress), changes in developmental stages (i.e. vegetative growth vs. flowering) or tissue types (i.e. roots vs. leaves) can be examined. These techniques have become more feasible in non-model organisms as automation and efficiency has reduced the cost of high throughput sequencing. Analyzing the complexity of a transcriptome is now possible in any organism and provides a quantifiable and robust method for investigating gene expression in specific tissues.

2.2.1. The evolution of transcriptome analysis

Along with gene expression analysis, the sequencing of cDNA has been an often used method for characterizing the transcriptome in plants because it represents a small, but information rich target compared to the full genome. Technologies for simultaneous analysis of both the sequence and expression of multiple genes were slow to develop, but are now becoming widely accessible. Differential Display Reverse Transcriptase Polymerase Chain Reaction (DDRT-PCR) (Liang & Pardee 1992) was one of the first methods for analysis of differential

gene expression on the scale of the full transcriptome and numerous genes were identified and cloned with this approach. DDRT-PCR uses one selective primer to amplify sub pools of mRNA that can be differentially visualized on sequencing gels. Despite the many successes of DDRT-PCR, the approach is labor intensive and can yield results that are not reproducible (Malhotra et al. 1998).

Additional technologies capable of analyzing the expression of multiple genes have been developed to address the growing interest in transcriptome analysis. Two technologies developed, cDNA microarrays (Schena et al. 1995) and Serial Analysis of Gene Expression (SAGE) (Velculescu et al. 1995), approached the detection of gene expression with completely different perspectives. In cDNA microarrays, fluorescently labeled cDNA from the organism of interest is hybridized to a DNA probe on a chip and relative fluorescence is detected. This data is continuous and does not directly give information about sequence variation. In SAGE, a library of joined cDNA fragments is generated and sequenced with the Sanger method (Velculescu et al. 1995),. The resulting sequences are separated and mapped to a genome or set of Expressed Sequence Tags (ESTs) (Sanger sequenced fragments of cloned cDNA (Adams et al. 1991)). This results in actual counts of aligned sequences or tags for a given gene. Initially, both cDNA microarrays and SAGE were very limited by the available technology. For instance, the first miniaturized microarray for plants contained just 45 *Arabidopsis* genes (Schena et al. 1995). The analysis of SAGE data was limited by sequencing cost, labor, and the availability of genome sequence or ESTs. Later advances in miniaturization allowed for the open reading frames (ORFs) from the entire yeast genome to be placed on just a few slides (Lashkari et al. 1997) and so that microarray analysis of transcriptomes has become a routine part of work in model organisms and some non-model organisms. For organisms with very limited resources, the

method of cDNA AFLP (Bachem et al. 1996) was pioneered about the same time as the microarray and has been used extensively in non-model species (Vuylsteke et al. 2007), along with DDRT-PCR, because these methods required the least amount of prior genome knowledge.

Massively Parallel Signature Sequencing (MPSS) was later developed (Brenner et al. 2000), which produced data similar to that of SAGE, yielding very short sequencing reads (initially 32mers) that could be aligned to a reference genome or set of ESTs. MPSS was promising because of an increase in sequencing depth compared to previous SAGE methods. However, the library preparation for MPSS required a complex bead cloning stage and the technology was never made widely available. Together, these count-based methods set the stage for current methods in transcriptome analysis by influencing ideas for new molecular methods and new computational advances that are now making the analysis of transcriptomes more feasible in non-model organisms.

New sequencing technologies continue to emerge and have been discussed elsewhere (Metzker 2010) along with many in-depth discussions of their potential impact beyond the analysis of transcriptomes (Davey et al. 2011; Schneeberger & Weigel 2011). Here the focus is on transcriptome analysis methods utilized in conjunction with three of the dominant next-generation sequencing technologies now widely in use around the world; Illumina's method of highly parallel sequencing by synthesis (Illumina, San Diego, CA); Applied Biosystem's SOLiD sequencing by ligation (Applied Biosystems, Foster City, CA); and Roche 454's pyrosequencing (Roche/454 Life Sciences, Branford, CT).

2.3. Sequencing and Analysis Choice

The choice of sequencing technology and data analysis approach is critical to the success of an experiment. All three of the above sequencing technologies produce a tremendous volume of high quality data, but each has specific practical applications. Illumina and SOLiD sequencing produce short, but high depth data sets. For the Illumina sequencing, the user can currently select the length of reads in the range of 36nt to 150nt that can be sequenced either from one end of a DNA fragment (single end reads) or from both ends of a DNA fragment (paired end reads). Longer reads and paired-end reads are typically selected in de novo assembly projects, but shorter reads are sometimes chosen for alignment to a reference genome. The confidence score for a given base in a sequence declines as the read length grows, which can hinder alignment and downstream analysis. The data from Illumina sequencing reads are represented as actual nucleotide sequence and the analysis can proceed directly to the alignment to a reference genome or to de novo assembly.

In the SOLiD system, the user can currently choose read lengths of 35nt to 75nt in length in either single end or paired end format. The SOLiD system sequences two bases at a time (thus there are 16 possible combinations to query) and any single base must be sequenced twice to identify the true sequence at a single position. This method is thought to improve the identification of sequencing errors in post-sequencing analysis. However, for researchers without a reference genome this 2-base encoding system is a drawback, because the resulting sequence is numerically encoded and will not be immediately recognizable to a biologist. It is only through subsequent analysis that biological relevance of a SOLiD sequencing read is restored. Usually SOLiD reads are aligned in their 2-base encoded format (so called “color space” format) to a 2-base encoding genome to convert the sequence back to nucleotide space, but without a reference genome it can require additional informatics to make sense of the sequencing. Direct conversion

of sequencing reads is possible, but all bases following a single error in color space will create errors in all subsequent bases of a read and this is not recommended. The reader is referred to the manufacturer's website for a more detailed explanation of double encoding. If a researcher in a non-model system chooses to use the SOLiD system the genome of a close relative may be the most direct option for downstream analysis.

The read length distribution of Illumina & SOLiD systems is very uniform and most reads are exactly the length requested by the researcher. In Roche's 454 sequencing, reads have a wider sequence length distribution and 454 reads are also encoded in normal nucleotide space. Most 454 reads are now longer than 500nt, with a mode around 700nt and a maximum length of over 1000nt. The long reads of the 454 sequencer typically result in high quality transcriptome assemblies, but these data sets are much lower depth per sequencing dollar spent. The analysis of high depth short-read data is fundamentally different from the analysis of low depth long reads and therefore the computational resources and analysis approaches differ greatly depending upon the choice of platform.

2.4. Sequencing Library Preparation

2.4.1. RNA quality & Fragmentation

The most critical step in the preparation of a transcriptome sequencing library is obtaining a high quality mRNA sample that has been depleted of rRNA. To remove rRNA a researcher can use any one of the commercially available kits for a poly (A) capture or other hybridization-based depletion methods leaving a sample highly enriched with mRNA. Next, the quality of an mRNA sample is gauged by analyzing the sample on denaturing formaldehyde gels for insense 18s and 28s rRNA bands and a high molecular weight smear representing the mRNA.

Lack of bright 18s and 28s bands or an abundance of low molecular weight smearing indicates sample degradation. Most protocols also call for the analysis of an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA) electropherogram. In this case the researcher checks the electropherogram for the same 18s and 28s rRNA peaks and a distribution of high molecular weight mRNA, but is also aided by Agilent's software, which compares a number of RNA integrity metrics and provides an RNA Integrity Number (RIN). Samples with RINs greater than eight are typically considered to be high quality. Finally, the fragmentation of mRNA or cDNA is performed with either nebulization or divalent cations under elevated pressure according to the protocols of the manufacturers kit. Typically, RNA is fragmented prior to conversion to cDNA in short read sequencing technologies to increase the uniformity of coverage across the transcript. In a 454-library, preparation the fragmentation is performed after normalization (equalization of transcript abundance so that highly expressed transcripts are not the only sequence obtained) of cDNA. Normalization is typically performed in 454 sequencing because libraries are not sequenced to as much depth as with Illumina or SOLiD. If libraries are not normalized the researcher risks obtaining sequence from only the most highly expressed genes. There are a number of commercial kits available for normalization of cDNA for this purpose. It is best practice to check with the manufacturer of the sequencing instrument for the latest protocols prior to library preparation.

2.5. Current transcriptome analysis approaches

At the most basic level there are now two approaches for transcriptome analysis using next generation sequencing technologies, align-then-assemble (alignment first) and assemble-then-align (assemble first) (Haas & Zody 2010). The alignment first methods are completely

dependent upon the existence of a reference genome while the assemble first algorithms are essentially de novo assembly algorithms that are only dependent on a reference genome if the identification of intron-exon junctions is required for the experimental goals. Comparisons of an alignment first method to two variations of assemble first methods are shown in Figure 1. In a non-model system both approaches can utilize the genome of a close relative. At the moment there is no clear answer as to which method is most appropriate for any given non-model species. A greater similarity between the gene sequences in the test genome and those of the related reference genome increases the utility of the alignment first approach has.

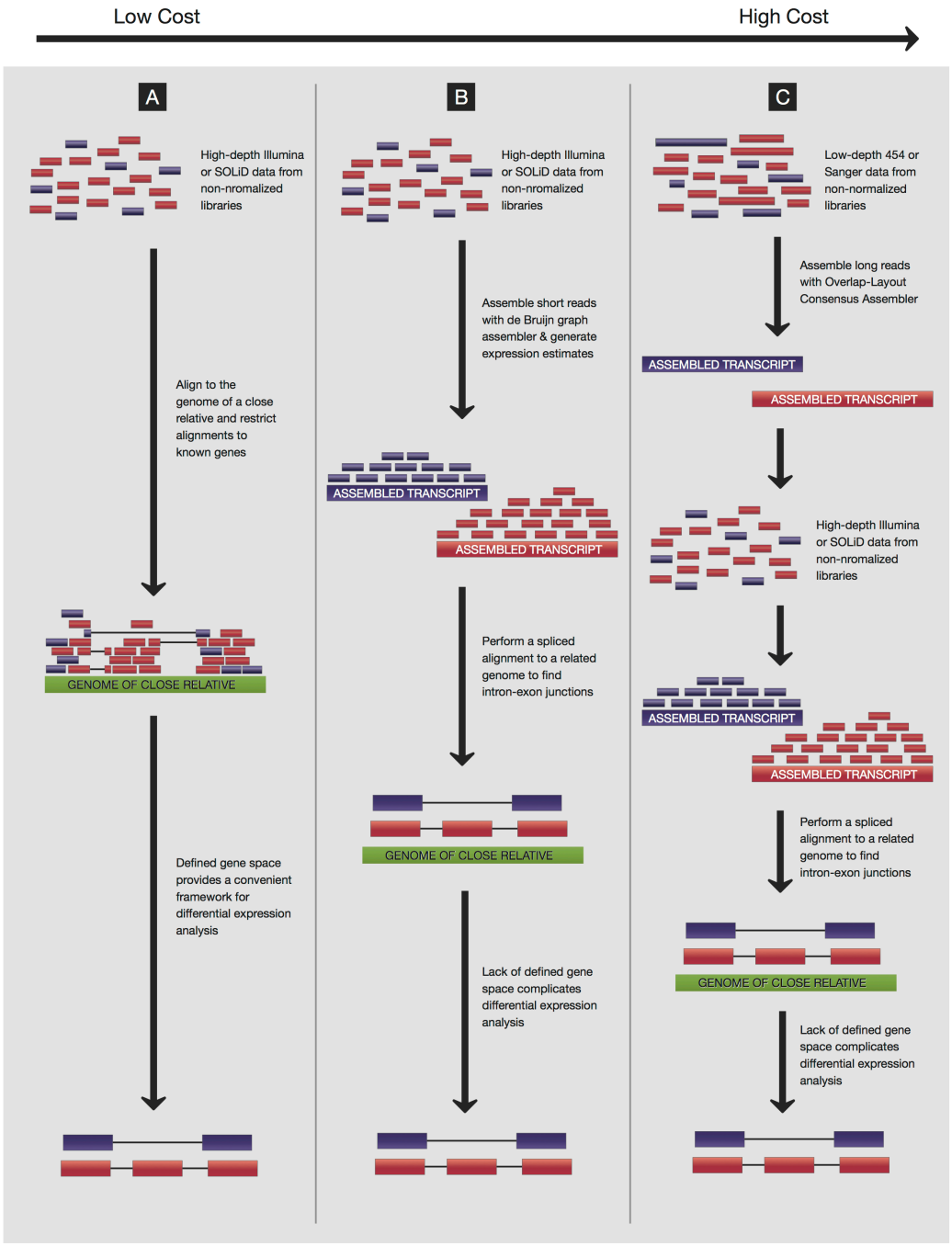

Figure 2.1. Comparison of three possible options for the analysis of the transcriptome data in non-model organisms. **1A**: Align-then-assemble: Short read data can be directly aligned to the genome of a close relative, using pre-defined gene-models from the organism's annotation. This is particularly useful for analysis of experiments because comparisons between gene models are unambiguous. However, exons and whole genes can be missed when reads fail to align. **1B**: Assemble-then-align variant 1: Short read data is assembled into transcripts in a manner that allows integrated detection of gene expression, but lack of defined gene space can result in ambiguous comparisons in the experimental context. The genome of a close relative can help define comparisons among large numbers of transcripts. The approach also requires additional sequencing depth compared to and align-then-assemble approach. **1C**: Assemble-then-align variant 2: Longer 454 reads or Sanger reads are assembled into transcripts. 454 data sets are

typically normalized prior to sequencing the initial assembly and therefore expression information is lost. Again, lack of defined gene space can result in ambiguous comparisons in the experimental context, but genome of a close relative can help define comparisons among large numbers of transcripts. The approach is the most expensive in terms of sequencing costs and computational time.

2.5.1. Align-then-assemble methods

No method yet exists that was specifically designed for an alignment to the genome of a close relative, but this approach presents a statistical framework to generate hypotheses related to a particular biological process. Alignment first methods can take advantage of predefined gene-space (i.e. the gene models of the related organism), allowing for a direct comparison between treatments and controls, and between replicates. Alignment first algorithms include Scripture (Guttman et al. 2010) and Cufflinks (Trapnell et al. 2010) and several installable packages for differential expression analysis in the R statistical software environment (a publically available statistical program). These R packages include DEGseq (L. Wang et al. 2009a), edgeR (Robinson et al. 2010) and the Two-Stage Poison Model (Auer & Doerge 2011). Each of these alignment-first methods depends upon on software tools such as Bowtie (Langmead et al. 2009) and/or Tophat (Trapnell et al. 2009) for initial mapping of sequences to the reference genome. Most of these packages also offer analysis methods for detection of differentially expressed genes in data sets with replicates (Table 2.1).

Table 2.1. Commonly used and freely available transcriptome analysis tools.

Algorithm or Pipeline Name	Type	Reference Dependent	Assembly paradigm	Supported OS	Recommended Data type
Trinity	Assembler	No	Greedy/ Eulerian	Linux	High depth
Trans-ABySS	Assembler	Either	Eulerian	Linux/Mac OS X	High depth
SOAPdenovo	Assembler	No	Eulerian	Linux	High Depth
Velvet	Assembler	No	Eulerian	Linux	High depth
Rnnotator (depends on Velvet)	Assembler	No	Eulerian	Linux	High depth
Oases (depends on Velvet)	Assembler	No	Eulerian	Linux	High depth
MIRA	Assembler	No	OLC variant	Linux	Long reads
CAP3	Assembler	No	OLC variant	Linux	Long reads
Newbler	Assembler	No	OLC variant	Linux	Long read
Cufflinks	Mapping First	Yes	mapping first	Linux/Mac OS X	high depth short reads
Scriptuture	Mapping First	Yes	mapping first	Linux	high depth short reads

For an alignment-first method there are a number of possible problems that may arise in aligning to a related genome rather than to the true reference genome of the actual test species, particularly for the short sequencing reads such as those produced by Illumina or SOLiD. Most commonly, there is poor alignment due to divergence of the genomes at the sequence level. Alignment and analysis can also be complicated when only a few exons or portions of exons are conserved between the species. Furthermore, multimapping (a read mapping to multiple

locations), especially in multigene families, can produce ambiguous or nonsensical outputs if alignment criteria are relaxed to deal with sequence divergence. Because of this, Type I error (detecting an expressed gene when it is not truly expressed) is likely when alignment to another genome is performed. If this problem is acknowledged, then interpretation can be limited to generating hypotheses at the level of the gene family rather than at the level of a specific gene. This type of analysis is especially useful when attempting to dissect a metabolic pathway or signaling pathway, because many genes are often conserved between close relatives and it is possible to detect differences in the upregulation of gene families or conserved genes acting at key nodes involved in specific pathways. Type II error (not detecting an expressed gene when it is truly expressed) will result due to failure to align in divergent sequences or due to the lack of a homologous gene in the related genome.

2.5.2. Case Studies for the Alignment First Approach

Toth et al. (2007) published one of the first studies to recognize the potential of the high-throughput alignment first approach for non-model organisms. They examined social organization in *Polistes metricu* (wasp) by aligning short 454 sequencing reads (the mean read length for the early applications was 120nt) from brain tissue to the *Apis mellifera* (honey bee) genome. Several genes thought to be involved in social organization in wasps were identified. Later, Collins et al. (2008) generated Illumina data from the allopolyploid plant *Pachycladon enysii* and aligned these sequences to the genome of *Arabidopsis thaliana* (Collins et al. 2008). Together, with de novo assembly they were able to identify genes belonging to the ancestral genomes of *P. enysii*, but also recognized the potential of the method applications for other non-model organisms.

The alignment first approach to a time course experiment involving *Phytophthora*-challenged red raspberry (*Rubus idaeus* L.) is presented here, highlighting many of the issues encountered in the analysis of non-model transcriptomes and non-model experimental systems. Red raspberry is highly heterozygous and currently has few publicly available genomic resources. The genome of the closely related *Fragaria vesca* L. (woodland or alpine strawberry) (Shulaev et al. 2010) was used as the reference genome because both strawberry and raspberry are members of the Rosaceae family, have similar genome size (240 Mpb compared to 280 Mbp for *R. idaeus*) (Arumuganathan and Earle, 1991; Shulaev et al., 2010) and the same chromosome number (x=7; 2n=2x=14).

Four pools (resistant control, resistant with *P. rubi*. inoculation, susceptible control, susceptible with *P. rubi*. inoculation) of mRNA were generated by combining the mRNA from four replicates at three time points (2 days, 5 days, and 20 days). The pool of samples from the exposed susceptible genotype had only two time points because by day 20 the plants were dead. Sequencing libraries were constructed from the four pools and a data set of over 70 million quality-filtered 86nt reads was generated on the Illumina GA IIx. This bulking scheme was designed to maximize discovery of genes expressed only at a particular phase of disease progression while minimizing cost.

Sequencing reads from the resulting four data sets were aligned independently to the *F. vesca* genome and gene expression and differential gene expression levels were calculated using a normalizing statistic called FPKM (Fragments mapped Per Kilobase of Exon per Million reads mapped). This provides a measure of expression level that accounts for variation in gene length. The union of the four data sets included 16,956 unique and putatively expressed genes. No technical replicates (i.e. subsamples sequenced separately) were used, but the expression levels

for genes that were quantified in both the treatment and control pools show similar expression for most genes (Figure 2). This demonstrated that sampling error and lane effects (variation between sequencing runs) did not significantly influence the expression analysis. The relationship between the expression for a given gene in the resistant control versus the resistant treatment was stronger (Pearson's correlation = 0.93) than that shown between the susceptible control and susceptible treatment (Pearson's correlation = 0.70) (Figure 2). The reduced relationship for the susceptible cultivar is likely due to the absence of the final time point in the susceptible treatment pool (the plants were dead at 20 days), but may also be due to more extreme changes in gene expression induced in highly diseased tissue.

In order to further increase the stringency of the analysis only genes whose 95% confidence interval excluded zero were considered to eliminate the analysis of genes with little or no expression. The interpretation of results focused on differences in the resistant genotype because the resistance response is thought to be induced thus requiring changes in gene expression. Furthermore, the ability to detect differentially expressed genes induced by the pathogen is more powerful between the resistant treatment and control because differences in the susceptible genotype could possibly be developmental. Differences in the susceptible genotype were also considered for key genes in defense response pathways, particularly when the response was the opposite of that in the resistant genotype.

More than 700 genes were significantly differentially expressed between the resistant control plants and the resistant plants exposed to the pathogen. This was characterized by a log fold expression change greater than 0.5 or less than -0.5 and a FPKM greater than three after Benjamini Hochberg False Discovery Rate correction for multiple testing as performed in Cufflinks. To generate hypotheses about the mechanism of resistance to root rot caused by *P.*

rubi in red raspberry, only genes that were previously implicated in disease resistance in other plants or for which a contrasting response between resistant and susceptible plants was observed were considered.

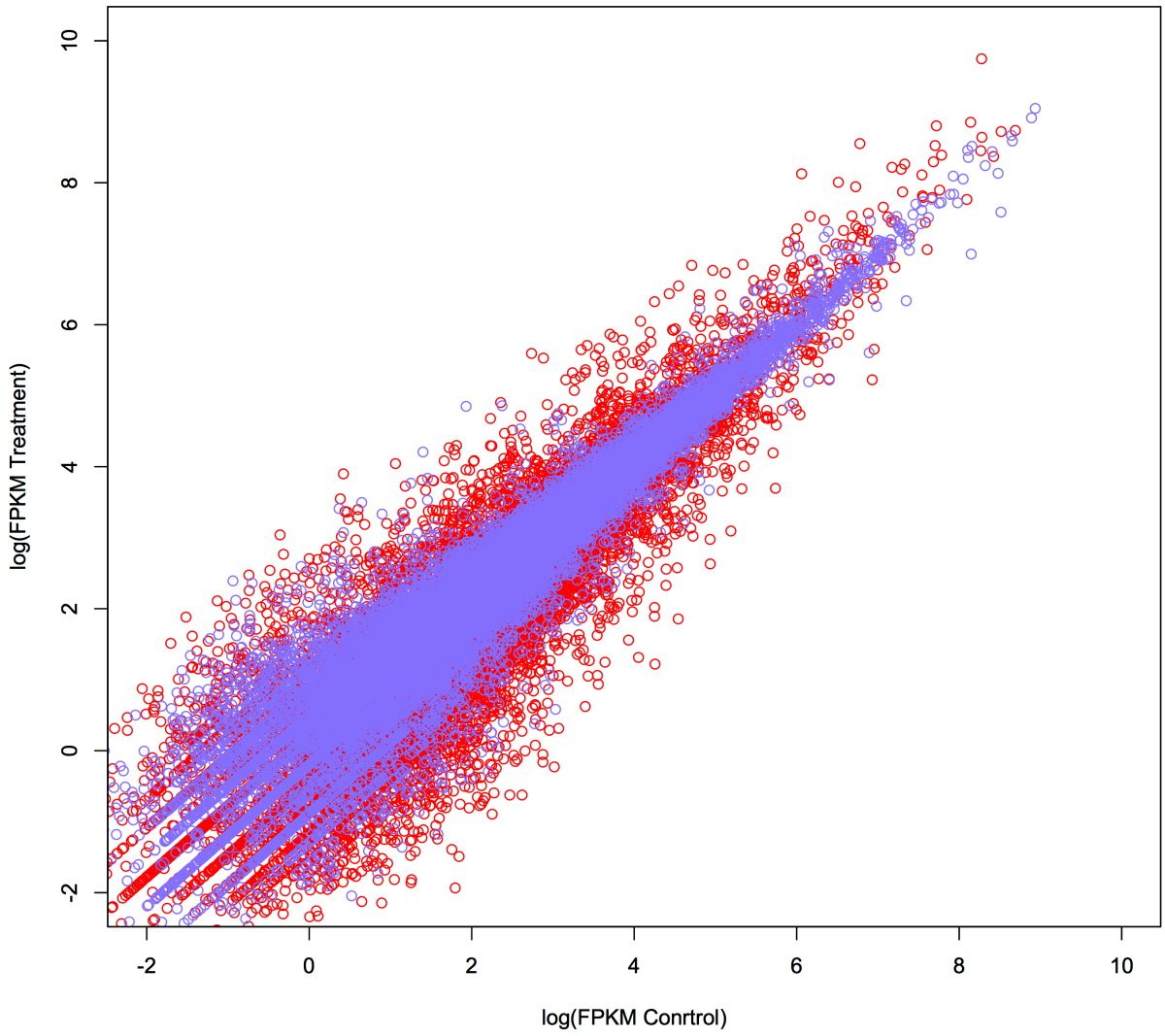

Figure 2.2 A comparison between treatments and controls by genotype for gene expression as predicted by an alignment-first method. In red, the susceptible control is compared to the susceptible treatment and shows increased divergence from correlation (Pearson's correlation = 0.73) when compared to the resistant genotype (blue, Pearson's correlation = 0.93). Overall, the similarity in expression levels between treatment and control demonstrates reproducibility of the alignment-first method.

In the resistant treatment the experiment detected increased expression of Pathogenesis Related proteins (PR proteins) that have long been implicated in disease resistance, WRKY transcription factors that may be involved in various defense response pathways (Panstruga et al. 2009) and ATP citrate synthase which may be rate limiting in a metabolism level defense response (Bolton 2009). These results are consistent with current models of plant-pathogen interaction and are useful for generating specific testable hypotheses for future studies.

2.6. Assemble-then-align methods

In the context of non-model organisms, the assemble first methods are essentially de novo assemblies, where the alignment to a close relative can help to validate the results and help identify exon-exon boundaries, which are essential for identification and validation of alternative splicing. Furthermore, knowing exon-exon boundaries in transcripts can aid the development of primers for follow up studies with quantitative reverse transcriptase polymerase chain reaction (RT PCR) to look at expression of specific genes in different tissues and/or time points or environments.

When analyzing data from non-model organisms it must be noted that they may not conform to the assumptions of assemblers developed for the genomes of humans or model organisms. Many non-model plants are heterozygous, polyploid or both, and assembly of the transcriptome from a highly heterozygous or polyploid plant could result in either the formation of chimeric transcripts (representing two haplotypes for instance) if the assembler lacked sensitivity or an extremely large set of transcripts that is more difficult to analyze (this occurs when the algorithm is highly sensitive and separately assembles divergent alleles and alternatively spliced transcripts).

Numerous transcriptomes from non-model plants have recently been analyzed in an assemble first manner utilizing two basic approaches dependent upon on the sequencing platform used and the type of algorithm used to assemble the data: 1) variations of Overlap Layout Consensus (OLC) assemblers; and 2) the Eulerian path assemblers. Examples of variations on OLC assemblers include Roche's GS de novo Assembler (Newbler) (Roche/454 Life Sciences, Branford, CT), MIRA (Chevreux et al. 2004), and CAP3 (Huang and Madan, 1999). The OLC assemblers produce very high quality assemblies for data sets with long reads and low depth by constructing a graphic representation of overlapping sequences and subsequently moving though the graph to assemble transcripts. However, as the number of reads increases, there is a need for increased memory efficiency and decreased computational time.

Eulerian path assemblers offer a more efficient method of assembly with the construction of a de Bruijn graph. In this case the de Bruijn graph is constructed by identifying all existing short subsequences of a defined length (k) in all reads and finding overlap of k-1 in among these to create links between all subsequences of the defined length to generate the graph. While non-intuitive, these graphs can be efficiently traversed with elegant mathematical principles (Pevzner et al. 2001). Early examples of Eulerian-based assemblers include Euler (Pevzner et al. 2001) and Velvet (Zerbino et al. 2008).

Transcriptome assemblers are often built on the backbone of a genome assembler designed for analyzing data from genomic DNA or are simply a genome assembler with altered parameters. Examples of genome assemblers adapted for the assembly of transcriptomes include miraEST (Chevreux et al. 2004), Trans-ABySS (Robertson et al. 2010), Oases (http://www.ebi.ac.uk/~zerbino/oases/) and Rnnotator (Martin et al. 2010). More recently, Trinity (Grabherr et al. 2011) was introduced, which specifically takes into account the

differences between the transcriptome and genome assembly problem. Trinity makes use of greedy assembly (finding overlapping similar sequences), followed by partitioning of overlapping components and creation of multiple graphs representing reads from individual genes. Finally it traverses these graphs to obtain linear transcript sequences. A summary of the most common de novo assembly methods is shown Table 1.

2.6.1. Case Studies for the Assemble First Approach

Some non-model species trancriptomes recently analyzed with de novo assemble first approaches include *Artemisia annua* (Artemisia) (W. Wang et al. 2009b), *Oryza longistaminata* (a wild species of rice) (Yang et al. 2010), *Fagopyrum esculentum* and *F. tataricum* (buckwheat) (Logacheva et al. 2011) and *Pteridium aquilinum* (bracken fern) (Der et al. 2011). While these studies have succeeded in producing high quality assemblies, the results do not capture gene expression levels and thus, are missing a key component of the transcriptome analyzed. This is because normalized 454 libraries were used for de novo assembly and thus the expression information is lost in the normalization of the library. It is possible, however, to quantify expression levels in addition to transcript sequence when non-normalized libraries are sequenced on the 454 platform as recently done in *Panax quinquefolius L.* (American ginseng) (Sun et al. 2010), but this requires greater sequencing depth than commonly used in 454 sequencing and thus a greater expense is associated with the approach. When long reads and low depth data sets are used to sequence normalized libraries, however, another independent approach can be used to obtain gene expression information after assembly. In these cases, the researcher can later target a specific gene with quantitative RT-PCR or can generate high depth short read data sets to map back to the assembled transcriptome (see Figure 1).

Some recent studies have also made use of high depth short-read data sets from Illumina for assemble first analysis utilizing non-normalized libraries assembled with Eulerian assemblers rather than OLC assemblers. Examples include *Eucalyptus grandis* (eucalyptus) (Mizrachi et al. 2010), *Cicer arietinum* (chickpea) (Garg et al. 2011) and *Eichhornia paniculata* and *E. paradoxa* (water hyacinth) (Ness et al. 2011). In the *Eucalyptus* transcriptome assembly Mizrachi et al. (2010) generated over 62 million paired-end sequencing reads ranging from 36nt to 60nt in length. Velvet was used to assemble these into 18,894 contigs greater than 200nt in length of which 15,713 contained predictable coding sequence. Likewise, for chickpea Garg et al. (2011) used 50,523,492 paired-end 72nt reads and 56,136,825 single-end reads of 51nt in length to generate over 53,409 contigs longer than 100nt by using Oases (which makes use of Velvet). They found 45,636 of these contigs showed similarity to other legume sequences.

An assemble first approach is demonstrated on here the same *Phytophthora*-challenged red raspberry experiment presented above. A de novo assembly of the same 70 million quality filtered 86nt reads was performed using the Trinity software. This resulted in an assembly of 78,830 transcripts greater than 300nt in length. Of these, 64,277 had a blastx hit to the NCBI (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov*/)* nr database with an e-value less than 1.0×10^{-3}, of which 20,761 were bacterial, 43,461 were eukaryotic organisms, and the remaining 55 blast hits were archea, viruses or unclassified organisms. Analyzing these data further revealed that nearly all sequences that had top hits to eukarytic organisms were to vascular plants, with the exception of over 500 sequences with top hits belonging to *Phytophthora*, which had been experimentally introduced into the system. A summary of the number of assembled transcripts aligning to plants is shown in Table 2.2.

Table 2.2. Top assembled transcript blast hits to plants from RNA-Seq data produced from *Phytophthora* challenged *Rubus idaeus* (red raspberry) root tissue at 4 time points.

Genus	*Number of Blast Hits*
Vitis	16429
Populus	8933
Ricinus	6253
Arabidopsis	1793
Glycine	1322
Malus	394
Medicago	384
Oryza	334
Prunus	223
Nicotiana	155
Solanum	152
Hordeum	138
Fragaria	134
Zea	130
Sorghum	128
Total	36902

Further, the Trinity program estimated the expression levels for transcripts in the assembly and the results were compared to the expression levels detected by the alignment first method. The mean FPKM for each gene detected in the alignment first method was plotted against the FPKM of the corresponding best match in the assembly (Figure 2.3). Gene expression estimates are similar for many genes, with a Pearson's correlation of 0.73, which is more strongly correlated than some estimates of the correlation between RNA-Seq and microarrays in yeast (Z. Wang et al. 2009c).

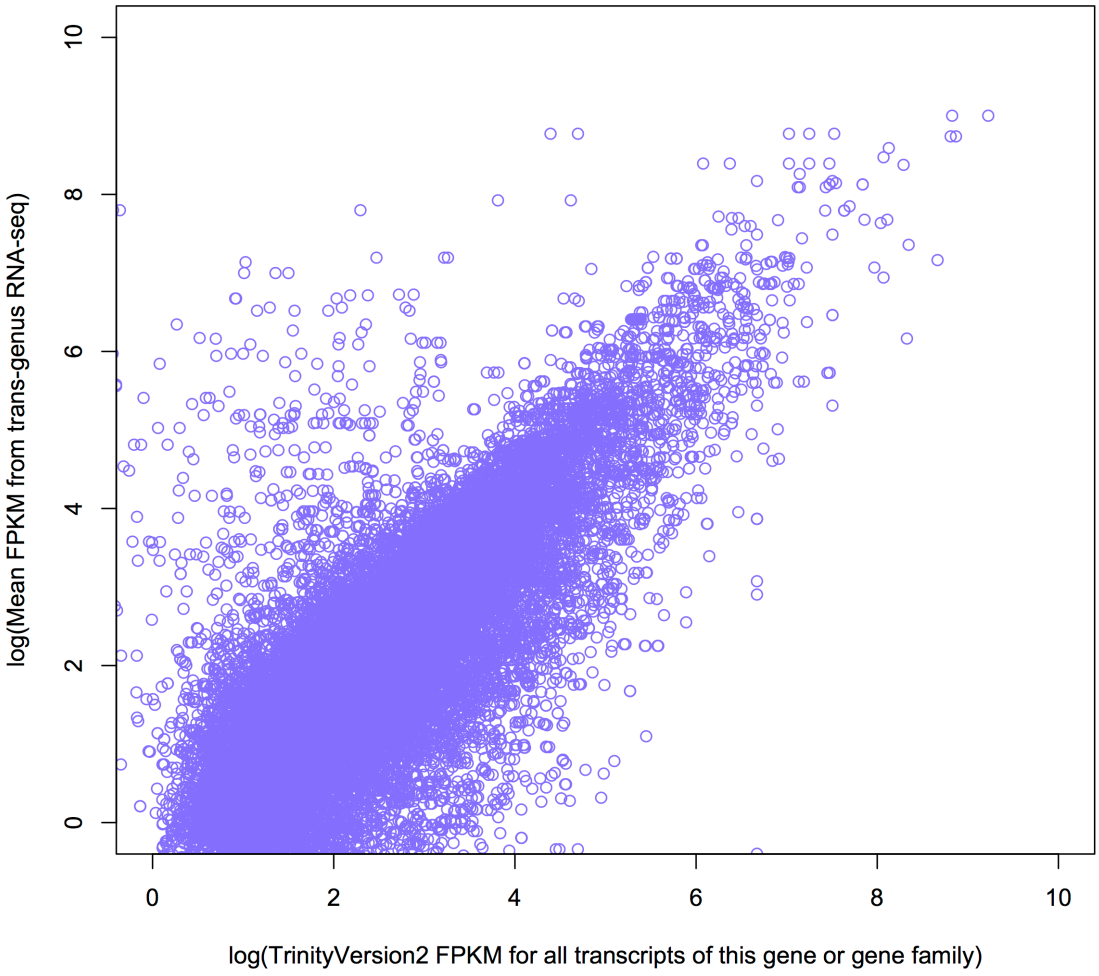

Figure 2.3. A comparison of an alignment-first method to an assemble-first method at the level of gene or gene family. In the alignment-first method (the y-axis) TopHat was used to align raspberry sequence to the ab initio gene models of strawberry and Cufflinks was used to calculate the gene expression level as reported in Fragments Mapped Per Kilobase of exon per Million Reads Mapped (FPKM). In the de novo assembly method (the x-axis) Trinity was used to assemble the sequence from all four conditions (resistant treatment, resistant control, susceptible treatment and susceptible control) and to directly estimate expression level in FPKM. This plot was generated by finding the de novo assembled transcript with the best match to the ab initio gene model from strawberry using the nucmer component of the MUMmer package. Pearson's correlation between the two methods = 0.73.

2.7. Discussion and Conclusions

Numerous transcriptomes have recently been sequenced in plants, and the number of transcriptome sequencing projects is likely to grow dramatically over the next few years in parallel with predicted decreases in sequencing costs. With this, the number of analysis possibilities for non-model plants will also grow. Understanding the technical considerations of different analysis approaches ensures researchers can choose tools that best meet their needs. Comparisons of gene expression estimates from the alignment-based method and the de novo assemblies for raspberry are encouraging, but much work needs to be done. For now, the pairing of both methods provides a powerful combination. While align-then-assemble methods provide a statistical framework for differential expression testing and hypothesis generation, the de novo approach holds more power for discovering unique sequences. De novo assembly also provides the possibility of simultaneously querying the transcripts and expression levels in multiple organisms in a system. Both methods are extremely powerful and allow researchers in non-model organisms to generate specific and testable molecular-level hypotheses about complex problems such as the dissection of metabolic pathways or plant-microbe interactions.

2.8. References

Adams, M.D. et al., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013), pp.1651–1656.

Auer, P.L. & Doerge, R.W., 2011. A Two-Stage Poisson Model for Testing RNA-Seq Data. *Statistical Applications in Genetics and Molecular Biology*, 10(1).

Bachem, C.W. et al., 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant Journal*, 9(5), pp.745–753.

Bolton, M.D., 2009. Primary Metabolism and Plant Defense—Fuel for the Fire. *Molecular Plant-Microbe Interactions*, 22(5), pp.487–497.

Brenner, S. et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 18(6), pp.630–634.

Chevreux, B. et al., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research*, 14(6), pp.1147–1159.

Collins, L. et al., 2008. An Approach to Transcriptome Analysis of Non-Model Organisms Using Short-Read Sequences. In *Genome Informatics 2008: Proceedings of the 19th International Conference*. Genome Informatics 2008: Proceedings of the 19th International Conference. pp. 3–14.

Davey, J.W. et al., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), pp.499–510.

Delcher, A.L. et al., 2002. Fast algorithms for large-scale genome alignment and comparison.

Nucleic Acids Research, 30(11), pp.2478–2483.

Der, J.P. et al., 2011. De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. *BMC Genomics*, 12(1), p.99.

Garg, R. et al., 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 18(1), pp.53–63.

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652.

Guttman, M. et al., 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnology*, 28(5), pp.503–510.

Haas, B.J. & Zody, M.C., 2010. Advancing RNA-Seq analysis. *Nature Biotechnology*, 28(5), pp.421–423.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), p.R25.

Lashkari, D.A. et al., 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 94(24), pp.13057–13062.

Liang, P. & Pardee, A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257(5072), pp.967–971.

Logacheva, M.D. et al., 2011. De novo sequencing and characterization of floral transcriptome in two species of buckwheat (Fagopyrum). *BMC Genomics*, 12, p.30.

Malhotra, K. et al., 1998. Interaction and effect of annealing temperature on primers used in

differential display RT-PCR. *Nucleic Acids Research*, 26(3), pp.854–856.

Martin, J. et al., 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics*, 11, p.663.

Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11(1), pp.31–46.

Mizrachi, E. et al., 2010. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics*, 11(1), p.681.

Ness, R.W., Siol, M. & Barrett, S.C.H., 2011. De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics*, 12(1), p.298.

Panstruga, R., Parker, J. & Schulze-Lefert, P., 2009. Snapshot: Plant Immune Response Pathways. *Cell*, (136), p.978.e2.

Pattison, J.A., Wilcox, W. & Weber, C.A., 2004. Assessing the resistance of red raspberry (Rubus idaeus L.) genotypes to Phytophthora fragariae var. rubi in hydroponic culture. *Hortscience*, 39(7), pp.1553–1556.

Pevzner, P.A., Tang, H. & Waterman, M., 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 98(17), p.9748.

Robertson, G. et al., 2010. De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), pp.909–912.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.

Schena, M. et al., 1995. Quantitative monitoring of gene expression patterns with a

complementary DNA microarray. *Science*, 270(5235), pp.467–470.

Schneeberger, K. & Weigel, D., 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends in Plant Science*, 16(5), pp.282–288.

Shulaev, V. et al., 2010. The genome of woodland strawberry (Fragaria vesca). *Nature Genetics*, 43(2), pp.109–116.

Sun, C. et al., 2010. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics*, 11, p.262.

Trapnell, C. et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), pp.511–515.

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–1111.

Velculescu, V., Zhang, L. & Vogelstein, B., 1995. Serial analysis of gene expression. *Science*, 270(5235), pp.484–487.

Vuylsteke, M., Peleman, J. & van Eijk, M.J.T., 2007. AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. *Nature Protocols*, 2(6), pp.1399–1413.

Wang, L. et al., 2009a. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26(1), pp.136–138.

Wang, W. et al., 2009b. Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics*, 10, p.465.

Wang, Z., Gerstein, M. & Snyder, M., 2009c. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), pp.57–63.

Yang, H., Hu, L. & Hurek, T., 2010. Global characterization of the root transcriptome of a wild species of rice, Oryza longistaminata, by deep sequencing. *BMC Genomics*. 11, pp.705.

Zerbino, D.R. et al., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5), p.821.

2.9. SUPPLEMENTAL METHODS

Forty clones of the resistant genotype cv. 'Latham', along with 40 clones of the susceptible genotype cv. 'Titan' were randomly assigned to eight independent hydroponic systems originally developed for screening raspberry to resistance to *Phytophthora fragariae* var. *rubi* (Pattison et al. 2004) now known as *P. rubi*. Four of these systems were challenged with single zoospore isolates of the pathogen. Methods for challenging the systems were as described previously (Pattison et al. 2004). The remaining four hydroponic systems remained un-inoculated to serve as controls. Three root samples of approximately one gram were collected into 23.88-micron thick aluminum foil packets. These were immediately flash frozen in liquid nitrogen and transferred to storage at -80 °C until RNA extraction was performed. Both resistant and susceptible cultivars were sampled 48 hours post inoculation and 5 days post inoculation. Tissue was also sampled after 20 days in all treatments and controls in both resistant and susceptible plants with the exception of the 'Titan' treatment, as root tissue was too highly degraded in the plants by this time to extract a high quality RNA sample.

Total RNA was extracted from approximately 25 mg root tissue from four randomly selected resistant (one from each replicate) and four randomly selected susceptible plants (one from each replicate) at each time point for both treatments and controls using Qiagen RNeasy Plant Mini Kits (Qiagen Inc., Valencia, CA) according to the manufacturers recommendations except with the addition of an extra buffer RPE wash and incubation for 30 seconds prior to centrifugation. Samples were further pooled within each available time point by variety and treatment creating a total of eleven samples (there is no sample for the final timepoint from

'Titan'). Each of these samples were then purified again with Qiagen RNeasy Plant Mini Kits according to the manufacturers recommendation for purification of RNA with the addition of an extra wash with buffer RPE and an incubation of 30 seconds before centrifugation. All eleven samples were run on a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California) and RNA Integrity Numbers (RINs) were calculated. Only samples with RIN scores higher than eight were used in library preparation.

Total mRNA from the 11 samples was combined within treatments and controls and by genotype across time points to create a total of 4 samples representing the resistant control, resistant treatment, susceptible control, and susceptible treatment. Sequencing libraries were prepared independently for each of these for samples according to the Illumina mRNA-Seq Protocol (Illumina, San Diego, CA), Part #1004898 (Rev. A September 2008). Cluster generation and single-end 86nt sequencing was performed according to the manufacturers recommendations on the Illumina Genome Analyzer IIx (Illumina, San Diego, CA) in independent lanes. Raw sequencing images from the Illumina Genome Analyzer IIx were processed using default settings on Illumina's 1.6.0 pipeline (RTA 1.6.32.0) with standard quality cross-calibration to a bacteriophage PhiX control lane.

For the alignment first method the strawberry genome assembly version 8 was used as the reference genome. Alignments were performed using Bowtie Version 0.12.3 (Langmead et al. 2009) and Tophat version 1.0.13 (Trapnell et al. 2009). These alignments were restricted to strawberry genome ab initio gene models (genemark_predictions_101209.gff3). Gene level expression and differential expression analysis were performed using Cufflinks version 0.8.0 (Trapnell et al. 2010) with default settings. De novo assembly of the transcriptome was performed using Trinity (Grabherr et al. 2011) using data from all four lanes of sequencing and

using default settings. The nucmer component of the MUMmer package (Delcher et al. 2002) was used to align all the Trinity transcripts to the strawberry ab initio gene models with default settings. The resulting delta file was filtered with the -q & -r options to identify the transcripts that were most likely to be the homologous match in the strawberry genome.

Chapter 3

DISSECTING RESISTANCE TO PHYTOPHTHORA ROOT ROT IN RUBUS IDAEUS (ROSACEAE) WITH RNASEQ IN THE ABSENCE OF A REFERENCE GENOME

3.1. Abstract

Resistance to Phytophthora root rot (PRR) (caused by Phytophthora rubi (Wilcox and Duncan) Man in 't Veld, comb. nov. in the red raspberry (Rubus idaeus L.) 'Latham' was investigated. Four sequencing libraries were prepared with mRNA from 'Latham' (resistant to PRR) and 'Titan' (susceptible to PRR) in controls and after inoculation. Over 70,600,000 86nt reads were sequenced. RNA-seq was performed using the draft genome of the woodland strawberry (Fragaria vesca L.) for alignment. This detected 16,956 unique and putatively expressed genes across the four data sets. The intersection of the sets included 12,746 genes. Differential expression analysis revealed changes in defense pathways between treatments and controls. Overlap in activation of defense responses between 'Latham' and 'Titan' was apparent, but many changes were unique to 'Latham'. Differences in PAMP triggered immunity (PTI) signal transduction included increased expression of MAPKK2 in 'Latham'. A homolog to PEN3 was up regulated in the 'Latham' treatment along with an elicitor inducible cytochrome P450. NPR1 was up regulated in the 'Latham' treatment and down regulated in the 'Titan' treatment. A WRKY transcription factor similar to WRKY51 had increased expression in the 'Latham' treatment and not significantly expressed in other conditions. Nineteen members of the PR-10 family were detected and 18 were significantly up regulated in 'Latham', while in 'Titan', only

10 were up regulated and 3 were down regulated. Other defense responses include decreased expression of negative regulators of cell death (LSD1 and gf14 kappa) in 'Latham' and increased expression in 'Titan'. The 'Latham' treatment also had increased expression of ATP citrate synthase, suggesting increased flux through the citric acid cycle (TCA), which may contribute to root vigor. Together, these results provide evidence that programmed cell death, activation of specific defense responses, and increased primary metabolism in 'Latham' are involved in increased resistance.

3.2. Introduction

In red raspberry (*Rubus idaeus L.*), there are a number of species reported to cause Phytophthora Root Rot (PRR), but *Phytophthora rubi* (Wilcox and Duncan) Man in 't Veld, comb. nov. is known to be the principal agent of economically damaging PRR throughout the world (Wilcox, 1989; Wilcox et al., 1993). Growers can implement a number of cultural methods to control disease, such as using raised beds, choosing a site with good drainage, and seasonal soil application of metataxyl but these only have minimal influence compared to genetic factors (Wilcox et al., 1999). For this reason genetic resistance sources have been examined. Initially, a number of research programs screened germplasm for varieties with resistance to PRR and several were identified. A resistance source of particular interest was identified in the cultivar 'Latham', which was developed in 1908 at the University of Minnesota. 'Latham' continues to be a source of resistance for breeders because it offers good resistance and has endured without any reports of resistance-breaking isolates.

The durable nature of the resistance in 'Latham' is likely due to the fact that the underlying genetic mechanisms of resistance are quantitative. However, other reasons for the

durability of resistance may include the homothallic nature of P. rubi (Wilcox et al., 1993) and its almost exclusively subterranean lifecycle (Wilcox, 1992). Thus, unlike aerially dispersed and heterothallic species such as P. infestans, P rubi may lack diversity and have limited ability to spread. Diversity within P. rubi and the biology of pathogenesis remains a relatively unexplored area of research, but it is interesting because P. rubi has an extremely limited host range (so far it is only reported to affect Rubus species) while some Phytophthora caused root rots, such as such as P. cinnamomi, affect thousands of species (Hardham, 2005). While durability of resistance can suggest either an extensively quantitative trait or limited diversity in the pathogen, some loci may be of greater importance in conditioning resistance and therefore could be considered targets for marker-assisted selection.

Pattison (2004) developed a hydroponic system to screen raspberry plants for resistance to PRR (Pattison et al., 2004). This hydroponic system was also used to study the inheritance of resistance of PRR and for phenotyping in mapping studies. Segregation ratios were examined in several populations derived from NY00-34 (‘Titan’ x ’Latham’) and a dominant two-gene model was proposed that explains major components of the resistance in ‘Latham’ (Pattison et al., 2007). Under this model, ‘Latham’ is thought to be homozygous for a dominant resistance gene at one locus and heterozygous at another locus. Other mapping studies in ‘Latham’ also found two major loci involved in resistance to P. rubi (Graham et al., 2011), but extensive comparison of these independent findings has not yet been performed.

Pattison et al. (2004; 2007) observed that roots of ‘Latham’ appear to be initially susceptible to PRR. However, after a period of susceptibility (where some roots die), new roots emerge and remain healthier despite the continued presence of the pathogen (Pattison et al., 2004; Pattison, 2004). The apparent change in the level of resistance suggests altered gene

expression following exposure to P. rubi. These changes may limit infection and spread in new growth, but may also be due to an increased rate of root growth. Simply limiting disease progress in this manner could be effective in the field, where spread of P. rubi and continued disease cycles depend on flooded conditions.

Others have also made observations suggesting that root re-growth is a significant factor in resistance. Laun et al. (1997) suggested that following an initial response to the pathogen an "intense regeneration of unaffected roots was visible" in later growth of the moderately resistant 'Winklers Samling'. This suggests that 'Latham' and 'Winklers Samling' may share some common mechanisms of resistance and that both may be linked to changes in gene expression that allow for more vigorous root growth with limited spread of the pathogen. Additionally, Graham et al. (2011) recently reported one of two QTL for resistance overlaps with root vigor QTL mapped in soils without PRR further suggesting root vigor may be involved in resistance to PRR in 'Latham'.

Graham et al. (2011) also sequenced 2 BACs and annotated numerous genes that may be involved in either defense or metabolic activity that could potentially contribute to vigor. However, each of these BAC sequences spans only a distance of ~145kb and the causal gene(s) underlying the QTL may not have been covered. Still, the BAC sequences are sufficiently large to cover much of the gene space near QTL identified by Graham et al. (2011), and it is likely that genes within these BACs or in nearby gene space may play a significant role in the ability of 'Latham' to endure attack by P.rubi. Furthermore, observations made by Laun et al. (1997) and Pattison et al. (2004; 2007) suggest that there may be a change in gene expression that leads to increased ability to endure disease.

Activation of defense responses can arise from linked defense pathways including effector triggered immunity (ETI) and pathogen associated molecular pattern (PAMP, (Medzhitov and Janeway, 1997)) triggered immunity (PTI) (Jones and Dangl, 2006; Boller and He, 2009). ETI occurs after the plant recognizes one or more effectors from the pathogen. Defense responses associated with ETI typically involve the hypersensitive response (HR) and salicylic acid (SA) mediated resistance which may involve WRKY or TGA transcription factors (Eulgem, 2005). ETI is thought to be effective at mediating resistance to biotrophic pathogens via HR, but not necrotrophic pathogens.

PTI is associated with recognition of PAMPs which are highly conserved portions of the pathogen and may result in induction of jasmonic acid (JA) mediated defense. PTI is an effective defense against necrotrophic pathogens via activation of ERF (Berrocal-Lobo et al., 2002) or MYC transcription factors (Lorenzo, 2004). PTI may also trigger HR (Kamoun, 2006) or defense responses in a JA independent manner that is likely to involve WRKY transcription factors (Ingle et al., 2006). Activation of the JA pathway suppresses the SA pathway and vice versa (Glazebrook, 2005).

This model is robust for pathogens that are clearly divided into biotrophic and necrotrophic categories, but for hemi-biotrophs, such as many Phytophthora species, more extensive modulation of defense responses must occur because Phytophthora species are known to occasionally escape the HR (Kamoun et al., 1999; Vleeshouwers et al., 2000; Chen and Halterman, 2011). Furthermore, under environmental conditions where plants are exposed to multiple pathogens (biotrophic, necrotrophic, and hemibiotrophic) defense responses must be somewhat generalized. Generalized responses include induction of pathogenesis related (PR) proteins, callose depostion, strengthening of the cell wall, secretion of toxic secondary

metabolites to the apoplast via the PEN3 ABC transporter (Bednarek et al., 2009) and localized cytoplasmic release of antimicrobial cargo such as PR-1 (Kalde et al., 2007)via the Ternary SNARE complex.

It remains unclear the extent to which ETI and PTI constitute distinct mechanisms (Jones and Dangl, 2006). However, because these pathways have unique initiation points they are extremely useful in the dissection of genome-wide expression data and may point to key differences in the initiation of defense responses between resistant and susceptible genotypes. This study leverages the use of the recently completed woodland strawberry (Fragaria vesca L.) genome to perform genome-wide expression profiling and identifies key genes in known plant defense pathways that are likely to contribute to resistance in the cultivar 'Latham.'

3.3. Methods

3.3.1. Plant Material

Tissue culture plants of 'Latham' and 'Titan' were obtained from Nourse Farms (South Deerfield, MA) and were maintained in a Murashige and Skoog basal medium (Murashige and Skoog, 1962) with minor modifications. Plants were transferred to rock wool cubes and allowed to acclimatize and grow for five weeks in hydroponic systems developed for screening red raspberry for resistance to PRR (Pattison et al., 2004). Growth chambers were maintained at 20 ºC with 16 hours of daylight. A total of 40 clones of the resistant cultivar 'Latham' and 40 clones of the susceptible cultivar 'Titan' were randomly assigned to eight separate hydroponic systems, with 5 clones of each variety in each of the 8 bins.

3.3.2. Phytophthora rubi isolation

Plants with symptoms of PRR were obtained from the northeast section of block DA005RASCAW of the Darrow Farm (New York State Agricultural Experiment Station (NYSAES), Cornell University, Geneva NY). Roots were rinsed free of soil and were stored in sterile distilled water during processing. Roots with symptoms were selected and sections of roots at the margin of disease roots and healthy roots were cut and briefly dipped in 70% ethanol before placement into a Phytophthora selective cornmeal agar prepared with pimaricin, ampicillin, rifampicin, and hymexazol (P5ARPH) in concentrations given by Jeffers and Martin (Jeffers and Martin, 1986). Plates were incubated at 20 ºC in the dark and were checked daily for emerging aseptate hyphae with a characteristic right angle branching pattern. Newly emerging hyphae were excised from the P5ARPH and were placed onto CV8A agar plates (Wilcox et al., 1993)with rifampicin (5mg/L) and sodium ampicillin (125mg/L). Phytophthora isolates were subcultured every two weeks. Verification of isolate identity was obtained by extracting DNA from hyphae with Qiagen DNeasy kits (Qiagen Inc., Valencia, CA) and amplification of ITS sequences with universal ITS primers 1 and 4 (White et al., 1990). Amplification was followed by sequencing of the PCR product with ITS1 or ITS4 in separate reactions on an Applied Biosystems Automated 3730 DNA Analyzer with Big Dye Terminator chemistry and AmpliTaq-FS DNA Polymerase (Applied Biosystems, Foster City, CA). The resulting sequences were BLASTED against the NCBI nr/nt database. Isolates with top BLAST hits to P. rubi type specimens were retained.

Non-sterile soil extract (NSE, (Jeffers and Aldwinckle, 1987)) was prepared to induce formation of sporangia [after numerous failed attempts at sporangium induction with sterile methods]. Sporangia formation was apparent after two days and the plates were kept at room temperature under ambient light. Any remaining NSE on the plate was poured off and the plate

was flooded with sterile water to induce zoospore release. Zoospores were serial diluted without quantification and plated on P5ARPH. The plates checked daily under a light transmission microscope for growth. After growth of hyphae was observed emerging from zoospores, the zoospore and associated hyphae were transferred to V8RA plates and grown at 20 º C in the dark.

After additional hyphae growth, samples were collected and DNA was extracted for ITS sequence amplification and sequencing as described above. The single zoospore sequences were again BLASTED to the NCBI nr/nt database and all isolates had sequences with top hits to the P. rubi strain ATCC 64968. Single zoospore isolates were tested for virulence in a smaller trial containing 5 'Latham' and 5 'Titan' plants and were able to reproduce the original disease symptoms observed in 'Titan'.

3.3.3. Experimental design, sampling, and RNA extraction

Four replicate hydroponic systems were challenged with single zoospore derived P. rubi isolates and four systems remained un-inoculated to serve as controls. For RNA extraction three root tissue samples of approximately one gram each were collected into 23.876 micron thick aluminum foil packets and were immediately flash frozen in liquid nitrogen and stored at -80 ºC. Both resistant and susceptible cultivars were sampled at 2 days post inoculation and at 5 days post inoculation. Tissue was also sampled after 20 days in all treatments and controls in both resistant and susceptible plants with the exception of the 'Titan' treatment because the root tissue was too highly degraded to by this time to extract a high quality RNA sample.

Total RNA was extracted from approximately 25 mg of root tissue from four randomly selected resistant (one from each replicate) and susceptible plants (one from each replicate) at each time point for both treatments and controls using Qiagen RNeasy Plant Mini Kits (Qiagen

Inc., Valencia, CA) according to the manufacturers recommendation except with the addition of an extra buffer RPE wash and incubation for 30 seconds prior to centrifugation. Samples were further pooled within each available time point by variety and treatment creating a total of eleven samples (there no sample for the final timepoint from 'Titan'). Each of these samples were then purified again Qiagen RNeasy Plant Mini Kits according to the manufacturers recommendation for purification of RNA with the addition of an extra wash with buffer RPE and an incubation of 30 seconds before centrifugation All eleven samples were then run on a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California) and RNA Integrity Numbers (RINs) were calculated. All RINs were above eight (see table M-1) indicating high quality samples.

3.3.4. Library Preparation and RNAseq

Library preparation for sequencing was performed according to the Illumina mRNAseq Protocol (Illumina, San Diego, CA), Part #1004898 (Rev. A September 2008). Single end 86nt sequencing was performed on the Illumina Genome Analyzer IIx (Illumina, San Diego, CA) at the Cornell University Life Sciences Core Laboratories Center according to the manufacturer's recommendations. Raw sequencing images from the Illumina Genome Analyzer IIx were processed using default settings on Illumina's 1.6.0 pipeline (RTA 1.6.32.0) with standard quality cross-calibration to a bacteriophage PhiX control lane.

The strawberry genome assembly version 8 was used as the reference genome for the so called Tuxedo RNAseq pipeline, which involves alignments performed using Bowtie Version 0.12.3 (Langmead et al., 2009) and Tophat version 1.0.13 (Trapnell et al., 2009) followed by Cufflinks version 0.8.0 (Trapnell et al., 2010). These alignments utilized strawberry genome ab initio gene models (genemark_predictions_101209.gff3). Each of the four data sets was aligned independently to the strawberry genome and gene expression was calculated using a normalizing

statistic called FPKM (fragments mapped per kilobase of exon per million reads mapped). Gene level expression and differential expression analysis were performed using Cufflinks version 0.8.0 (Trapnell et al., 2010) with default settings.

3.4. Results

Sequencing on the Illumina GAIIx resulted in 15,213,629 86nt reads for the 'Latham' control, 18,363,149 86nt reads for the 'Titan' control, 18,923,742 86nt reads for the 'Latham' treatment, and 18,145,135 86nt reads for the 'Titan' treatment. This sequencing depth is approximately equal to 5-fold genome coverage for each condition. Analysis of the raw sequencing data was conducted to reveal any influence of the sequencing run (lane effects) on further analysis. Figure 3.1 – Figure 2.4 show the descriptive statistics for per base Phred scores across sequencing reads for each data sets. Phred scores, were originally defined in the program "Phred" as $q = -10 \times \log10\ p$, where p is the probability estimated error probability for a given base (Ewing and Green, 1998). Phred scores are as high as 50 (an error rate of 1/100,000) are common in Sanger sequencing. Phred score ranges in Illumina data currently have a maximum quality of approximately 40, but for the GAIIx a Phred score of 35 is often the maximum observed Phred score.

The mean quality scores in final bases of the 'Latham' sequence data were lower than the mean quality scores in the 'Titan' sequence data. Furthermore, greater dispersion is shown by an increased interquartile range (IQR) in 'Latham' compared to 'Titan'. Differences in sequence quality and by genotype may be due to slight variations in chemistry from library preparation or variation in the physical properties between flow cell lanes. Differences by genotype could also be due to variation in phenolic compounds, sugars, or other plant metabolites, but this is unlikely

because the sample is diluted many fold during library preparation. The mean quality scores are within the acceptable range (above a Phred score of 20) across most of the read (Figure 3.1 – Figure 3.4), but alignment rates differed between the 'Latham' and 'Titan' (Table 3.1). This is likely due to poor quality scores in the final bases of these reads which increased the number of mismatches and thus lead to a higher failure to align. Normalization of with FPKM (fragments mapped per kilobase of exon per million reads mapped) was done to account for differences in mapping caused by differing quality scores between data sets.

Table 3.1 Illumina sequencing reads obtained and percent alignment by sequence origin/lane with default Bowtie/TopHat settings and untrimmed 86nt reads.

Sequence Origin	Total Reads Obtained	Number aligned to the F. vesca genome	Percent Alignment to the F. vesca genome
Latham Control	15213629	3267724	21%
Titan Control	18363149	6461806	35%
Latham Treatment	18923742	3455446	18%
Titan Treatment	18145135	5834537	32%

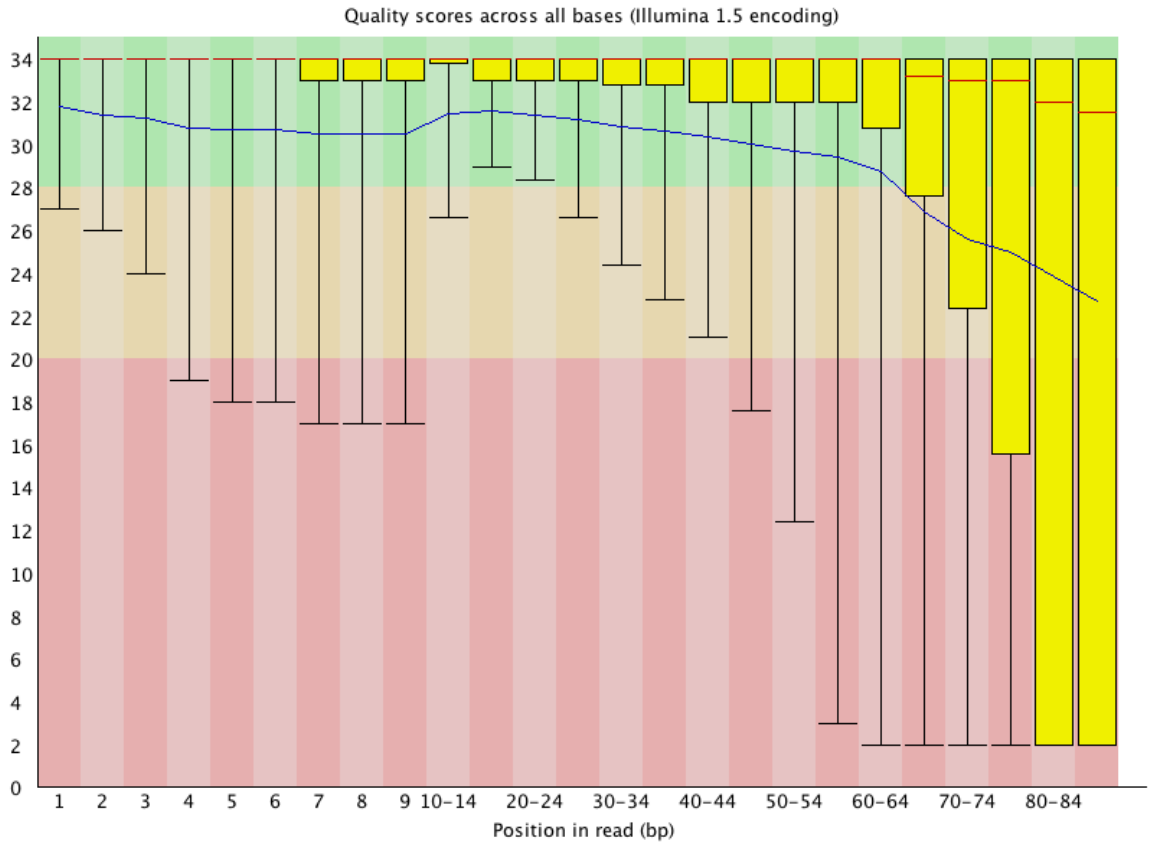

Figure 3.1 Per base sequence quality for the 'Latham' control. The y-axis is the Phred score and the x-axis is the base number in the sequencing read. The red line shows the median Phred score, the yellow boxes show the interquartile range, the blue line is the mean quality score and the whiskers show the 10th and 90th percentiles.

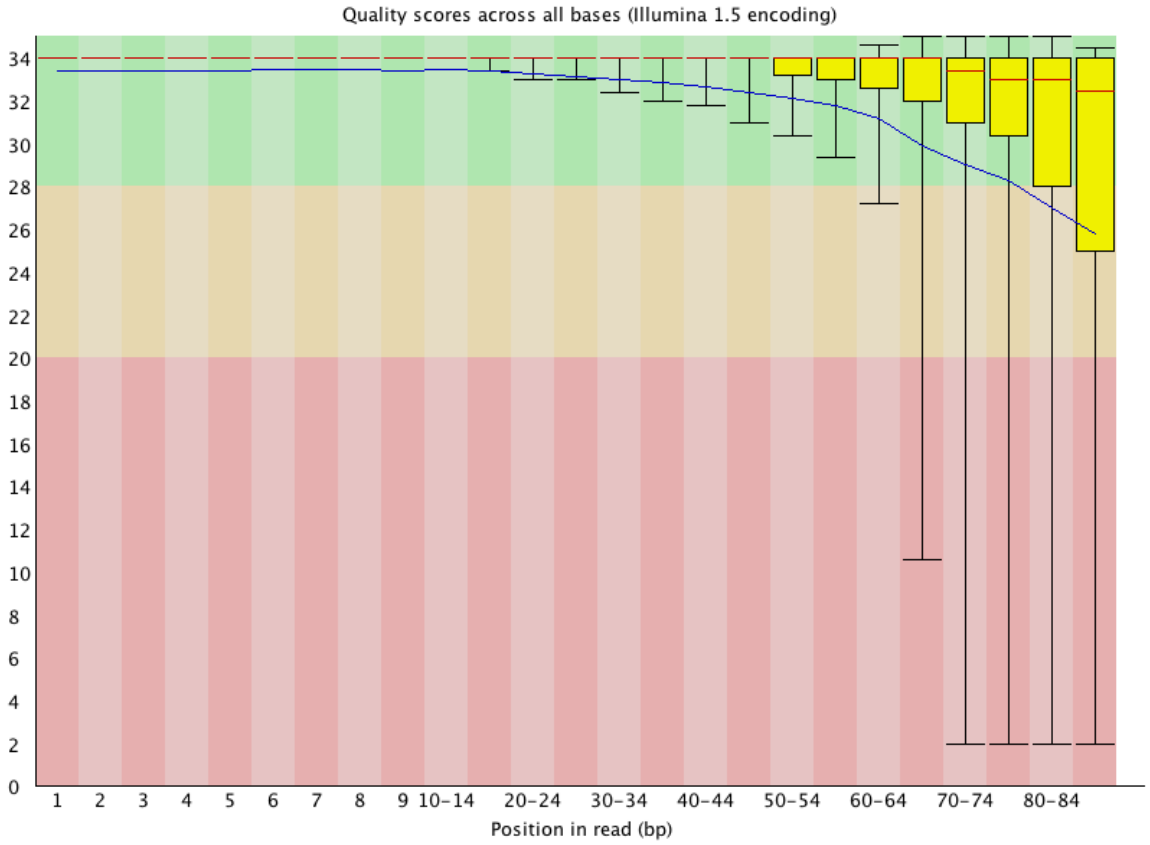

Figure 3.2 Per base sequence quality for the 'Titan' control. The y-axis is the Phred score and the x-axis is the base number in the sequencing read. The red line shows the median Phred score, the yellow boxes show the interquartile range, the blue line is the mean quality score and the whiskers show the 10th and 90th percentiles.

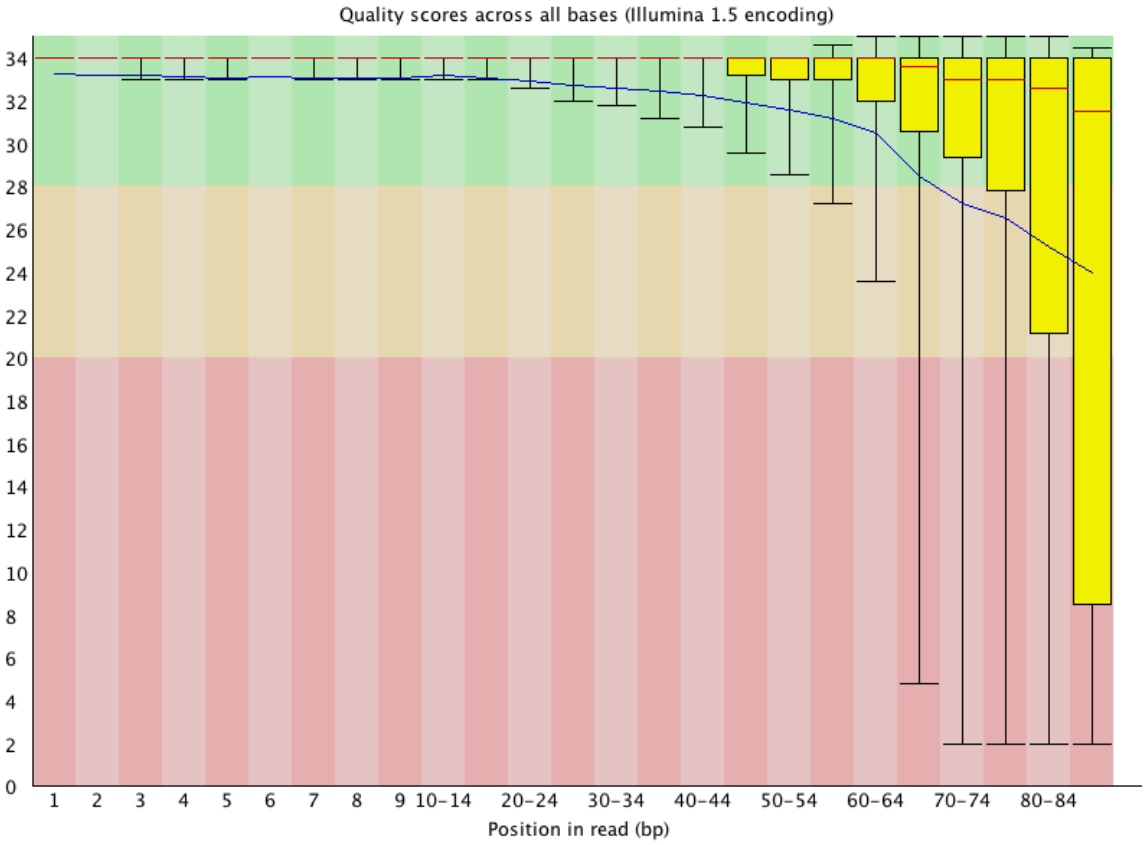

Figure 3.3 Per base sequence quality for the 'Latham' treatment. The y-axis is the Phred score and the x-axis is the base number in the sequencing read. The red line shows the median Phred score, the yellow boxes show the interquartile range, the blue line is the mean quality score and the whiskers show the 10th and 90th percentiles.

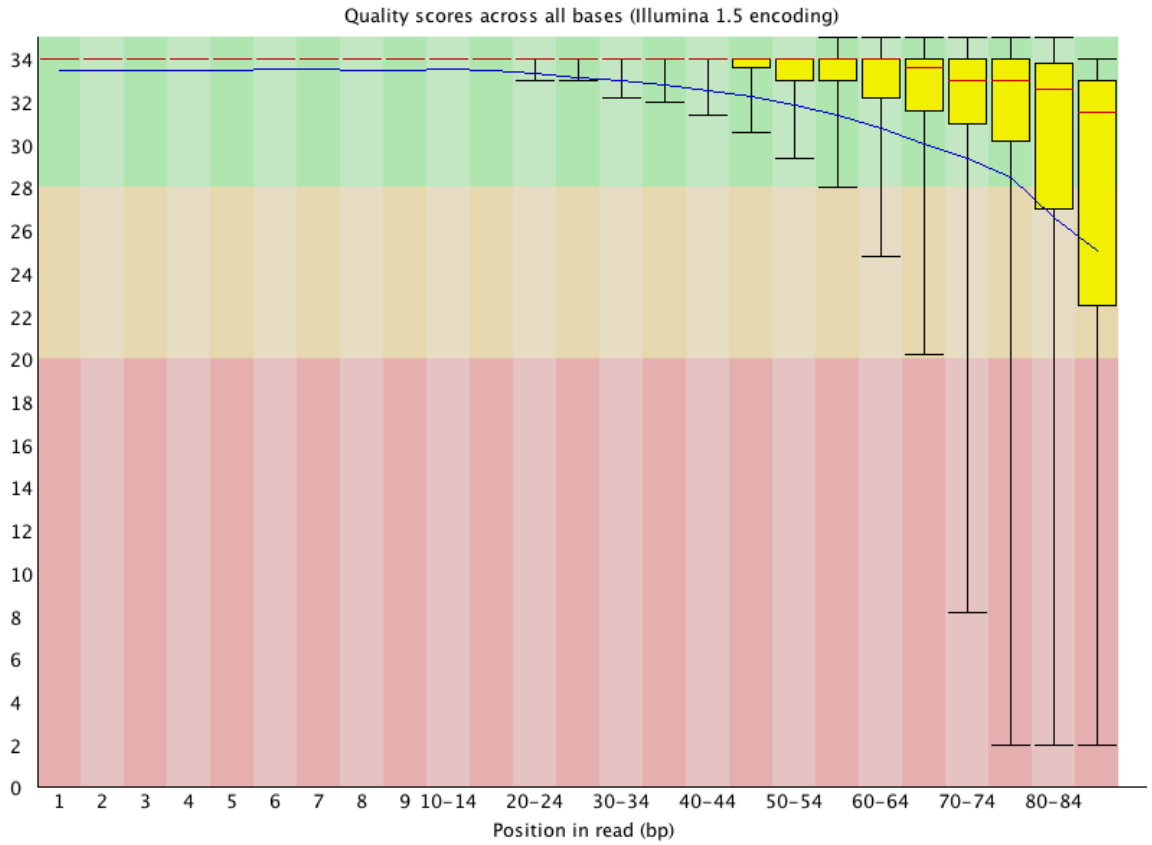

Figure 3.4 Per base sequence quality for the 'Titan' treatment. The y-axis is the Phred score and the x-axis is the base number in the sequencing read. The red line shows the median Phred score, the yellow boxes show the interquartile range, the blue line is the mean quality score and the whiskers show the 10th and 90th percentiles.

Analysis of the raw data also reveals deviation from the theoretical per sequence GC content, particularly in the treatments (Figure 3.5 – 3.8). These differences are unlikely to be due to changes in gene expression as a gene would have to be extremely highly differentially expressed to cause a peak given the number of sequencing reads. Instead, this suggests that RNA from both the pathogen and other organisms may be included in these data. The susceptible treatment shown in Figure 3.8 is particularly asymmetric, which is likely due to increased quantity of both pathogen and perhaps secondary infections of opportunistic pathogens or saprophytes.

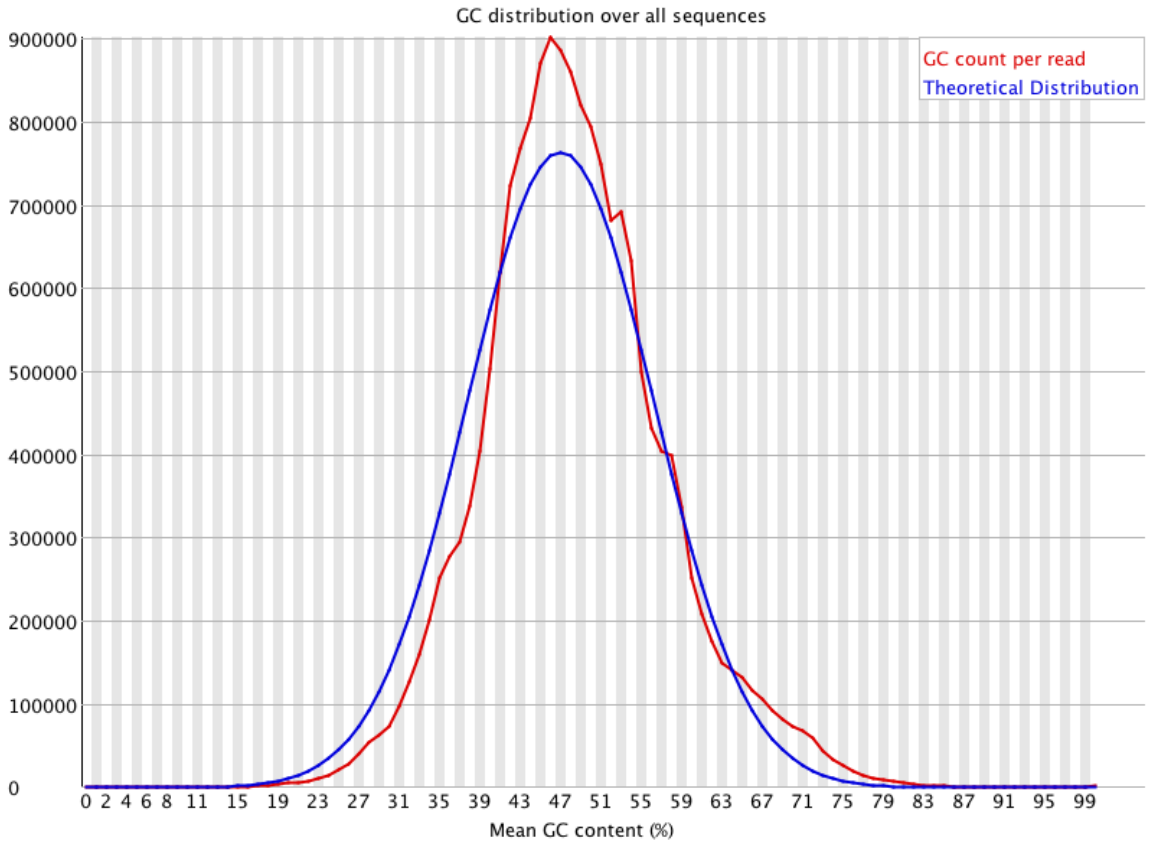

Figure 3.5 Per read GC content of Illumina sequence data for the 'Latham' control. The blue line shows the theoretical modeled normal distribution for each data set and the red line shows the actual GC count per read.

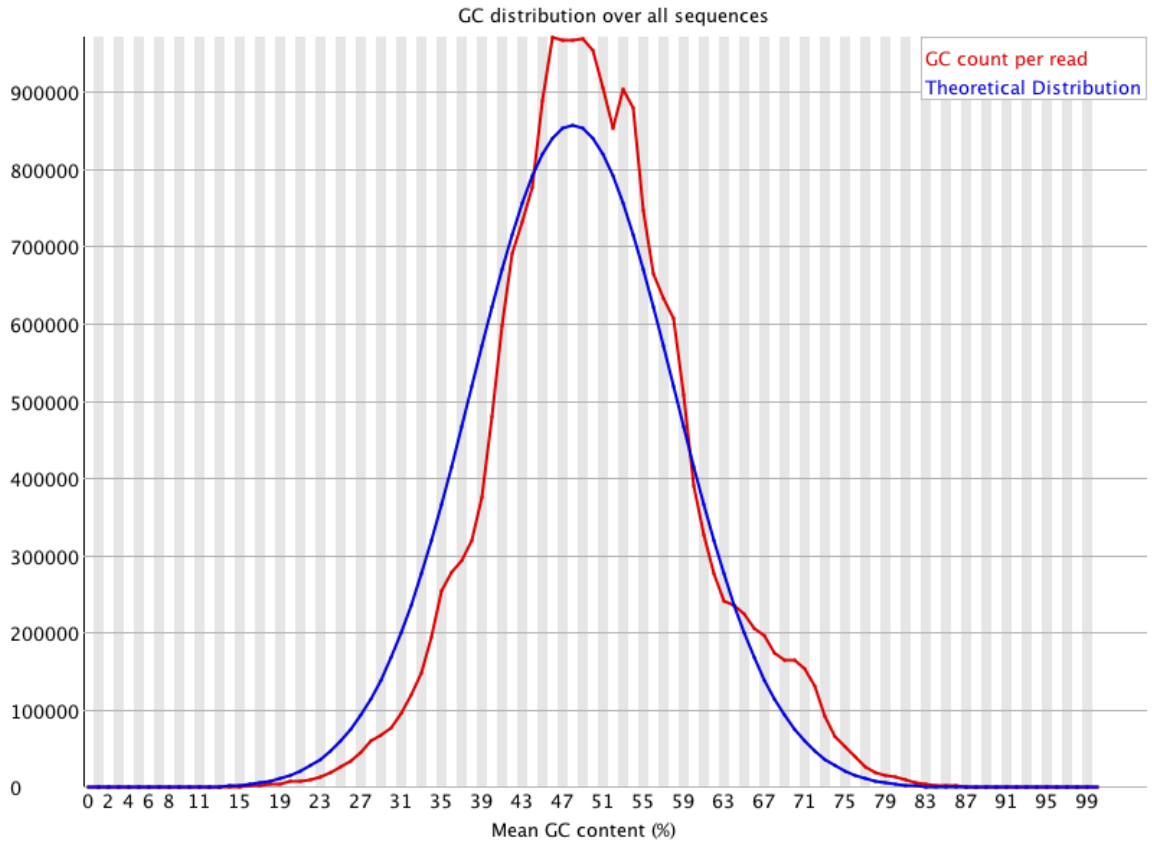

Figure 3.6 Per read GC content of Illumina sequence data for the 'Titan' control. The blue line shows the theoretical modeled normal distribution for each data set and the red line shows the actual GC count per read.

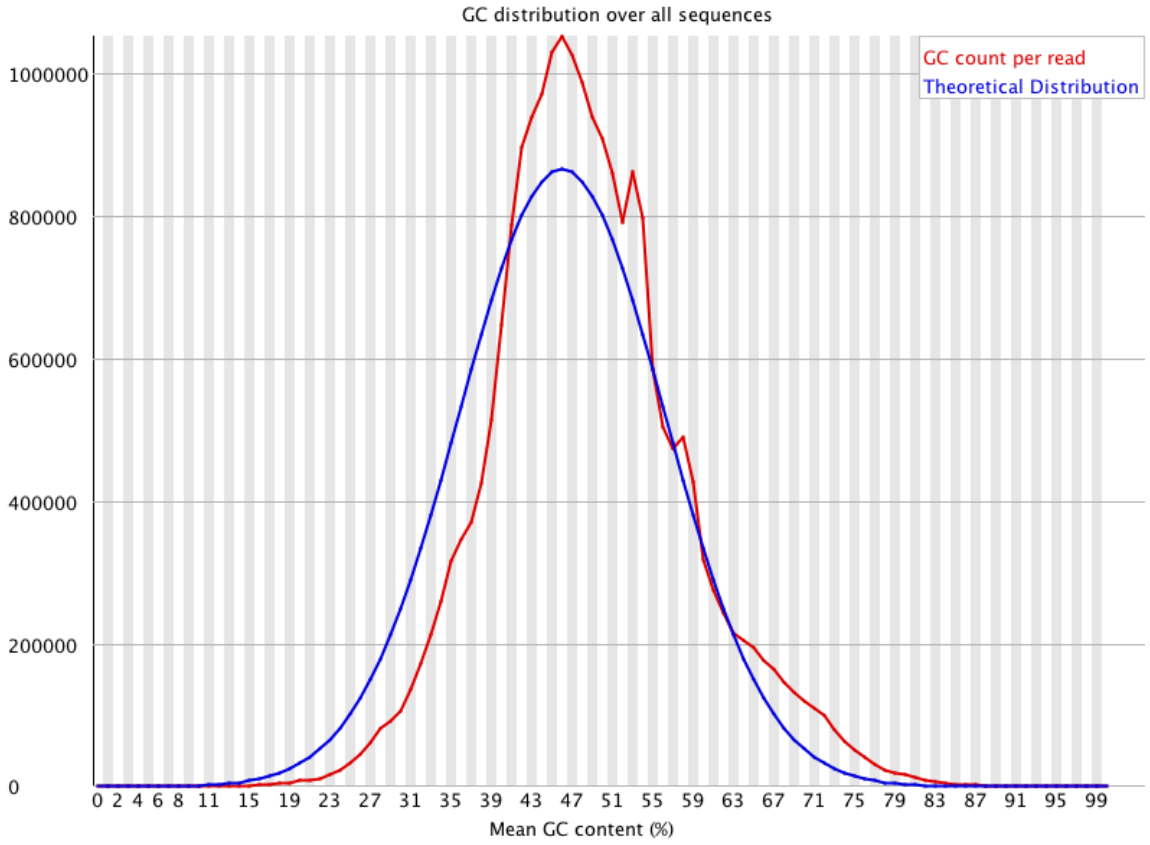

Figure 3.7 Per read GC content of Illumina sequence data for the 'Latham' treatment. The blue line shows the theoretical modeled normal distribution for each data set and the red line shows the actual GC count per read.

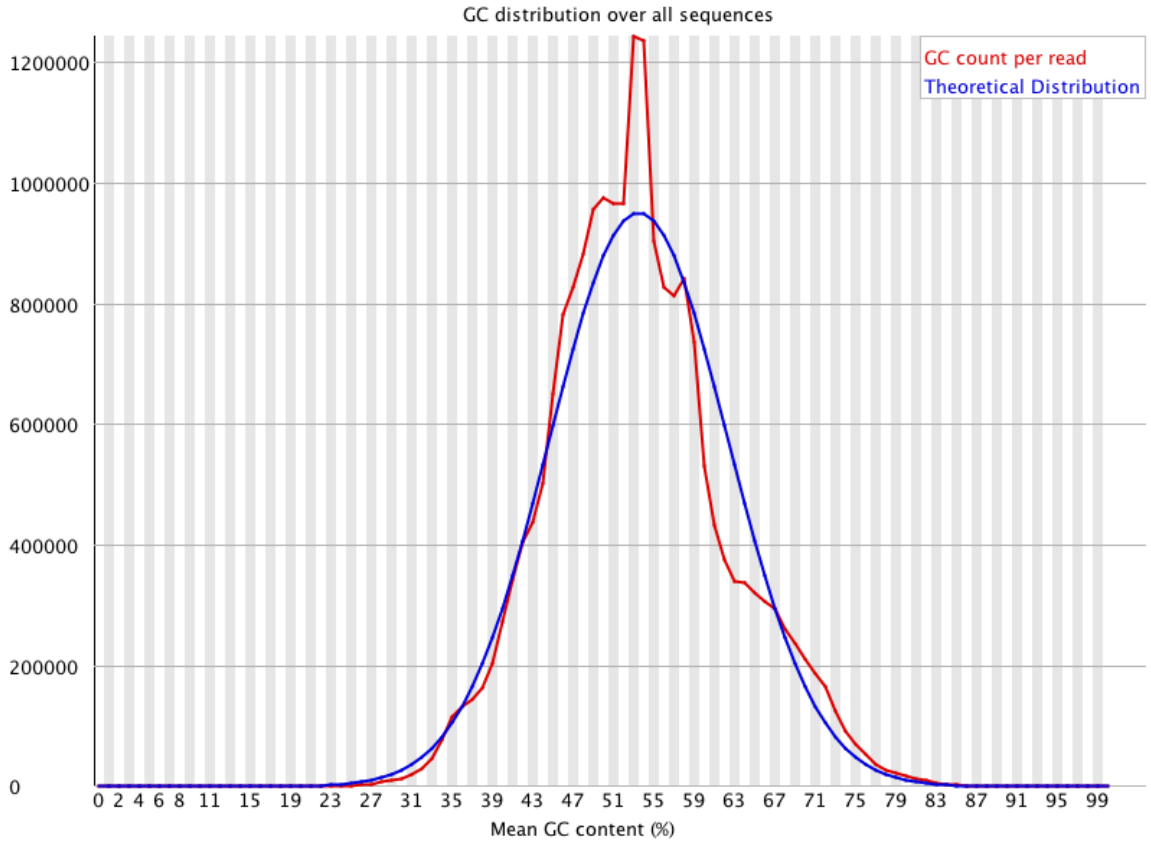

Figure 3.8 Per read GC content of Illumina sequence data for the 'Titan' treatment. The blue line shows the theoretical modeled normal distribution for each data set and the red line shows the actual GC count per read.

The analysis of the untrimmed 86nt Illumina data with Bowtie/TopHat and Cufflinks (default settings) detected 15,164 expressed genes in the 'Latham' control, 14,896 expressed genes in the 'Titan' control, 15,716 expressed genes in the 'Latham' treatment, and 13,991 expressed genes in the 'Titan' treatment. The union of the four data sets included 16,956 unique and putatively expressed genes with an intersection of 12,746 genes (Figure 3.9).

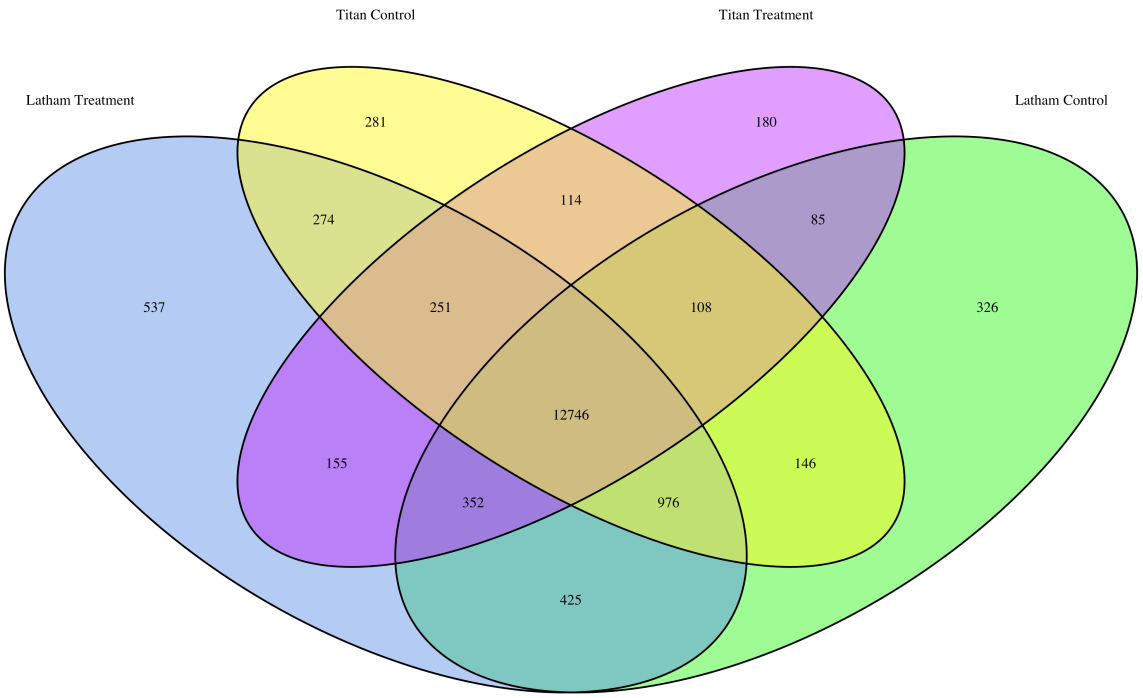

Figure 3.9 A four-way Venn diagram showing overlap of the 16,956 genes detected in the RNAseq experiment. The intersection of the four data sets contains 12,746 genes.

Most genes that were uniquely detected in any one condition had low FPKM values and were likely chance detections of genes with very low expression levels. This suggests that the sequencing depth was not adequate for genes with the lowest expression levels.

The correlation between the predicted FPKM for a gene in the 'Latham' control versus the 'Latham' treatment was stronger (Figure 3.10, Pearson's correlation = 0.9339) than that between the 'Titan' control and 'Titan' treatment (Figure 3.11, Pearsons correlation = 0.7080). Reduced correlation between the susceptible control and treatment was likely due to the absence of the third developmental time point in the susceptible treatment (roots were too necrotic to obtain an RNA sample), but may also be due to more dramatic changes in gene expression in response to P. rubi.

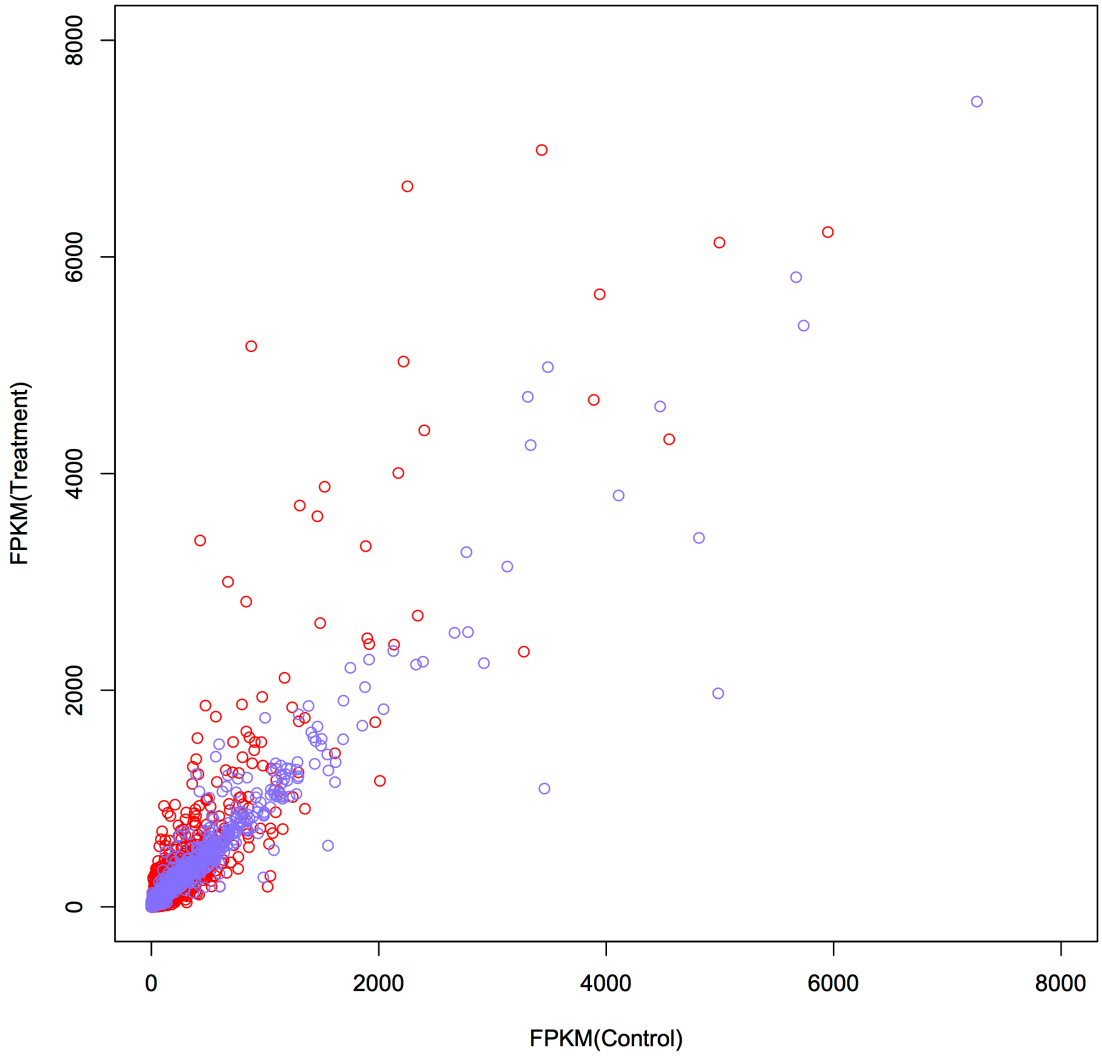

Figure 3.10. A comparison between treatments and controls by genotype for gene expression as predicted when the RNA-seq reads were aligned to ab initio gene models from the strawberry genome. Linear regression between the susceptible treatment and control had a lower correlation ('Titan', shown in red, Pearson's correlation = 0.841409) compared to the linear regression performed between the resistant treatment and control ('Latham', shown in purple, Pearson's correlation= 0.9664046). Increased divergence between the susceptible treatment and susceptible control may be due do to pathogen-induced changes in expression or may be due to differences caused by the lack of the final developmental time point where the roots were too degraded to collect an RNA sample. The similarity in expression levels between the treatment and the control for 'Latham' demonstrates reproducibility of this method.

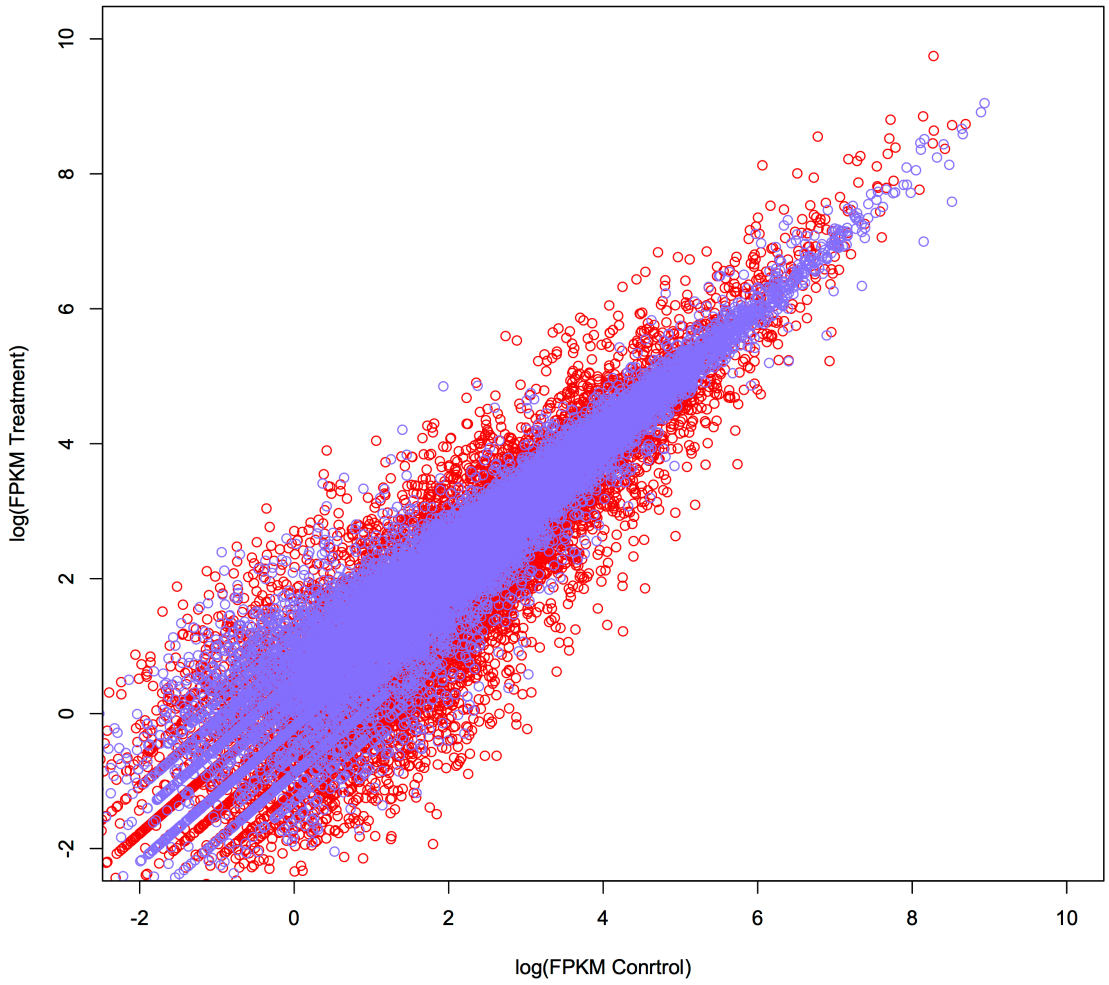

Figure 3.11. A comparison between treatments and controls on a log scale by genotype for gene expression as predicted when the RNA-seq reads were aligned to ab initio gene models from the strawberry genome. In red, the susceptible ('Titan') control is compared to the susceptible treatment and shows increased divergence from correlation when compared to the resistant ('Latham') genotype (purple). Increased divergence between the susceptible treatment and susceptible control may be due do to pathogen-induced changes in expression or may be due to differences caused by the lack of the final developmental time point where the roots were too degraded to collect a meaningful RNA sample. The similarity in expression levels between treatment and control demonstrates reproducibility of the method.

To identify genes differentially expressed after exposure to P. rubi, the F. vesca ab initio models for the 16,956 genes representing the union of the four data sets were BLASTED and re-

annotated using Blast2go. Twenty BLAST hits, Gene Ontology Terms, and Enzyme codes were obtained for all sequences. Differential expression analysis was performed to compare treatments to controls using Cufflinks with Benjamini Hochberg false discovery rate correction for multiple testing. Terms associated with major disease resistance pathways were used to filter the BLAST hits in order to identify genes potentially playing a role in disease resistance. This level of analysis included genes with GO terms associated with systemic acquired resistance (SAR), the salicylic acid (SA) pathway or the jasmonic acid (JA) pathway and also simply any defense response. A summary of genes and their GO terms that met these criteria is found in Appendix 1. Numerous genes were up regulated in the PR-10 family and these genes are shown separately in Table 3.2 for clarity. The NPR-1 family proteins are shown separately in Table 3.3.

Table 3.2 PR-10 genes (mal d) expressed in 'Latham' and 'Titan' showing generalized up regulation in 'Latham' with a mix of up or down

Gene ID	GO Description	Latham Control	Latham Treatment	Significantly up or down regulated in Latham?	Titan Control	Titan Treatment	Significantly up or down regulated in Titan?
gene32193	major allergen mal d	566.23	1386.24	up	394.18	1363.63	up
gene07038	mal d	7603.17	8494.85	up	4553.48	4316.65	down
gene07025	mal d	208.03	417.40	up	180.69	195.63	-
gene07040	mal d	3310.25	4708.11	up	1915.67	2425.42	up
gene07023	mal d	26.25	30.48	-	8.69	19.00	-
gene07045	mal d	36.01	66.74	up	27.11	3.88	down
gene07046	mal d	421.81	1065.36	up	385.54	779.23	up
gene07024	mal d	226.60	358.85	up	140.14	116.00	-
gene07041	mal d	671.72	1214.05	up	392.14	838.73	up
gene05091	mal d	255.95	459.31	up	99.65	138.24	up
gene07042	mal d	624.51	1065.58	up	454.34	246.21	down
gene07043	mal d	112.93	182.39	up	80.19	222.89	up
gene05092	mal d	171.07	291.15	up	106.40	87.32	-
gene07044	mal d	93.83	255.71	up	143.64	142.80	-
gene04930	major allergen mal d	999.72	1744.94	up	886.76	1325.15	up
gene05157	major allergen mal	248.44	657.62	up	280.98	715.40	up

	d						
gene07039	mal d	38.62	76.84	up	19.48	30.26	-
gene07022	mal d	67.45	227.73	up	89.00	159.60	up
gene07035	mal d 1-like	510.32	729.33	up	238.30	654.23	up

Table 3.3 NPR1-like genes expressed in both the resistant and susceptible genotypes, showing increased expression in the resistant treatment compared to the resistant control and opposite changes between the susceptible treatment and control.

Gene ID	GO Description	FPKM Resistant Control	FPKM Resistant Treatment	Significantly up or down regulated in Latham?	FPKM Suseptible Control	FPKM Suseptible Treatment	Significantly up or down regulated in Titan?
gene20034	NPR1-like protein	21.47	31.43	up	15.06	9.01	down
gene28686	NPR1 NIM1-like regulatory protein	11.2	19.52	up	5.51	4.48	-

Other significant defense responses include aspects of primary or secondary metabolism, but would not pick up GO terms associated with defense because they are also ordinary parts of plant growth and development. These include those genes involved in lignin biosynthesis, secondary metabolism of defense compounds or primary metabolism. Table 3.4 shows cytochrome P450 genes that are differentially expressed in 'Latham' because these genes are often involved in production of toxic secondary metabolites.

Table 3.5 shows genes differentially expressed in 'Latham' that may be involved lignin biosynthetic processes, while Table 3.6 shows genes differentially expressed in 'Latham' that may be involved in the Tricarboxylic Acid Cycle (TCA).

Table 3.4 Cytochrome P450 genes and their expression in the four conditions. Most are have increased expression in both 'Titan' and 'Latham' treatments with the exception of an elicitor induced cytochrome P450 from rice which is notated with an asterisk.

Gene ID	Latham Control	Latham Treatment	Significantly up or down regulated in Latham?	Titan Control	Titan Treatment	Significantly up or down regulated in Titan?	GO Terms
gene05277 *	161.73	248.33	up	138.91	19.62	down	F:heme binding; F:flavonoid 3',5'-hydroxylase activity; F:electron carrier activity; P:oxidation reduction
gene29802	244.28	653.65	up	184.45	227.72	up	F:heme binding; F:electron carrier activity; P:oxidation reduction; F:taxane 13-alpha-hydroxylase activity
gene28893	214.08	558.45	up	174.12	225.49	up	F:heme binding; F:electron carrier activity; P:oxidation reduction; F:taxane 13-alpha-hydroxylase activity
gene20071	126.98	276.33	up	84.09	121.82	up	F:heme binding; F:electron carrier activity; P:oxidation reduction; F:taxane 13-alpha-hydroxylase activity
gene20070	77.83	167.41	up	55.75	92.38	up	F:heme binding; F:electron carrier activity; P:oxidation reduction; F:taxane 13-alpha-hydroxylase activity
gene27843	55.38	77.93	up	43.43	88.93	up	C:microsome; P:flavonoid biosynthetic process; F:flavonoid 3'-monooxygenase activity; C:plasma membrane; F:electron carrier activity; F:heme binding; F:p-coumarate 3-hydroxylase activity; P:lignin biosynthetic process; C:integral to membrane; P:coumarin biosynthetic process; C:endoplasmic reticulum; C:mitochondrion; P:oxidation reduction
gene26894	3.59	19.76	up	17.64	12.21	-	F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:monooxygenase activity; F:iron ion binding
gene27088	6.25	16.69	up	2.52	0.97	-	F:oxidoreductase activity; F:iron ion binding

gene23938	2.92	10.62	up	4.27	6.38	-	F:heme binding; F:isoflavone 2'-hydroxylase activity; F:electron carrier activity; P:oxidation reduction
gene04522	6.03	8.91	up	4.73	5.57	-	F:oxidoreductase activity; F:iron ion binding

Table 3.5 Differentially expressed genes in 'Latham' with GO annotations associated with lignin biosynthesis and their corresponding values in 'Titan'

GeneID	FPKM Latham Control	FPKM 3 Latham Treatment	Significantly up or down regulated in Latham?	FPKM Titan Control	FPKM Titan Treatment	Significantly up or down regulated in Titan?	gene id	GO DATA
gene02141	1295.85	1776.75	up	1047.23	723.89	down	caffeoyl- 3-o-methyltransferase	F:caffeoyl-CoA O-methyltransferase activity; F:metal ion binding; P:lignin biosynthetic process; P:methylation
gene28114	58.5506	82.79	up	48.0575	95.02	up	laccase 3	F:laccase activity; C:endomembrane system; F:copper ion binding; F:L-ascorbate oxidase activity; C:apoplast; P:lignin catabolic process; P:oxidation reduction
gene27843	55.3832	77.93	up	43.4308	88.93	up	cytochrome p450	C:microsome; P:flavonoid biosynthetic process; F:flavonoid 3'-monooxygenase activity; C:plasma membrane; F:electron carrier activity; F:heme binding; F:p-coumarate 3-hydroxylase activity; P:lignin biosynthetic process; C:integral to membrane; P:coumarin biosynthetic process; C:endoplasmic reticulum; C:mitochondrion; P:oxidation reduction
gene07425	12.6463	21.49	up	7.639	9.00	-	protein	P:arsenite transport; F:cofactor binding; F:zinc ion binding; C:membrane; P:indole glucosinolate catabolic process; P:immune

								response; ; P:phytochelatin biosynthetic process; P:callose deposition in cell wall during defense response; P:lignin biosynthetic process; P:response to arsenic; P:defense response to bacterium; C:cytosol; F:oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; F:cadmium ion binding; P:cell death; P:response to cadmium ion; P:oligopeptide transport; F:arsenite transmembrane-transporting ATPase activity; F:copper ion binding; F:glutathione gamma-glutamylcysteinyltransferase activity
gene23954	4.16121	15.20	up	6.31403	7.87	-	cinnamyl alcohol dehydrogenase	P:lignin biosynthetic process; F:cinnamyl-alcohol dehydrogenase activity; P:oxidation reduction; F:zinc ion binding
gene18728	5.83944	12.08	up	1.98104	6.48	up	laccase 110a	P:response to copper ion; F:copper ion binding; C:cytoplasmic membrane-bounded vesicle; F:L-ascorbate oxidase activity; C:apoplast; F:laccase activity; P:vegetative to reproductive phase transition of meristem; P:lignin catabolic process; P:oxidation reduction
gene16163	1.06348	5.25	up	0.435691	3.62	up	putative laccase [Rosa hybrid cultivar]	F:laccase activity; F:copper ion binding; C:apoplast; P:lignin catabolic process; P:oxidation reduction

Table 3.6 Differentially expressed genes in ‘Latham’ with GO annotations associated with the Tricarboxylic Acid Cycle (TCA) and the corresponding values in ‘Titan’

GeneID	FPKM Latham Control	FPKM 3 Latham Treatment	Significantly up or down regulated in Latham?	FPKM Titan Control	FPKM Titan Treatment	Significantly up or down regulated in Titan?	Description	GO DATA
gene05086	595.43	1,501.22	up	421.88	477.55	up	atp citrate lyase	P:acetyl-CoA biosynthetic process; P:cellular carbohydrate metabolic process; F:binding; C:citrate lyase complex; C:plasma membrane; F:ATP citrate synthase activity; F:succinate-CoA ligase (ADP-forming) activity; F:lyase activity; C:cytosol P:carotenoid biosynthetic process; P:starch biosynthetic process; F:ATP citrate synthase activity; P:positive regulation of flower development; P:aging; P:anthocyanin accumulation in tissues in response to UV light; P:chlorophyll biosynthetic process; C:cytosol; P:positive regulation of cell size; P:regulation of embryonic development; C:citrate lyase complex; F:lyase activity; P:leaf
gene04096	510.06	1,007.01	up	386.37	212.69	down	atp-citrate lyase a-1	

								development; P:wax biosynthetic process; P:acetyl-CoA biosynthetic process; F:ATP binding; F:succinate-CoA ligase (ADP-forming) activity P:glyoxylate cycle; C:cytosol; F:NAD or NADH binding; F:copper ion binding; P:NADP metabolic process; P:response to salt stress; P:defense response to bacterium; C:plasmodesma; P:tricarboxylic acid cycle; F:isocitrate dehydrogenase (NADP+) activity; F:protein binding; F:magnesium ion binding; P:isocitrate metabolic process; P:response to zinc ion; P:response to cadmium ion; C:plasma membrane; C:apoplast
gene00517	333.91	387.72	up	209.12	195.09	-	nadp-specific isocitrate dehydrogenase	
gene17266	221.70	270.03	up	154.85	207.19	up	nad+ dependent isocitrate dehydrogenase subunit 1	F:NAD or NADH binding; F:magnesium ion binding; P:tricarboxylic acid cycle; F:isocitrate dehydrogenase (NAD+) activity; F:zinc ion binding; P:isocitrate metabolic process; C:mitochondrion C:cytosol; F:copper ion binding; C:cell wall; P:succinate metabolic process;
gene09138	118.67	142.85	up	89.44	70.46	down	succinyl-ligase	C:mitochondrion; F:ATP citrate synthase activity; P:succinyl-CoA metabolic process; P:tricarboxylic acid

gene00585	93.27	118.64	up	62.76	90.10	up	citrate synthase	cycle; F:succinate-CoA ligase (ADP-forming) activity; F:ATP binding; P:response to cadmium ion; F:succinate-CoA ligase (GDP-forming) activity C:plasmodesma; P:fatty acid beta-oxidation; F:citrate (Si)-synthase activity; P:tricarboxylic acid cycle; C:glyoxysome; C:cytosol; C:plastid; P:glyoxylate cycle

3.5. Discussion

Novel implementation of RNAseq has revealed numerous changes in the gene expression in ‘Latham’ during defense against P. rubi. These changes suggest that resistance in ‘Latham’ is an active process involving PTI, ETI and metabolic changes. Not surprisingly, many differentially expressed genes were either up regulated or down regulated by a similar magnitude between the controls and treatments in both genotypes and thus it is clear that the immune response of ‘Titan’ is also not completely compromised but is lacking key components compared to ‘Latham’. However, there are dramatic differences between genotypes in expression of key genes in well-studied defense pathways.

The response of ‘Latham’ to P. rubi is not typical of R-gene mediated resistance, but there is evidence that an R-gene may be involved. This is evident in observations of changes in expression of conserved genes that act either upstream or downstream to effector recognition. Several members of the chaperone complex, including heat shock protein (HSP) 90 and heat

shock cognate (HSC) 70 were detected and annotated. These proteins are involved in both folding of R-genes and breakdown of R-genes after effector recognition(Shirasu, 2009). Neither HSP nor HSC was differentially expressed in 'Latham' or 'Titan', but SGT1, which interacts with HSP90 and HSC70 in the chaparone complex was significantly up regulated in the 'Latham' treatment (FPKM=259.3) compared to the 'Latham' control (FPKM=200.2) while being down regulated in the 'Titan' treatment (FPKM=85.5) compared to the 'Titan' control (FPKM=128.4). SGT1 has been shown to be important in R protein accumulation following induction with the oomycete pathogen Hyaloperonospora parasitica (a downy mildew affecting Brassica species)(Azevedo et al., 2006). Thus the observed changes in SGT1 may suggest that R gene monitoring is increased in 'Latham' and possibly suppressed in 'Titan'.

Further evidence suggesting that ETI may be involved is the activation of key genes in the systemic acquired resistance (SAR) pathway. These genes include both activators of SA mediated response and repressors of JA mediated response. First, the non-expressor of PR1 (NPR1) was examined because it is highly conserved among plants (Chern et al., 2001) and is critical in SA mediated resistance. NPR1 exists in the cytoplasm as an oligomer of two NPR1 subunits joined by disulfide bonds and upon a change in redox the oligomer is split into two monomeric subunits that move into the nucleus (Mou et al., 2003). After entry into the nucleus NPR1 interacts with TGA transcription factors to induce expression of PR genes (Zhou et al., 2000). Two NPR1-like genes (gene20034 and gene28686) were detected in the experiment and were both significantly up regulated in the resistant treatment compared to the resistant control after FDR control, while being down regulated or not differentially expressed between the susceptible treatment and control (Table 3.3).

Genes similar to TOPLESS (TPL), a flexible negative regulator of auxin and JA mediated signaling (Pauwels et al., 2010; Causier et al., 2011), were also detected (genes; 15951, 12546, 31001). All three copies were up regulated in the 'Latham' treatment compared to the 'Latham' control. One copy, gene31001, was up regulated in the 'Titan' treatment (although with a lower log fold change compared to 'Latham') and the other two were not differentially expressed in 'Titan'. Repression of auxin or JA signaling via TPL requires NINJA (similar to gene11828), which is the actual transcriptional repressor. NINJA was detected, but was not differentially expressed in either 'Latham' or 'Titan'. Because NINJA is known to be induced by MeJA (Pauwels et al., 2010) it appears that neither 'Latham' nor 'Titan' has a highly activated JA pathway in response to P. rubi. However, the increased expression of TLP in 'Latham' may suggest repression of the JA and/or the auxin pathway. Potential JA antagonism is consistent with activation of SA mediated defense.

Several negative regulators of cell death were down regulated in the resistant treatment, while being up regulated in the susceptible treatment, including a 14-3-3 family protein and LSD1. The 14-3-3 family has diverse functions (Bridges, 2005), but its members have long been known for their role in inhibition of apoptotic cell death in animals (Fu and Subramaniam, 2000). Gene 12042 was annotated as 14-3-3-like protein gf14 kappa and is similar to a 14-3-3-like protein gf14-e, which was recently shown to function as a negative regulator of cell death in rice (Manosalva and Bruce, 2011). This gf14 kappa protein (gene 12042) is significantly down regulated in the 'Latham' treatment (FPKM = 230.8) compared to the 'Latham' control (FPKM=354.0) and is significantly and up regulated in the 'Titan' treatment (FPKM=216.5) compared to the 'Titan' control (FPKM=162.5).

Another negative regulator of cell death detected was LSD 1 which encodes a zinc finger protein that either suppresses a pro-death pathway or promotes an anti-death pathway (Dietrich et al., 1997). The LSD 1 like protein (gene10679) is expressed with an FPKM of 4.4 in the 'Latham' control and is zero in the resistant treatment. Conversely, this gene is not expressed in the susceptible control, but is induced and expressed with an FPKM of 13.7 in the susceptible treatment.

Furthermore, possible activation of PTI in 'Latham' was observed with a significantly increased expression of MAPKK2 (gene12639) in the 'Latham' treatment (FPKM = 31.37) compared to the 'Latham' control (FPKM=49.19), while MAPKK2 expression in 'Titan' was slightly reduced in the treatment (FPKM=21.43) compared to the 'Titan' control (FPKM=22.26).

While increased metabolic activity in 'Latham' may lead to increased root vigor, it is likely to be actively conditioned by mechanisms of resistance involving defense pathways. Several other genes that potentially act downstream from PAMP recognition were differentially expressed. Examples of PTI specific responses include increased expression of genes involved in callose deposition and lignin biosynthesis (Quentin et al., 2009). Specific genes differentially expressed include udp-glucosyl transferase 74b1 (gene00713) and lignin biosynthesis genes such as caffeoyl- 3-o-methyltransferase (gene02141) were induced.

Other defense response genes detected include those involved in the indole glucosinolate pathway. A key example is the ABC transporter PEN3 (similar to gene25666), which delivers a toxic load of secondary metabolites to the apoplast (Bednarek et al., 2009). The PEN3 ABC transporter was up regulated in the 'Latham' treatment (FPKM=103.9) compared to the 'Latham' control (FPKM=20.3) and was only slightly upregulated in the 'Titan' treatment (FPKM=24.3) compared to the 'Titan' control (FPKM=14.9). PEN3 may work synergistically with the

observed up regulation of elicitor induced cytochrome P450 (gene05277). Cytochrome P450s are specifically important in synthesis of the secondary metabolites delivered by PEN3, including the synthesis of 4-methoxyindol-3-ylmethylglucosinolate which was recently shown to be dependent upon a cytochrome P450 and responsible for broad spectrum defense against fungi (Bednarek et al., 2009).

Other possible defense responses observed may have arisen from either ETI or PTI, including, increased expression of WRKY transcription factors. While the WRKY transcription factor family is quite large (Eulgem & Somssich, 2007), many WRKY transcription factors have been implicated in various plant defense pathways (Eulgem and Somssich, 2007; Pandey and Somssich, 2009; van Verk et al., 2011). The WRKY transcription factor encoded by gene21970 is particularly noteworthy because the log fold change was ~2.69 between the 'Latham' treatment (FPKM=13.4) and control (FPKM=0.9) and was slightly down regulated from between the 'Titan' control (FPKM=1.9) and the 'Titan' treatment (FPKM=0.8).

To examine this WRKY transcription factor further the amino acid sequence of the ab initio gene21970 was BLASTED to the NCBI nr database. The sequence is similar to Arabidopsis WRKY51 that has been shown to be regulated in a SA dependent manner and was one of only a few WRKY transcription factors examined that were strongly influenced by a mutation in NPR1 (Dong et al., 2003). This WRKY transcription factor may therefore be involved in the induction of defense genes such as the PR genes.

In the Rosaceae family, the PR-10 genes have been extensively studied because of their involvement in plant defense and because of their role in food allergies (Ebner et al., 1991; Gao et al., 2005). Nineteen members of the PR-10 family were detected in this experiment and 18 were significantly up regulated in 'Latham', while in 'Titan', only 10 were up regulated and 3

were down regulated. In apple leaves it has been shown that several classes of PR-10 are inducible with salicylic acid (Ziadi et al., 2001). PR-10 was also shown to be induced after inoculation with the hemi-biotrophic pathogen Venturia inequalis (apple scab) in a resistant apple cultivar following an "intense necrotic resistance reaction of leaf cells" (Chevalier et al., 2008). Furthermore, the induction of PR-10 genes through either the JA pathway or the SA pathway has been demonstrated in rice (McGee & Hamer, 2001) and hot pepper (Park et al. 2004).

The function of PR-10 proteins is not completely understood, but several studies suggest ribonuclease activity, which defends against viruses, and antimicrobial activity with an unknown mode of action against bacteria, fungi, and oomycetes. Recombinant Capsicum annuum (hot pepper) PR-10 was effective against both Tobacco Mosaic Virus (TMV) and against Phytophthora capsici in culture (Park et al. 2004). In Oxalis tuberosa Mol. PR-10 was tested against various plant pathogens and was found to inhibit the growth of P. cinnamomi in a dose-dependent fashion, but not growth of six other Phytophthora species at the concentrations tested (Flores et al. 2002). Interestingly, P. cinnamomi and P. rubi are closely related, both being members of clade 7 of a recent phylogenetic study of Phytophthora species (Blair et al. 2008) while the other species tested by Flores et al. belong to other clades. Differential activation of PR-10 genes suggests that some of the PR-10 genes are activated by different pathways or divergent transcription factors.

Together, these results show that the response of 'Latham' is an active process and is likely to involve both ETI and PTI. Based on these results, it is hypothesized that an R-gene plays a role in resistance, but that P. rubi is successful at escaping HR in 'Latham'. Furthermore,

by observing decreased expression in key genes, P. rubi may subvert PTI in ‘Titan’, while either a stronger PTI response in ‘Latham’ or ETI may lead to more successful disease resistance.

Specific changes in the gene expression of ‘Latham’ included a dramatic increase in PEN3 and an elicitor-induced cytochrome P450, which may act synergistically to limit penetration of P. rubi zoospores. Together with other defense responses, such as callose deposition, lignification, and increased expression of PR genes, the PEN3-cytochrome P450 response may be largely responsible for the observation that roots in ‘Latham’ have induced resistance.

Furthermore it is well established that enhanced lignification and callose deposition are metabolically expensive defense responses. Thus increased metabolic flux through the TCA as indicated by increased expression of ATP citrate synthase may allow ‘Latham’ to successfully respond. These data are in alignment with the work by Pattison et al. (2004; 2007) which suggested a change in gene expression and the work by Graham et al. (2011) which suggests the PRR resistance QTL in ‘Latham’ associate with vigor QTL in identified without pathogen exposure. Alignment of these data to the raspberry genome (upon completion) along with the BAC sequences of Graham et al. (2011) and markers from Pattison et al. (2007) will help to further illuminate the mechanisms of resistance in ‘Latham’.

3.6. References

Alkharouf, N.W. et al. 2006. Timecourse microarray analyses reveal global changes in gene expression of susceptible Glycine max (soybean) roots during infection by Heterodera glycines (soybean cyst nematode). *Planta*, 224(4), pp.838–852.

Azevedo, C. et al. 2006. Role of SGT1 in resistance protein accumulation in plant immunity. *The EMBO journal*, 25(9), pp.2007–2016.

Barber, R. et al. 2005. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics*, 21, pp.389–395.

Bednarek, P. et al. 2009. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, 323(5910), pp.101–106.

Berrocal-Lobo, M. Molina, A. & Solano, R. 2002. Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *Plant Journal*, 29(1), pp.23–32.

Blair, J.E. et al. 2008. A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. *Fungal genetics and biology*, 45(3), pp.266–277.

Boller, T. & He, S.Y. 2009. Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. *Science*, 324(5928), pp.742–744.

Bridges, D. 2005. 14-3-3 Proteins: A Number of Functions for a Numbered Protein. *Science's STKE*, 2005(296), pp.re10–re10.

Causier, B. et al. 2011. The TOPLESS Interactome: A Framework for Gene Repression in

Arabidopsis. *Plant Physiology*. Prepublication.

Chen, Y. & Halterman, D.A. 2011. Phenotypic characterization of potato late blight resistance mediated by the broad-spectrum resistance gene RB. *Phytopathology*, 101(2), pp.263–270.

Chern, M.S. et al. 2001. Evidence for a disease-resistance pathway in rice similar to the NPR1-mediated signaling pathway in Arabidopsis. *Plant Journal*, 27(2), pp.101–113.

Chevalier, M. et al. 2008. Specific activation of PR-10 pathogenesis-related genes in apple by an incompatible race of Venturia inaequalis. *Biologia Plantarum*, 52(4), pp.718–722.

Dietrich, R. et al. 1997. A Novel Zinc Finger Protein Is Encoded by the Arabidopsis LSD1 Gene and Functions as a Negative Regulator of Plant Cell Death. *Cell*, 88(5), pp.685–694.

Dong, J. Chen, C. & Chen, Z. 2003. Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Molecular Biology*, 51(1), pp.21–37.

Ebner, C. et al. 1991. Common epitopes of birch pollen and apples--studies by western and northern blot. *The Journal of allergy and clinical immunology*, 88(4), pp.588–594.

Eulgem, T. 2005. Regulation of the Arabidopsis defense transcriptome. *Trends in Plant Science*. 10(2), pp. 1360-1385

Eulgem, T. & Somssich, I.E. 2007. Networks of WRKY transcription factors in defense signaling. *Current Opinion In Plant Biology*, 10(4), pp.366–371.

Ewing, B. & Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3), pp.186–194.

Flores, T. et al. 2002. Ocatin. A novel tuber storage protein from the andean tuber crop oca with antibacterial and antifungal activities. *Plant Physiology*, 128(4), pp.1291–1302.

Fu, H. & Subramaniam, R. 2000. 14-3-3 Proteins: Structure, Function, and Regulation. *Annual Review of Pharmacological Toxicology*. 40(1), pp.617.

Gao, Z. de Weg, W. & Schaart, J. 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (Malus domestica). *Theoretical and Applied Genetics*. 11, pp.171-183.

Glazebrook, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Reviews in Phyyopsyhology*, 43, pp.205–227.

Graham, J. et al. 2011. Towards an understanding of the nature of resistance to Phytophthora root rot in red raspberry. *Theoretical And Applied Genetics*, 123(4), pp.585–601.

Hardham, A. 2005. Phytophthora cinnamomi. *Molecular Plant Pathology*, 6(6), pp.589–604.

Ingle, R. Carstens, M. & Denby, K. 2006. PAMP recognition and the plant-pathogen arms race. *Bioessays*, 28, pp.880–889.

Jeffers, S.N. & Aldwinckle, H. 1987. Enhancing detection of Phytophthora cactorum in naturally infested soil. *Phytopathology*, 77(10), pp.1475–1482.

Jeffers, S.N. & Martin, S. 1986. Comparison of 2 Media Selective for Phytophthora and Pythium Species. *Plant Disease*, 70(11), pp.1038–1043.

Jones, J.D.G. & Dangl, J.L. 2006. The plant immune system. *Nature*, 444, pp.323–329.

Kalde, M. et al. 2007. The syntaxin SYP132 contributes to plant resistance against bacteria and secretion of pathogenesis-related protein 1. In *PNAS*. PNAS. pp. 11850–11855.

Kamoun, S. 2006. A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annual Review of Phytopathology*, 44, pp.41–60.

Kamoun, S. Huitema, E. & Vleeshouwers, V. 1999. Resistance to oomycetes: a general role for the hypersensitive response? *Trends in Plant Science*, 4(5), pp.196–200.

Laxalt, A. Cassia, R. & Sanllorenti, P. 1996. Accumulation of cytosolic glyceraldehyde-3-phosphate dehydrogenase RNA under biological stress conditions and elicitor treatments in

potato. *Plant molecular Biology*. 30, pp.961-972.

Lorenzo, O. 2004. JASMONATE-INSENSITIVE1 Encodes a MYC Transcription Factor Essential to Discriminate between Different Jasmonate-Regulated Defense Responses in Arabidopsis. *The Plant Cell Online*, 16(7), pp.1938–1950.

Manosalva, P. & Bruce, M. 2011. Rice 14-3-3 protein (GF14e) negatively affects cell death and disease resistance. *The Plant Journal*. 68, pp. 777–787.

McGee, J. & Hamer, J. 2001. Characterization of a PR-10 pathogenesis-related gene family induced in rice during infection with Magnaporthe grisea. *Molecular Plant-Microbe Interactions*. 14(7), pp. 877–886.

Medzhitov, R. & Janeway, C.A. 1997. Innate immunity: impact on the adaptive immune response. *Current Opinion in Immunology*, 9(1), pp.4–9.

Mou, Z. Fan, W. & Dong, X. 2003. Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. *Cell*, 113(7), pp.935–944.

Murashige, T. & Skoog, F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, 15. pp.473-479.

Pandey, S. & Somssich, I.E. 2009. The role of WRKY transcription factors in plant immunity. *Plant Physiology*, 150(4), p.1648.

Park, C.-J. et al. 2004. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant Journal*, 37(2), pp.186–198.

Pattison, J.A. 2004. Investigating the Inheritance of Phytophthora Root Rot (Phytophthora fragariae var. rubi) Resistance in Red Raspberry (Rubus Idaeus L.) Using Molecular Markers, Unpublished Dissertation, Cornell University. pp.1-117

Pattison, J.A. Samuelian, S.K. & Weber, C.A. 2007. Inheritance of Phytophthora root rot

resistance in red raspberry determined by generation means and molecular linkage analysis. *Theoretical And Applied Genetics*, 115(2), pp.225–236.

Pattison, J.A. Wilcox, W. & Weber, C.A. 2004. Assessing the resistance of red raspberry (Rubus idaeus L.) genotypes to Phytophthora fragariae var. rubi in hydroponic culture. *Hortscience*, 39(7), pp.1553–1556.

Pauwels, L. et al. 2010. NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature*, 464(7289), pp.788–791.

Quentin, M. et al. 2009. Imbalanced Lignin Biosynthesis Promotes the Sexual Reproduction of Homothallic Oomycete Pathogens. *PLoS Pathogens*, 5(1), p.e1000264.

Shirasu, K. 2009. The HSP90-SGT1 Chaperone Complex for NLR Immune Sensors. *Annual Review of Plant Biology*, 60(1), pp.139–164.

Tajima, H. et al. 1999. Over-expression of GAPDH induces apoptosis in COS-7 cells transfected with cloned GAPDH cDNAs. *Neuroreport*, 10(10), p.2029.

van Verk, M.C. Bol, J.F. & Linthorst, H.J.M. 2011. WRKY transcription factors involved in activation of SA biosynthesis genes. *BMC Plant Biology*, 11, p.89.

Vleeshouwers, V. et al. 2000. The hypersensitive response is associated with host and nonhost resistance to Phytophthora infestans. *planta*, 210, pp.853–864.

White, T. et al. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols A guide to methods and applications*, pp.315–322.

Wilcox, W. 1989. Identity, virulence, and isolation frequency of seven Phytophthora spp. causing root rot of raspberry in New York. *Phytopathology*, 79(1), pp.93–101.

Wilcox, W. 1992. Phytophthora Root and Crown Rots. *Disease Identification Sheet No. 7*.

Wilcox, W. et al. 1993. Identity of a Phytophthora species attacking raspberry in Europe and

North America. *Mycological Research*, 97, pp.817–831.

Wilcox, W. Pritts, M. & Kelly, M. 1999. Integrated control of Phytophthora root rot of red raspberry. *Plant Disease*, 83(12), pp.1149–1154.

Zhou, J.M. et al. 2000. NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid. *Molecular Plant-Microbe Interactions*, 13(2), pp.191–202.

Ziadi, S. et al. 2001. Characterization in apple leaves of two subclasses of PR-10 transcripts inducible by acibenzolar-S-methyl, a functional analogue of salicylic acid. *Physiological and Molecular Plant Pathology*. 59, pp.33-43.

3.7. Supplemental Figures

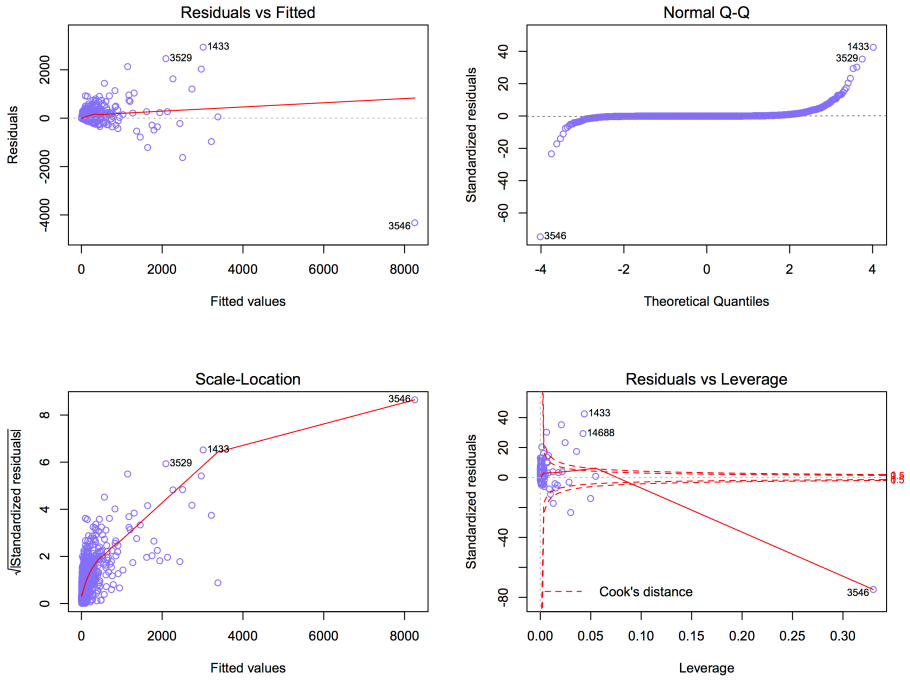

Figure 3.12. Titan linear regression diagnostics for figures 3.10 and 3.11 showing that data are following a fairly normal distribution with some diversion at the tails and illustrating that some outliers may be negatively influencing regression calculations.

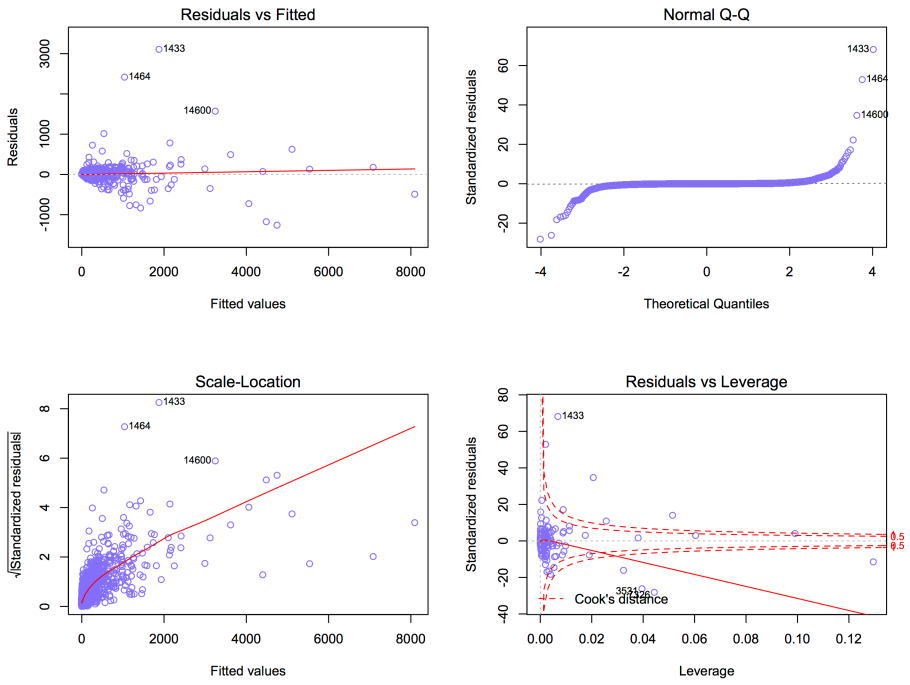

Figure 3.13. Latham linear regression diagnostics for figures 3.10 and 3.11 showing that data are following a fairly normal distribution with some diversion at the tails and illustrating that some outliers may be negatively influencing regression calculations.

3.7.1. Supplemental Note

In the susceptible genotype, 'Titan', there was an extreme outlier (Supplemental Figure 3.1) represented by glyceraldehyde 3-phosphate dehydrogenase (GAPDH, gene07063), which was the most highly expressed gene identified in the experiment. The FPKM for GAPDH was over 17,084, which is more than twice the level of the most highly expressed gene in the resistant treatment. GAPDH was not significantly differentially expressed between the resistant treatment and control, which had FPKM values of approximately 5813 and 5670 respectively. GAPDH is often regarded as a housekeeping gene and is thus often used as a control in quantitative RT-PCR, but it has been recognized that this is inappropriate in many cases (Barber et al. 2005). In human cell lines it was shown that overexpression of GAPDH induces cell death (Tajima et al.

1999). GAPDH was also previously shown to be up regulated in susceptible genotypes upon infection with other *Phytophthora* species such as *P. infestans* (Laxalt et al. 1996), and upon infection by the biotrophic nematode *Heterodera glycines* (Alkharouf et al. 2006). In the work by Laxalt et al. the same response was observed when tubers were treated with either *P. infestans* or a crude extract of *P. infestans* cell wall components, suggesting that the GAPDH was induced via a PTI pathway.

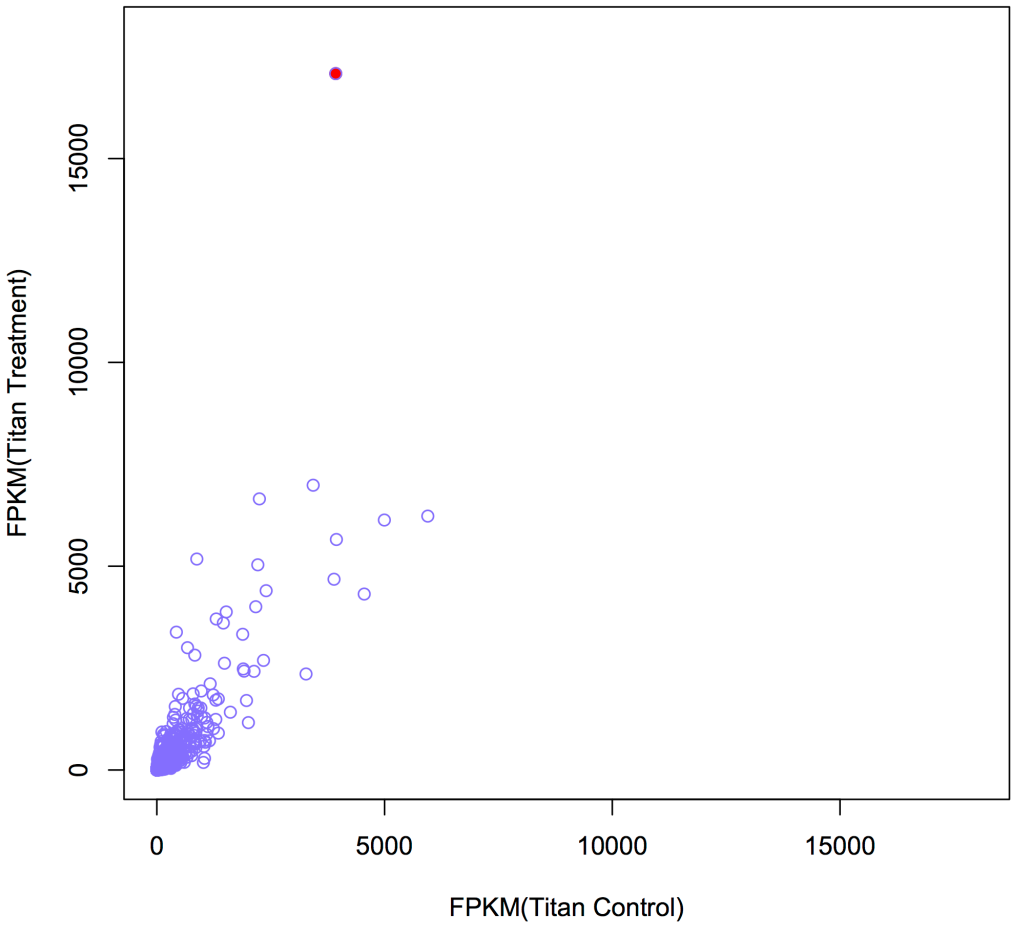

Figure 3.14. A comparison between treatments and controls for the susceptible genotype 'Titan', for gene expression as predicted when RNA-seq reads were aligned to ab initio gene models from the strawberry genome. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is shown as an extreme outlier in red.

Chapter 4

DISSECTING RESISTANCE TO *PHYTOPHTHORA RUBI* IN 'LATHAM' RED RASPBERRY WITH DE NOVO ASSEMBLY

4.1. Abstract

Phytophthora root rot (PRR) (caused by *Phytophthora rubi* (Wilcox and Duncan) Man in 't Veld, comb. nov. is a devastating soil born disease of red raspberry (*Rubus idaeus* L.). To investigate resistance in the cultivar 'Latham' and susceptibility of the cultivar 'Titan', a de novo assembly of RNAseq data was performed with over 70,600,000 86nt sequencing reads derived from a *P. rubi* challenge experiment. These data were assembled in three ways to generate an assembly for 'Latham', an assembly for 'Titan', and a combined assembly that included all the sequencing reads from both 'Latham' and 'Titan.' Assembly of the 'Latham' data alone resulted in 87,946 transcripts longer than 200 nt with an N50 of 974 and an N20 of 2,050. The 'Titan' assembly contained 71,512 transcripts greater than 200nt in length with an N50 transcript length of 1,002 nt and an N20 of 2,246. Assembly of the entire data set together resulted in 80,290 transcript sequences longer than 200nt with an N50 transcript size of 1,459 and an N20 of 3,548. A striking diversity in NPR1 transcripts and expression levels were revealed between the 'Latham' and 'Titan' assemblies. Furthermore, a greater diversity of transcripts annotated as R-genes were found in 'Latham' (523) compared to 'Titan' (314). Among these were several TIR-NBS-LRR type R-genes that are similar to RDR genes implicated in resistance to black spot caused by the hemibiotrophic fungus *Diplocarpon rosae* in rose (*Rosa multiflora*). These R-

genes are highly expressed in ‘Latham’, but not ‘Titan’. Together, these assemblies will aid in further investigation of the resistance to *P. rubi* and will also serve as a resource for rapid annotation of the *R. ideaus* genome.

4.2. Introduction

Phytophthora rubi (Wilcox and Duncan) Man in ’t Veld, comb. nov. causes the most economically damaging form of Phytophthora Root Rot (PRR) in red raspberry (*Rubus idaeus* L.) throughout the world (Wilcox et al. 1993; Wilcox 1989). Growers can implement a number of cultural methods to control disease, but these have only minimal impact compared to genetic resistance (Wilcox et al. 1999). For this reason, germplasm has been screened for sources of resistance (Crandall 1977; Pattison et al. 2004). Markers linked to loci conferring resistance would be useful for marker-assisted selection (MAS). Several previous studies have examined resistance in the cultivar ‘Latham’ and found that resistance was likely to follow a dominant two gene model (Pattison et al. 2007).

Pattison et al. (2007) also engaged in mapping with resistance gene analog polymorphisms using degenerate primers derived from conserved portions of known R-genes. A cloning effort from this work produced 75 resistance gene analogs (Samuelian et al. 2008). However, this is likely to be only a fraction of the R-gene content for raspberry since the *R. idaeus* genome is over 280MB (Arumuganathan & Earle 1991) and the small 145MB genome of *Arabidopsis thaliana* (Arumuganathan & Earle 1991) is known to contain over 200 NBS-LRR genes (Meyers et al. 2003).

Furthermore, Pattison et al. (2007,2004) suggested that resistance is likely to be induced following exposure to *P. rubi*. Induction of defense responses is common in plants and may

involve either pathogen associated molecular pattern (PAMP) triggered immunity (PTI) or effector triggered immunity (ETI) (Jones & Dangl 2006).

To investigate the resistance response in 'Latham', an RNAseq experiment was performed (see Chapter 3) using the woodland strawberry genome (*Fragaria vesca* L.) as a reference for alignment of over 70,600,000 sequencing reads. The alignment first results indicate that an R-gene/ETI plays a role in resistance, but that *P. rubi* is successful at escaping HR in 'Latham'. By observing decreased expression in key genes it was also hypothesized that *P. rubi* may be successful at subverting PTI in 'Titan', while either stronger PTI in 'Latham' or ETI may lead to more successful disease resistance.

Other key changes observed in the RNAseq experiment included a dramatic increase in PEN3 and an elicitor-induced cytochrome P450 in 'Latham'. PEN3 and secondary metabolites produced via a cytochrome P450 mediated pathway were recently shown to confer broad spectrum resistance to fungi (Bednarek et al. 2009) and may act to limit penetration of *P. rubi* zoospores. Together with other defense responses, such as callose deposition, lignification, and increased expression of PR genes, the PEN3-cytochrome P450 response may be partially responsible for the observation that roots in 'Latham' have induced resistance.

In order to confirm results from the alignment of data to the strawberry genome and to develop resources for further study of resistance to *P. rubi* the RNAseq data set was analyzed using an assembly-based approach. Sequences and expression levels from 'Titan' and 'Latham' were compared to further elucidate the resistance mechanisms at work in the 'Latham'.

4.3. Methods

4.3.1. RNA Isolation and Quality Analysis Summary

[Abridged: For complete methods see chapter 3]

Total RNA was extracted from approximately 25 mg of root tissue at each time point from four randomly selected resistant (one from each replicate) and four randomly selected susceptible plants (one from each replicate) for both treatments and controls using Qiagen RNeasy Plant Mini Kits (Qiagen Inc., Valencia, CA) according to the manufacturer's recommendation except for the addition of an extra buffer RPE wash and incubation for 30 seconds prior to centrifugation. Samples were further pooled within each available time point by variety and treatment creating a total of eleven samples (there is no sample for the final timepoint from 'Titan' due to root degradation). Each of these samples were then purified again with Qiagen RNeasy Plant Mini Kits according to the manufacturer's recommendation for purification of RNA with the addition of an extra wash with buffer RPE and an incubation of 30 seconds before centrifugation. All eleven samples were analyzed on a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California) so that RNA Integrity Numbers (RINs) could be calculated.

4.3.2. Sequencing Library Preparation and Sequencing Summary

[Abridged: For complete methods see chapter 3]

The experimental design implemented here was intended to maximize information gain for hypothesis generation in the investigation of the mechanisms of resistance to *P. rubi* in the red raspberry cultivar 'Latham'. To do this four lanes of sequencing data were produced from four distinct categories representing a resistant control (cv. 'Latham'), a susceptible control (cv. 'Titan'), a resistant treatment (cv. 'Latham' plus inoculation with *P.rubi*), and a susceptible treatment (cv. 'Titan' plus inoculation with *P.rubi*). To maximize the discovery of genes expressed only at a particular phase of disease progression mRNA was bulked from three time points (2 days, 5 days, and 20 days) across four biological replicates for each of the four classes

except in the case of the susceptible treatment because the roots were too highly degraded by the third time point to obtain a high quality RNA sample.

Library preparation for sequencing was performed according to the Illumina mRNAseq Protocol (Illumina, San Diego, CA), Part #1004898 (Rev. A September 2008). Single end 86nt sequencing was performed on the Illumina Genome Analyzer IIx (Illumina, San Diego, CA) at the Cornell University Life Sciences Core Laboratories Center according to the manufacturer's recommendations. Raw sequencing images from the Illumina Genome Analyzer IIx were processed using default settings on Illumina's 1.6.0 pipeline (RTA 1.6.32.0) with standard quality cross-calibration to a bacteriophage PhiX control lane.

4.3.3. De novo assembly and annotation

A de novo assembly of the RNAseq data was performed using Trinity (Grabherr et al. 2011) with a kmer size of 25 and default settings. For the combined assembly, all four data sets were concatenated into a single file and the assembly was performed with a singular file. In the 'Latham' and 'Titan' individual assemblies data the treatment and control lanes were concatenated prior to assembly and assemblies were performed with Trinity using a kmer size of 25 and default settings. Blast2go was used to annotate transcripts in each data set.

BLAST searches were performed for each sequence using an e-value threshold of 1.0×10^{-3} and up to 20 BLAST hits were recorded for each sequence into an XML database. Sequence descriptions, gene ontology (GO) terms and enzyme codes were assigned. Further analysis was performed using custom perl scripts. Assembly contiguity statistics were calculated using the abyss-fac script from ABySS version 1.3.1 (Simpson et al. 2009). The nucmer component of the MUMmer package (Delcher et al. 2002) was used to align all the Trinity transcripts to the strawberry ab initio gene models with default settings. The resulting delta file was filtered with

the -q & -r options to identify the transcripts that were most likely to be the homologous match in the strawberry genome.

4.4. Results

An assessment of contiguity in each assembly was performed to help examine how well the transcriptomes were assembled. The N80 (the contig length at which 80 percent of the bases in the assembly are contained in contigs larger than this value), N50 (the contig length at which fifty percent of the bases in the assembly are contained in contigs larger than this value), and N20 (the contig length at which 20 percent of the bases in the assembly are contained in contigs larger than this value) statistics were recorded for each assembly.

Assembly of the 'Latham' data alone resulted in 87,946 transcripts longer than 200 nt with an N80 of 342, an N50 of 974 and N20 of 2,050. The total size of the 'Latham' only transcriptome was 54.8 MB. Assembly of the 'Titan' data alone resulted in 71,512 transcripts greater than 200nt in length with an N80 of 345, an N50 of 1,002 nt and N20 of 2246. The total size of the 'Titan' only transcriptome was 45.4 MB in length. Assembly of the entire data set together resulted in 80,290 transcript sequences longer than 200nt in length with an N80 of 555nt, an N50 transcript size of 1459 and an N20 of 3548. The total size of the combined assembly was 78.6 MB in length.

To further examine how well a given transcript was assembled in the combined assembly, the transcripts were aligned to the ab initio gene models from the *F. vesca* genome to identify the most likely homologous gene in strawberry. The predicted gene lengths based on the *F. vesca* ab initio gene models were then plotted against the length of the transcript assembled by Trinity in the combined assembly (Figure 4.1). Many genes were assembled near the predicted gene length,

but most transcripts were shorter than the predicted gene length. A number of assembled transcripts also exceeded the predicted ab initio gene length and these sequences may represent chimeric transcripts, but may also be an indication that the strawberry ab initio gene model either does not accurately represent the transcript from raspberry or was not well annotated in the *F. vesca* genome.

The combined assembly contained 64,277 transcripts greater than 300nt with blastx hits to the NCBI nr database with an e-value less than 1.0×10^{-3}. Of these, 20,761 were to bacteria, 43,461 were eukaryotic organisms, and the remaining 55 blast hits were archea, viruses or unclassified organisms. Nearly all of the sequences with top hits to eukaryotic organisms were to vascular plants, with the exception of over 1000 sequences with top hits belonging to the oomycetes *Phytophthora* or *Albugo* (Table 4.1). Given the high number of transcripts with blast hits to bacteria or oomycetes, an analysis of the number of sequencing reads contributing to each data set was conducted.

Most of the transcripts (16,141) with BLAST hits to bacteria belonged to the genus *Burkholderia.* These sequences were used to create a Bowtie index for alignment and the sequencing reads were mapped to the transcripts and recorded (Table 4.2) in order to estimate the fraction of the data set that was derived from bacterial nucleic acids. Similarly, the 512 sequences with top BLASTx hits to *Phytophthora* species were used to created a bowtie index and each data set was aligned independently to estimate the number of reads arising from *Phytophthora* (Table 4.3).

Differences in sequences and expression levels were apparent between the individual assemblies. First, differing numbers and expression levels of NPR1 transcripts were detected between ‘Latham’ and ‘Titan’ (Table 4.4). Furthermore, annotations of ‘Latham’ revealed 523

transcripts of the LRR type compared to only 314 in 'Titan'. The 50 most highly expressed LRR genes for 'Latham' and 'Titan' are shown in Tables 4.5 and 4.6. Diverse sets of WRKY transcription factors were detected in both 'Latham' and 'Titan' assemblies and a summary of WRKY transciption factors with GO terms associated with defense is shown in Table 4.7. Twelve PR-10 family genes (mal d) were assembled in 'Titan' and 11 were assembled in 'Latham'(Table 4.8).

Table 4.1 Summary of BLAST hits for the top 15 genera represented by the combined transcriptome with descriptive statistics for e-values, including quartiles 1-3 (Q1-3). Vascular plant genera are designated with *, oomycetes with **, and bacteria with ***. These 15 genera represent approximately 85% of the top blast hits or 54472 /64,277

	Number of Hits	e-value min	e-value mean	e-value max	e-value Q1	e-value Q2	e-value Q3
Vitis*	16429	0.00E+00	6.19E-06	9.51E-04	3.06E-130	3.38E-62	1.33E-29
Burkholderia***	16141	0.00E+00	3.97E-06	9.51E-04	3.91E-101	9.33E-52	2.12E-32
Populus*	8920	0.00E+00	5.52E-06	9.66E-04	2.61E-132	4.39E-63	9.51E-30
Ricinus*	6253	0.00E+00	6.73E-06	9.44E-04	1.63E-140	1.31E-66	3.04E-31
Arabidopsis*	1793	0.00E+00	7.32E-06	9.24E-04	1.24E-88	2.13E-40	5.88E-18
Glycine*	1322	0.00E+00	4.70E-06	9.42E-04	5.97E-87	1.96E-50	1.04E-27
Albugo**	552	0.00E+00	3.53E-06	9.35E-04	2.34E-80	1.17E-54	2.38E-39
Ralstonia***	508	0.00E+00	5.19E-06	6.88E-04	4.89E-59	9.75E-41	7.27E-25
Phytophthora**	512	0.00E+00	7.92E-06	9.57E-04	6.35E-71	4.41E-46	2.20E-25
Malus*	394	0.00E+00	1.05E-05	8.17E-04	4.25E-123	5.49E-77	7.88E-41
Medicago*	384	0.00E+00	6.16E-06	8.92E-04	1.72E-88	2.94E-49	8.23E-24
Pseudomonas***	378	0.00E+00	7.06E-06	7.50E-04	5.00E-56	9.49E-36	2.02E-20
Cupriavidus***	335	0.00E+00	8.06E-06	9.25E-04	5.73E-58	5.82E-38	8.03E-24
Oryza*	334	0.00E+00	1.71E-05	9.01E-04	4.48E-46	1.07E-23	7.73E-12
Prunus*	223	0.00E+00	1.33E-05	7.22E-04	9.14E-145	1.96E-57	1.27E-18

Table 4.2. Total reads obtained for each data set and the number and percent of reads mapping to Burkholderia sequences.

Data Set	Total_Reads	Aligned to Burkholderia	Percent
Resistant Control	15213629	161500	1%
Susceptible Control	18363149	390451	2%
Resistant Treatment	18923742	438053	2%
Susceptble Treatment	18145135	548116	3%

Table 4.3. Total reads obtained for each data set and the number and percent of reads mapping to *Phytophthora* sequences.

Lane	Total_Reads	Aligned to Phytophthora	Percent
Resistant Control	15213629	3132	0.02%
Susceptible Control	18363149	284	0.00%
Resistant Treatment	18923742	48648	0.20%
Susceptible Treatment	18145135	20297	0.10%

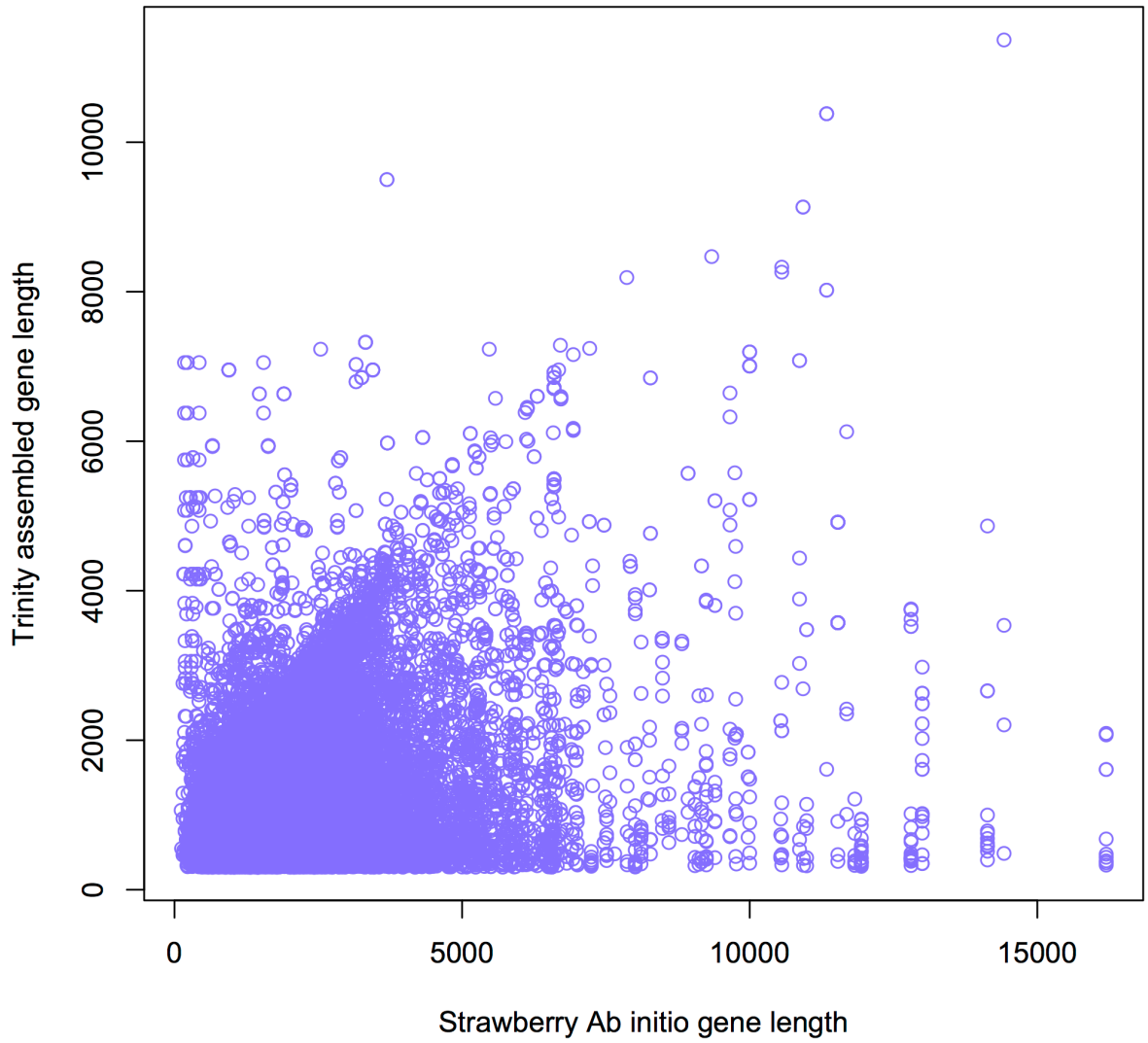

Figure 4.1. The best matches were found between the nucleotide sequences of the strawberry ab initio gene models and the combined Trinity assembly. The sequence length predicted by the ab initio gene models are shown in comparison to the actual length of the corresponding transcript from Trinity. The plot shows that for many genes the transcript length assembled by Trinity closely matched the predicted ab initio gene length. However, most transcripts were significantly shorter than the predicted gene length. A small number of transcripts were significantly longer than the predicted gene length, which may either represent divergence from the strawberry genome, invalid chimeric transcripts, or incomplete annotation of the *F. vesca* genome.

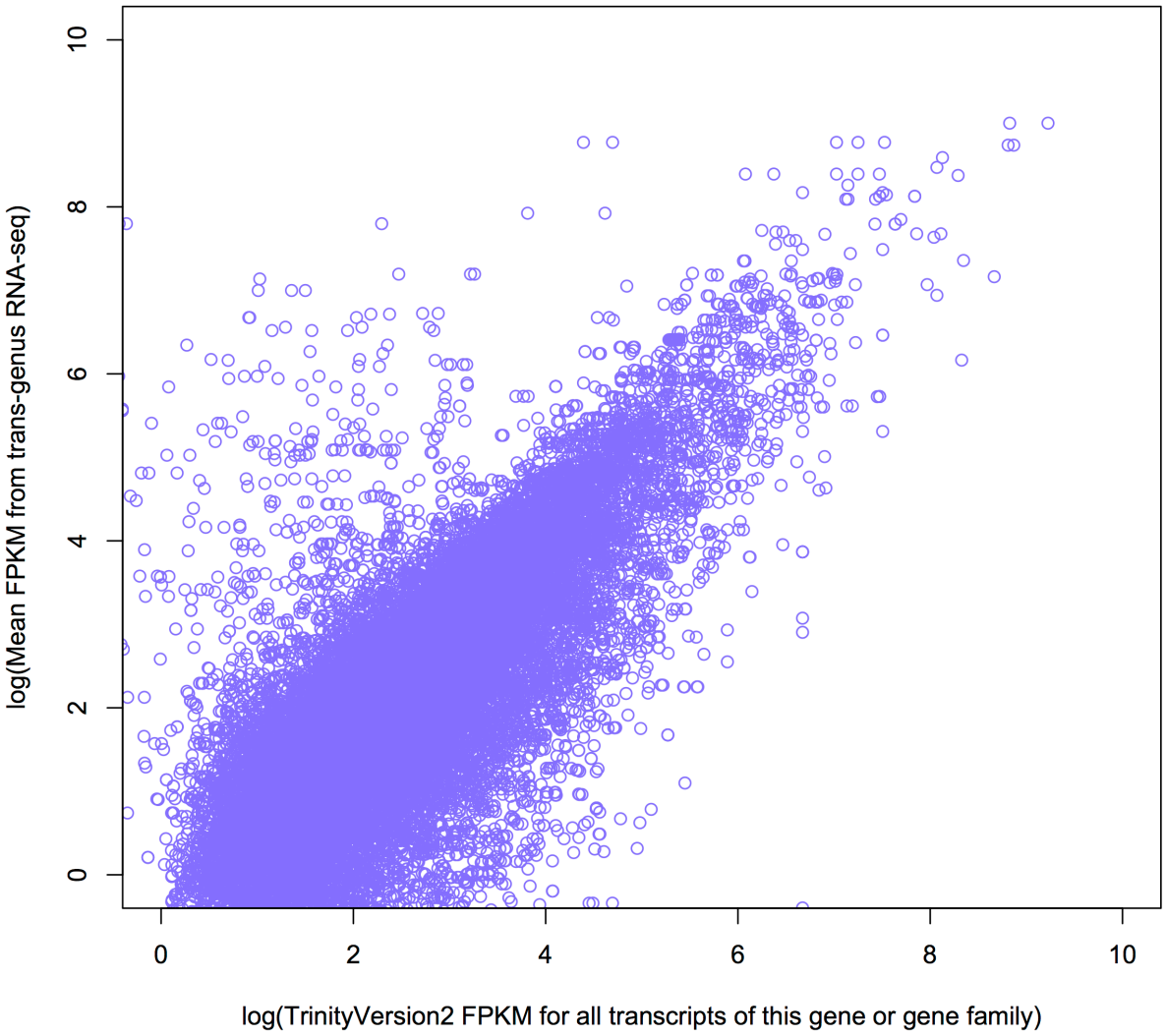

Figure 4.2. The best matches were found between the nucleotide sequences of the strawberry ab initio gene models and the combined Trinity assembly. The sequence FPKM predicted by the alignment-first method from chapter 3 is shown in comparison to the FPKM of the corresponding transcript from Trinity. Pearson‘s correlation between the two methods = 0.73

Table 4.4. Transcripts annotated as NPR1 or NPR1-like from both the 'Titan' or 'Latham' assemblies and corresponding FPKM.

TITAN				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp9708_c0_seq1	npr1 nim1-like regulatory protein	9.9	708	-
comp9400_c0_seq1	npr1 nim1-like regulatory protein	8.9	656	-
comp9400_c0_seq2	npr1-like protein	8.9	206	-
comp20396_c0_seq1	bop npr1 nim1-like regulatory protein	7.4	700	P:polarity specification of adaxial/abaxial axis; P:proximal/distal pattern formation; P:leaf morphogenesis; P:floral meristem determinacy; P:bract formation; P:floral organ abscission; F:protein binding; P:flower morphogenesis; P:nectary development; P:positive regulation of transcription, DNA-dependent; C:cytoplasm; C:nucleus
comp20595_c0_seq1	bop npr1 nim1-like regulatory protein	8.9	866	P:polarity specification of adaxial/abaxial axis; P:proximal/distal pattern formation; P:leaf morphogenesis; P:floral meristem determinacy; P:bract formation; P:floral organ abscission; F:protein binding; P:flower morphogenesis; P:nectary development; P:positive regulation of transcription, DNA-dependent; C:cytoplasm; C:nucleus
comp11013_c0_seq1	npr1-like protein	13.6	2120	-
comp29858_c0_seq1	npr1-like protein	3.3	222	-
comp18903_c0_seq1	npr1 nim1-like regulatory protein	5.6	1000	-
comp18903_c0_seq2	npr1 nim1-like regulatory protein	4.6	822	-
LATHAM				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp26245_c0_seq1	npr1 nim1-like regulatory protein	10.2	285	-
comp7299_c0_seq1	npr1 nim1-like regulatory protein	3.8	1394	-
comp7299_c0_seq2	npr1 nim1-like regulatory protein	4	1170	-
comp7299_c0_seq3	npr1 nim1-like regulatory protein	5.2	931	-
comp7299_c0_seq4	npr1 nim1-like regulatory protein	6.1	707	-
comp7299_c0_seq5	npr1 nim1-like regulatory protein	20.3	659	-
comp7299_c0_seq6	npr1 nim1-like regulatory protein	17.4	487	-

comp7299_c0_seq7	npr1 nim1-like regulatory protein	12.8	418	-
comp7299_c0_seq9	npr1-like protein	7.1	334	-
comp7299_c0_seq11	npr1-like protein	3.8	292	-
comp7299_c0_seq13	npr1 nim1-like regulatory protein	14.2	259	-
comp6995_c0_seq1	npr1-like protein	29.1	2769	-
comp10742_c0_seq1	bop npr1 nim1-like regulatory protein	11.6	1746	P:polarity specification of adaxial/abaxial axis; P:proximal/distal pattern formation; P:leaf morphogenesis; P:floral meristem determinacy; P:bract formation; P:floral organ abscission; F:protein binding; P:flower morphogenesis; P:nectary development; P:positive regulation of transcription, DNA-dependent; C:cytoplasm; C:nucleus
comp10742_c0_seq2	bop npr1 nim1-like regulatory protein	11.4	893	P:polarity specification of adaxial/abaxial axis; P:proximal/distal pattern formation; P:leaf morphogenesis; P:floral meristem determinacy; P:bract formation; P:floral organ abscission; F:protein binding; P:flower morphogenesis; P:nectary development; P:positive regulation of transcription, DNA-dependent; C:cytoplasm; C:nucleus

Table 4.5. The protein annotation for the 50 most highly expressed LRR type sequences in 'Titan'.

Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp2247_c0_seq1	cc-nbs-lrr resistance protein	53.6	1259	P:cellular process
comp128_c0_seq5	tir-nbs-lrr resistance protein	45.6	246	P:response to stimulus
comp4510_c0_seq1	f-box lrr-repeat protein 5	35.7	1447	P:cellular response to auxin stimulus; P:heat acclimation; C:SCF ubiquitin ligase complex; F:ubiquitin-protein ligase activity; P:protein ubiquitination; F:protein binding; P:positive regulation of cell division; P:SCF-dependent proteasomal ubiquitin-dependent protein catabolic process; P:lateral root primordium development; C:nucleus
comp4317_c0_seq1	cc-nbs-lrr resistance protein	34.4	727	F:hydrolase activity; F:nucleotide binding
comp3362_c0_seq1	tir-nbs-lrr resistance protein	29.9	1717	F:transmembrane receptor activity; P:defense response; P:apoptosis; F:ATP binding; P:signal transduction; P:innate immune response; C:intrinsic to membrane; F:nucleotide binding;

				F:nucleoside-triphosphatase activity
comp4608_c0_seq1	lrr receptor-linked protein	25.3	1574	P:circadian rhythm; F:transmembrane receptor protein tyrosine kinase activity; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway
comp6119_c0_seq1	lrr receptor-like serine threonine-protein kinase	22.2	2168	P:protein amino acid phosphorylation; P:oxidation reduction; F:protein serine/threonine kinase activity; C:integral to membrane; F:2-alkenal reductase activity; F:receptor activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; C:cytoplasmic membrane-bounded vesicle; F:ATP binding; C:plasma membrane
comp6315_c0_seq1	f-box lrr-repeat protein	21.1	1353	F:molecular_function; P:biological_process; C:cellular_component
comp7542_c0_seq1	nbs-lrr resistance protein	20.5	1730	P:defense response; C:apoplast; F:nucleotide binding
comp7551_c0_seq1	f-box lrr-repeat	16.6	1237	C:GPI-anchor transamidase complex
comp4317_c0_seq4	tir-nbs-lrr type disease resistance protein	15	463	P:defense response; P:cellular process
comp14722_c0_seq1	brassinosteroid lrr receptor kinase	12.9	945	F:receptor activity; F:phosphoprotein phosphatase activity; C:cytoplasmic membrane-bounded vesicle; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; C:mitochondrion
comp13542_c0_seq1	f-box lrr-repeat protein	12.7	1907	F:molecular_function; P:biological_process; C:cellular_component; F:heat shock protein binding
comp9354_c0_seq5	cc-nbs-lrr class disease resistance protein	11.3	350	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity; C:cellular_component
comp6595_c0_seq3	tir-nbs-lrr class disease resistance protein	11.1	201	P:primary metabolic process; C:mitochondrion; C:nucleolus; P:defense response to bacterium; P:cellular macromolecule metabolic process; C:plasmodesma; P:growth; C:cytosolic small ribosomal subunit; P:ribosome biogenesis; P:gene expression; C:chloroplast; C:plasma membrane
comp14455_c0_seq1	lrr receptor protein	10.5	2252	F:receptor activity; C:plant-type cell wall; C:plasmodesma; C:plasma

				membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:oxidation reduction; F:2-alkenal reductase activity
comp17480_c0_seq1	tir-nbs-lrr resistance protein	10.4	499	P:defense response; P:cellular process
comp18120_c0_seq1	lrr receptor-like serine threonine-protein kinase rfk1	10.1	273	F:receptor signaling protein serine/threonine kinase activity; P:protein amino acid phosphorylation; F:ATP binding; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:oxidation reduction; F:2-alkenal reductase activity
comp4967_c0_seq1	lrr receptor-linked protein	10	492	F:kinase activity
comp18700_c0_seq1	cc-nbs-lrr resistance protein	9.9	549	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp18640_c0_seq1	f-box lrr-repeat protein 14	9.8	671	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process
comp19569_c0_seq1	cc-nbs-lrr resistance protein	9.7	598	P:defense response; P:apoptosis; F:ATP binding
comp19477_c0_seq1	lrr receptor-linked protein	9.5	1208	P:circadian rhythm; F:transmembrane receptor protein tyrosine kinase activity; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway
comp16967_c0_seq1	f-box lrr-repeat protein 14	9.4	511	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process
comp26538_c0_seq1	cc-nbs-lrr resistance protein	9	307	P:response to stimulus; P:cellular process
comp8973_c0_seq1	nbs-lrr resistance protein	9	951	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp4317_c0_seq3	tir-nbs-lrr class disease resistance protein	9	512	P:defense response; P:cellular process; F:nucleotide binding
comp6950_c0_seq1	f-box lrr-repeat protein	8.9	499	F:molecular_function; P:biological_process; C:cellular_component
comp6495_c0_seq1	lrr receptor-like serine threonine-protein kinase rkf3	8.9	311	C:plasma membrane; F:protein binding; F:receptor signaling protein serine/threonine kinase activity; P:protein amino acid phosphorylation; F:ATP binding
comp13353_c0_seq1	tir-nbs-lrr resistance protein	8.9	257	P:defense response; P:cellular process; F:nucleotide binding
comp16443_c0_seq1	lrr receptor protein	8.6	520	F:receptor activity; C:plant-type cell wall; C:plasmodesma; C:plasma

				membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:oxidation reduction; F:2-alkenal reductase activity
comp15940_c0_seq1	lrr receptor protein kinase	8.6	481	F:receptor activity; C:endomembrane system; C:cytoplasmic membrane-bounded vesicle; F:protein tyrosine kinase activity; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity
comp11414_c0_seq1	cc-nbs-lrr resistance protein	8.5	286	P:apoptosis; P:defense response; F:ATP binding; F:hydrolase activity
comp5881_c0_seq2	lrr receptor protein	8.4	1805	F:receptor activity; C:plant-type cell wall; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:oxidation reduction; F:2-alkenal reductase activity
comp20692_c0_seq1	nbs-lrr resistance protein	8	653	P:apoptosis; F:ATP binding
comp15272_c0_seq1	nbs-lrr disease resistance protein	7.9	666	P:apoptosis; F:ATP binding; P:defense response
comp20913_c0_seq1	nbs-lrr resistance protein	7.9	357	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp23809_c0_seq1	tir-nbs-lrr resistance protein	7.9	577	P:defense response; P:cellular process; F:nucleotide binding
comp7715_c0_seq2	cc-nbs-lrr resistance protein	7.8	1148	F:nucleotide binding
comp21000_c0_seq1	f-box lrr-repeat protein	7.8	464	F:molecular_function; P:biological_process; C:cellular_component; C:plastid
comp24790_c0_seq1	nbs-lrr resistance protein	7.8	355	F:nucleoside-triphosphatase activity; P:defense response; P:apoptosis; F:ATP binding
comp5881_c0_seq1	lrr receptor protein	7.7	2219	F:receptor activity; C:plant-type cell wall; C:plasmodesma; C:plasma membrane; F:protein kinase activity; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; P:oxidation reduction; F:2-alkenal reductase activity
comp28589_c0_seq1	cc-nbs-lrr resistance protein	7.6	233	P:cellular process; P:response to stimulus
comp14234_c0_seq1	tir-nbs-lrr resistance protein	7.6	529	P:cellular process; P:response to stimulus; F:binding
comp31210_c0_seq1	cc-nbs-lrr resistance protein	7.5	456	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp8589_c0_seq1	tir-nbs-lrr resistance protein	7.4	1472	F:transmembrane receptor activity; P:defense response; P:apoptosis;

comp24315_c0_seq1	f-box lrr-repeat protein	7.2	465	F:ATP binding; P:signal transduction; P:innate immune response; C:intrinsic to membrane F:molecular_function; P:biological_process; C:cellular_component; C:endomembrane system
comp23273_c0_seq1	lrr receptor-linked protein	7.2	282	P:protein amino acid phosphorylation; F:protein serine/threonine kinase activity; C:mitochondrion; C:integral to membrane; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:circadian rhythm; C:cytoplasmic membrane-bounded vesicle; F:ATP binding; F:transmembrane receptor protein tyrosine kinase activity; C:plasma membrane
comp21623_c0_seq2	cc-nbs-lrr resistance protein	7.1	267	P:defense response; P:apoptosis; F:ATP binding
comp32842_c0_seq1	nbs-lrr resistance protein	7	318	P:defense response; P:apoptosis; F:ATP binding

Table 4.6. The protein annotation for the 50 most highly expressed LRR type sequences in 'Latham'.

Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp137_c0_seq8	tir-nbs-lrr resistance protein	301	276	F:transmembrane receptor activity; P:defense response; P:apoptosis; F:ATP binding; P:signal transduction; P:innate immune response; C:intrinsic to membrane
comp137_c0_seq10	tir-nbs-lrr resistance protein	201.9	254	P:defense response; P:cellular process
comp4954_c0_seq1	nbs-lrr resistance protein	38.1	2854	P:defense response; C:apoplast; F:nucleotide binding
comp1598_c0_seq1	cc-nbs-lrr resistance protein	34.2	796	P:cellular process
comp5413_c0_seq1	lrr receptor-linked protein	30.1	1491	P:circadian rhythm; F:transmembrane receptor protein tyrosine kinase activity; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway
comp1598_c0_seq2	cc-nbs-lrr resistance protein	26.9	737	P:cellular process
comp7188_c0_seq1	lrr receptor-like serine threonine-protein kinase	24.9	2666	P:protein amino acid phosphorylation; P:oxidation reduction; F:protein serine/threonine kinase activity; C:integral to membrane; F:2-alkenal reductase activity; F:receptor activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; C:cytoplasmic membrane-bounded vesicle; F:ATP binding; C:plasma membrane
comp5932_c0_seq2	f-box lrr-repeat protein 5	24.3	1423	P:cellular response to auxin stimulus; P:heat acclimation; C:SCF ubiquitin ligase complex; F:ubiquitin-protein ligase activity; P:protein ubiquitination; F:protein binding; P:positive regulation of cell division; P:SCF-dependent proteasomal ubiquitin-dependent protein catabolic process; P:lateral root primordium development; C:nucleus
comp8692_c0_seq1	nls-tir-nbs-lrr resistance protein	23.9	1318	P:defense response; P:cellular process
comp5932_c0_seq1	f-box lrr-repeat protein 5	23.2	1458	P:cellular response to auxin stimulus; P:heat acclimation; C:SCF ubiquitin ligase complex; F:ubiquitin-protein ligase activity; P:protein ubiquitination; F:protein binding; P:positive regulation of cell division; P:SCF-dependent proteasomal ubiquitin-dependent protein catabolic process; P:lateral root primordium development; C:nucleus
comp8666_c0_seq1	lrr receptor protein	21.4	2471	F:receptor activity; C:plant-type cell wall; C:plasmodesma; C:plasma membrane; F:protein kinase activity; P:protein amino acid phosphorylation; F:ATP binding;

				C:integral to membrane; P:oxidation reduction; F:2-alkenal reductase activity
comp10190_c0_seq1	tir-nbs-lrr resistance protein	20.2	1128	P:response to stimulus
comp8734_c0_seq1	nbs-lrr resistance protein	18.5	3265	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp12727_c0_seq1	nbs-lrr resistance protein	17.7	1598	P:apoptosis; F:ATP binding; P:defense response; F:binding
comp12972_c0_seq1	cc-nbs-lrr resistance protein	16.9	278	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity; F:phosphoprotein phosphatase activity
comp8785_c0_seq2	f-box lrr-repeat protein	16.8	398	F:molecular_function; P:biological_process; C:cellular_component
comp5789_c0_seq1	cc-nbs-lrr resistance protein	16.6	2392	F:nucleotide binding
comp12903_c0_seq1	cc-nbs-lrr resistance protein	16.3	756	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp13955_c0_seq1	nbs-lrr resistance protein	16.3	1431	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp15033_c0_seq1	nbs-lrr resistance protein	15.2	994	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp12165_c0_seq4	lrr receptor-like serine threonine-protein kinase rkf3	15	471	C:plasma membrane; F:protein binding; F:receptor signaling protein serine/threonine kinase activity; P:protein amino acid phosphorylation; F:ATP binding
comp5789_c0_seq3	cc-nbs-lrr resistance protein	14.7	1948	F:nucleotide binding
comp14262_c0_seq1	cc-nbs-lrr resistance protein	14.2	799	P:defense response; P:apoptosis; F:ATP binding
comp15931_c0_seq1	f-box lrr-repeat protein	13.9	1003	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process; F:molecular_function; P:biological_process; C:cellular_component; F:lyase activity
comp5031_c0_seq2	f-box lrr-repeat	13.7	1070	C:cellular_component
comp13346_c0_seq1	lrr receptor-like serine threonine-protein kinase	13.4	2273	F:receptor activity; P:protein amino acid phosphorylation; F:ATP binding; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway
comp5893_c0_seq8	tir-nbs-lrr resistance protein	13.4	282	P:defense response; F:hydrolase activity; P:cellular process; F:nucleotide binding
comp5789_c0_seq9	nbs-lrr resistance protein	13.2	715	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp3046_c0_seq6	tir-nbs-lrr resistance protein	13.2	1062	P:defense response; F:transmembrane receptor activity; P:apoptosis; F:nucleotide binding; F:nucleoside-triphosphatase activity; F:ATP binding; C:intrinsic to membrane; P:signal transduction; P:innate immune response
comp16760_c0_seq1	f-box lrr-repeat protein 14	13	1409	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process

comp23844_c0_seq1	cc-nbs-lrr resistance protein	12.6	280	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity; F:phosphoprotein phosphatase activity; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp16082_c0_seq1	lrr receptor-linked protein	12.4	1147	P:circadian rhythm; F:transmembrane receptor protein tyrosine kinase activity; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway
comp8911_c0_seq1	f-box lrr-repeat protein	12.3	1832	F:molecular_function; P:biological_process; C:cellular_component; F:heat shock protein binding
comp4552_c0_seq1	cc-nbs-lrr resistance protein	12.2	2692	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity; F:phosphoprotein phosphatase activity
comp8911_c0_seq2	f-box lrr-repeat protein	12	1767	F:molecular_function; P:biological_process; C:cellular_component; F:heat shock protein binding
comp23114_c0_seq3	tir-nbs-lrr class disease resistance protein	11.5	253	P:defense response; P:cellular process; F:nucleotide binding
comp14758_c0_seq1	cc-nbs-lrr resistance protein	11.1	561	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity; F:molecular_function; P:response to other organism
comp7838_c0_seq4	lrr receptor-like serine threonine-protein kinase erl1	11.1	339	P:stomatal complex morphogenesis; P:embryo sac development; P:ovule development; C:cytoplasmic membrane-bounded vesicle; F:kinase activity
comp9194_c0_seq1	f-box lrr-repeat protein 12	11	1845	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process
comp14262_c0_seq2	nbs-lrr resistance protein	11	456	P:defense response; P:apoptosis; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity
comp18929_c0_seq1	nbs-lrr disease resistance protein	10.9	904	P:apoptosis; F:ATP binding; P:defense response
comp12367_c0_seq8	cc-nbs-lrr resistance protein	10.8	218	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
comp24222_c0_seq1	tir-nbs-lrr resistance protein	10.8	372	P:defense response; P:cellular process; F:nucleotide binding
comp13158_c0_seq1	nbs-lrr resistance protein	10.6	652	P:plant-type hypersensitive response; C:plasma membrane; F:protein binding
comp9194_c0_seq2	f-box lrr-repeat protein 12	10.4	1736	F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process
comp19079_c0_seq1	f-box lrr-repeat protein	10.3	451	F:molecular_function; P:biological_process; C:cellular_component
comp13918_c0_seq1	nbs-lrr resistance protein	10.3	852	F:nucleotide binding
comp22371_c0_seq1	tir-nbs-lrr resistance protein	10.3	329	P:innate immune response; P:apoptosis; C:intrinsic to membrane; F:ATP binding;

comp778_c0_seq1	lrr receptor protein	10.2	3394	P:signal transduction; F:transmembrane receptor activity F:receptor activity; C:plant-type cell wall; C:endomembrane system; C:plasmodesma; C:plasma membrane; P:protein amino acid phosphorylation; F:ATP binding; C:integral to membrane; F:protein serine/threonine kinase activity; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:oxidation reduction; F:2-alkenal reductase activity

Table 4.7. WRKY transcription factors associated with defense in 'Titan' and 'Latham' assemblies.

TITAN				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp664_c0_seq1	wrky transcription factor 17	211.1	1260	P:defense response to bacterium; F:transcription factor activity; F:sequence-specific DNA binding; P:response to chitin; C:nucleus; F:calmodulin binding; P:regulation of transcription, DNA-dependent
comp2140_c0_seq1	wrky transcription	26.7	1461	P:defense response to bacterium; F:sequence-specific DNA binding; P:response to chitin; F:transcription factor activity; P:positive regulation of transcription, DNA-dependent; C:nucleus
comp1229_c0_seq2	wrky transcription factor 11	22.7	445	P:defense response to bacterium; F:transcription factor activity; F:sequence-specific DNA binding; P:response to chitin; C:nucleus; F:calmodulin binding; P:regulation of transcription, DNA-dependent
comp7857_c0_seq1	wrky transcription	16.3	599	P:jasmonic acid mediated signaling pathway; P:defense response to bacterium; P:regulation of transcription, DNA-dependent; C:plastid; F:sequence-specific DNA binding; F:transcription factor activity; P:defense response to fungus
comp3810_c0_seq2	wrky transcription factor 22	15.2	1108	P:response to gibberellin stimulus; P:defense response to bacterium; P:nitric oxide mediated signal transduction; F:protein binding; P:negative regulation of transcription, DNA-dependent; C:nucleus
comp9524_c0_seq1	wrky transcription	13.5	1606	F:zinc ion binding; F:transcription factor activity; P:salicylic acid mediated signaling pathway; P:positive regulation of transcription, DNA-dependent
comp11962_c0_seq1	wrky transcription factor 72	11.4	998	P:transcription, DNA-dependent; P:defense response
comp21895_c0_seq1	wrky	8.5	262	F:transcription factor activity; F:sequence-

	transcription			specific DNA binding; P:defense response; P:regulation of transcription, DNA-dependent
comp21837_c0_seq1	wrky transcription	7.8	503	F:transcription factor activity; F:sequence-specific DNA binding; P:defense response; P:regulation of transcription, DNA-dependent
comp26898_c0_seq1	wrky transcription	6.7	487	F:transcription factor activity; F:sequence-specific DNA binding; P:defense response; P:regulation of transcription, DNA-dependent
LATHAM				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp684_c0_seq1	wrky transcription factor 17	279.5	1247	P:defense response to bacterium; F:transcription factor activity; F:sequence-specific DNA binding; P:response to chitin; C:nucleus; F:calmodulin binding; P:regulation of transcription, DNA-dependent
comp5988_c0_seq1	wrky transcription	45	773	P:jasmonic acid mediated signaling pathway; P:defense response to bacterium; P:regulation of transcription, DNA-dependent; C:plastid; F:sequence-specific DNA binding; F:transcription factor activity; P:defense response to fungus
comp8303_c0_seq1	wrky transcription	34.1	1495	P:defense response to bacterium; F:sequence-specific DNA binding; P:response to chitin; F:transcription factor activity; P:positive regulation of transcription, DNA-dependent; C:nucleus
comp1685_c0_seq2	wrky transcription	33.8	1309	P:defense response; P:response to other organism; F:binding; P:regulation of cellular process; P:transcription, DNA-dependent; P:response to organic substance
comp7859_c0_seq1	wrky transcription	32	566	P:defense response; P:response to other organism; F:binding; P:regulation of biological process
comp13532_c0_seq1	wrky transcription factor 72	20	989	P:defense response
comp14786_c0_seq1	wrky transcription	15.4	1751	F:zinc ion binding; F:transcription factor activity; P:salicylic acid mediated signaling pathway; P:positive regulation of transcription, DNA-dependent
comp17098_c0_seq1	wrky transcription factor 41	10.3	471	F:sequence-specific DNA binding; P:response to chitin; F:protein binding; F:transcription factor activity; P:leaf senescence; P:regulation of defense response; C:chloroplast; P:positive regulation of transcription, DNA-dependent; P:defense response to bacterium, incompatible interaction; C:nucleus

Table 4.8. PR-10 (mal d) genes of in ‘Titan’ and ‘Latham’ assemblies.

TITAN				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp13_c1_seq1	major allergen mal d	5982.3	593	P:response to biotic stimulus; P:defense response
comp15_c0_seq1	mal d	5596.2	996	P:response to biotic stimulus; P:defense response
comp23_c0_seq1	mal d	3521.9	872	P:response to biotic stimulus; P:defense response
comp49_c0_seq1	major allergen mal d	424.8	425	P:response to biotic stimulus; P:defense response
comp49_c0_seq2	major allergen mal d	356.4	422	P:response to biotic stimulus; P:defense response
comp49_c0_seq3	major allergen mal d	566.3	316	P:response to biotic stimulus; P:defense response
comp49_c0_seq4	mal d	1371.8	261	P:response to biotic stimulus; P:defense response
comp49_c0_seq5	mal d	387.2	244	P:response to biotic stimulus; P:defense response
comp971_c0_seq1	mal d 1-like	118.1	990	P:response to biotic stimulus; P:defense response
comp2454_c0_seq1	major allergen mal d	37.5	303	P:response to biotic stimulus; P:defense response
comp2454_c0_seq2	major allergen mal d	49.4	287	P:response to biotic stimulus; P:defense response
comp9781_c0_seq1	mal d 1-associated protein	9.8	488	C:nucleus
LATHAM				
Seq. Name	Seq. Description	FPKM	Seq. Length	GOs
comp10_c0_seq1	mal d	9936.7	1120	P:response to biotic stimulus; P:defense response
comp26_c0_seq1	mal d	4491.5	881	P:response to biotic stimulus; P:defense response
comp37_c0_seq1	major allergen mal d	4003.9	670	P:response to biotic stimulus; P:defense response
comp66_c0_seq3	major allergen mal d	2013.5	248	P:response to biotic stimulus; P:defense response
comp66_c0_seq2	major allergen mal	827.5	322	P:defense response; P:response to biotic stimulus; P:ripening

comp90_c0_seq1	d major allergen mal d	805.7	399	P:response to biotic stimulus; P:defense response
comp90_c0_seq2	major allergen mal d	732	308	P:response to biotic stimulus; P:defense response
comp66_c0_seq1	major allergen mal d	236.4	475	P:defense response; P:response to biotic stimulus; P:ripening
comp906_c0_seq1	mal d 1-like	209.3	1066	P:response to biotic stimulus; P:defense response
comp79_c0_seq2	major allergen mal d	84.6	206	P:response to biotic stimulus; P:defense response
comp12722_c0_seq1	mal d 1-associated protein	21.1	1082	C:nucleus

4.5. Discussion

Examination of de novo assemblies from 'Latham' and 'Titan' has revealed key differences in both the surveillance and potential defense responses to recognition of *P. rubi*. Increased surveillance in 'Latham' is apparent not only by the increased number of LRR transcripts observed, but also by the expression levels of the LRR genes. The increase in the total number of R-gene transcripts detected in Latham' (523) compared to 'Titan' (314) is likely to be due to increased heterozygosity rather than an actual increase in the number of R genes expressed. Although there may be an increased number of expressed R genes in 'Latham' or an increase in the number of alternatively spliced isoforms and further investigation is needed.

The two most highly expressed TIR-NBS-LRRs (TNLs) in 'Latham' are similar to genes recently implicated in defense against the hemibiotrophic *Diplocarpon rosae* in rose (Terefe-Ayana et al. 2011). These are the 'Latham' transcripts comp137_c0_seq8 and comp137_c0_seq10, which show similarity to muRDR1d from *Rosa multiflora and have FPKMs of 301 and 201.9 respectively*. By contrast the two most highly expressed R-genes in Titan are a

CC-NBS-LRR with an FPKM of 53.6 and a TIR-NBS-LRR, which is also similar to other muRDR1 genes in rose, but has an FPKM of only 45.6.

Downstream from pathogen recognition the de novo assembly detected a larger number of distinct WRKY transcription factors associated with defense GO terms in the 'Titan' assembly, but increased expression levels of fewer WRKY transcription factors associated with defense in the 'Latham' assembly. A close examination of these WRKY transcription factors is warranted to investigate how they may differ in function and to identify the homologs in the strawberry genome, which were differentially expressed in Chapter 3. Increased expression of WRKY transcription factors after inoculation is likely to be the result of PTI given that the activation of the PAMP triggered MAPK cascade typically results in activation of WRKY transcription factors (Ingle et al. 2006).

NPR1 is typically thought to act downstream of effector recognition in the activation of SAR and thus its increased detection in the 'Latham' treatment in Chapter 3 is probably an indication that an R-gene is involved in the resistance response. This is clearly demonstrated in the de novo assembly. In 'Latham' there were 14 assembled transcripts annotated as NPR1 with FPKMs ranging from 3.8 to 29.1, while in the 'Titan' assembly there were only 9 transcripts assembled with NPR1 annotations and FPKMs ranging from 3.3 to 13.6. Furthermore, the 'Latham' assembly had five transcripts that were assembled to lengths greater than 1000nt, while the 'Titan' assembly had only two transcripts with lengths greater than 1000nt. Table 4.4 illustrates these differences. Again, the increased diversity of transcripts in 'Latham' could be due to heterozygosity, but increased expression levels together with the increased contiguity in these transcripts is a strong indicator that there are either differing isoforms or other copies of NPR1 are activated in 'Latham.'

Overall, the assembly results are consistent with those from the alignment-first approach performed in Chapter 3. The correlation between the methods (Pearson's corr = 0.73, Figure 4.2) supports the alignment-first approach and strengthens the hypotheses proposed in Chapter 3. These data will set the stage for rapid annotation of genes involved in resistance to *P. rubi* upon completion of the red raspberry genome. Further investigation of these assemblies is warranted to investigate alternative splicing patterns in highly expressed genes and variation in coding regions between genotypes. Pairing these data with markers from QTL studies and the raspberry genome will aid in refined hypotheses about resistance to *P. rubi* and should provide useful markers for selection of resistant plants in breeding programs.

4.6. References

Arumuganathan, K. & Earle, E., 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9(3), pp.208–218.

Bednarek, P. et al., 2009. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, 323(5910), pp.101–106.

Crandall, P., 1977. Red raspberry root rot research [Phytophthora erithroseptica]. *Annual Report - Proceedings of the Oregon Horticultural Society*, 68, p.57.

Delcher, A.L. et al., 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, 30(11), pp.2478–2483.

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652.

Ingle, R., Carstens, M. & Denby, K., 2006. PAMP recognition and the plant-pathogen arms race. *Bioessays*, 28, pp.880–889.

Jones, J.D.G. & Dangl, J.L., 2006. The plant immune system. *Nature*, 444, pp.323–329.

Meyers, B.C. et al., 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*, 15(4), pp.809–834.

Pattison, J.A., Samuelian, S.K. & Weber, C.A., 2007. Inheritance of Phytophthora root rot resistance in red raspberry determined by generation means and molecular linkage analysis. *Theoretical And Applied Genetics*, 115(2), pp.225–236.

Pattison, J.A., Wilcox, W. & Weber, C.A., 2004. Assessing the resistance of red raspberry (Rubus idaeus L.) genotypes to Phytophthora fragariae var. rubi in hydroponic culture. *Hortscience*, 39(7), pp.1553–1556.

Samuelian, S.K. et al., 2008. Isolation and linkage mapping of NBS-LRR resistance gene analogs in red raspberry (Rubus idaeus L.) and classification among 270 Rosaceae NBS-LRR genes. *Tree Genetics & Genomes*, 4(4), pp.881–896.

Simpson, J.T. et al., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19(6), pp.1117–1123.

Terefe-Ayana, D. et al., 2011. Frontiers: Mining Disease-Resistance Genes in Roses: Functional and Molecular Characterization of the Rdr1 Locus. *Frontiers in Plant Science*, 2, pp.1–12.

Wilcox, W., 1989. Identity, virulence, and isolation frequency of seven Phytophthora spp. causing root rot of raspberry in New York. *Phytopathology*, 79(1), pp.93–101.

Wilcox, W. et al., 1993. Identity of a Phytophthora species attacking raspberry in Europe and North America. *Mycological Research*, 97, pp.817–831.

Wilcox, W., Pritts, M. & Kelly, M., 1999. Integrated control of Phytophthora root rot of red raspberry. *Plant Disease*, 83(12), pp.1149–1154.

Chapter 5

ASSESSING RESISTANCE TO PHYTOPHTHORA ROOT ROT IN RUBUS IDAEUS L. (ROSACEAE) BY PROFILING OF PRIMARY METABOLITES

5.1. Abstract

Mechanisms of resistance to Phytophthora root rot (PRR) (caused by *Phytophthora rubi*) in the red raspberry (*Rubus idaeus L.*) 'Latham' were investigated through primary metabolite profiling with gas chromatography/mass spectrometry (GC/MS). These data were interpreted together with the results of a genome-wide expression study. Changes in sucrose, glucose, citrate and malate were quantified in root samples from controls and at 2 days, 5 days, and 20 days post inoculation with *P. rubi*. Few differences were statistically significant between genotypes or between treatments and controls of a single genotype at specific times; however, temporal trends were consistent between genotypes, treatments, and controls. In all genotypes, treatments and controls, both sucrose and glucose levels declined over time and citrate and malate levels peaked at five days. When specific times were not considered there were statistically significant differences in citrate levels between the 'Latham' treatment and the 'Titan' treatment ($p = 0.0164$) and between the 'Latham' control and 'Titan' control ($p = 0.0008$), suggesting that 'Latham' may have a more active flux through the citric acid cycle (TCA). This finding is consistent with previous reports that citrate is rate limiting in the TCA and that an increased flux through the TCA can be a factor in disease resistance. Therefore increased citrate concentrations could be a factor in root vigor in 'Latham' and thus a component of quantitative disease

resistance. This observation agrees with recent work suggesting that root vigor may play a role in 'Latham's' resistance to PRR.

5.2. Introduction

It is well recognized that active defenses against pathogens are important in all parts of the plant, but in roots it is also commonly suggested that vigor plays a role in resistance. This may be partly due to biases that arise because root characteristics are difficult to observe. Even vigorous roots would succumb to virulent pathogens in the absence of some form of defense and therefore plant immune responses must also play a role in cases where vigor is thought to be involved.

In red raspberry (*Rubus idaeus* L.), *Phytophthora rubi* causes economically damaging Phytophthora Root Rot (PRR) (Wilcox et al. 1993; Wilcox 1989). Recent studies suggest that resistance to PRR in 'Latham' is quantitative, but that a dominant two gene inheritance model can explain most of the observed variation (Pattison et al. 2007). Later, it was recognized that one QTL for root vigor in 'Latham' scored in the absence of disease overlapped with one of the two QTL for resistance (Graham et al. 2011).

Graham et al. (2011) sequenced two BACs (one from each QTL) and annotated several genes within each sequence, but did not emphasize the annotations in the context of disease resistance pathways. Instead, it was noted that several genes within these BAC sequences may play a role in root development and thus major conclusions suggested that vigor was likely a key player in resistance. However, identifying overlapping QTL for resistance and vigor does not necessarily imply causal relationships at the level of the genes underlying those QTL (i.e. different genes under the QTL could contribute separately to the resistance or root vigor).

Furthermore, given the mapping resolution limits in raspberry imposed by available populations and marker density it is likely that polymorphisms influencing the QTL may be outside the sequenced BACs.

Another possible complication with interpretation of BAC sequence data is that a distant cis-acting regulatory sequence could underlie the QTL as the causal polymorphism and thus a gene-centric BAC interpretation could miss the gene influenced by this sequence. That said, the observations made by Graham et al. (2011) regarding vigor are intriguing and vigor may play a role in resistance. This fits with the long standing recognition that resistance can be metabolically expensive (Tian et al. 2003; Bolton 2009; Zangerl et al. 1997; Smedegaard-Petersen & Tolstrup 1985; Heil et al. 2000) at least under certain conditions (Heidel et al. 2004).

Resistance mechanisms involving cell wall modifications have been shown to be particularly metabolically costly (Niemann et al. 1992; Heidel et al. 2004). In previous chapters here, it was observed that, in addition to genes from known disease resistance pathways, several genes known to be involved in primary metabolism were up regulated in 'Latham' following exposure to *P. rubi*. It is possible that changes in primary metabolism between 'Latham' and the susceptible cultivar 'Titan' following exposure to *P. rubi* may coincide with differences in resistance to PRR. Gas chromatography/mass spectrometry (GC/MS) analysis of primary metabolites are investigated here and analyzed in the context of the genome-wide expression studies conducted in previous chapters.

5.3. Methods

5.3.1. Plant Material [see also Chapter 3]

A total of 40 clones of the resistant cultivar 'Latham' and 40 clones of the susceptible

cultivar ‘Titan’ were randomly assigned to eight separate hydroponic systems developed for screening raspberry for resistance to *P. rubi* (Pattison et al. 2004). Four of these systems were challenged with single zoospore derived isolates of *P. rubi* using methods described previously (Pattison et al. 2004) and four hydroponic systems remained un-inoculated to serve as controls.

Approximately one gram of root tissue was collected from two randomly selected clones of ‘Latham’ and ‘Titan’ from each of the eight hydroponic systems (4 controls and 4 treatments) at 0, 2, 5 and 20 days post inoculation. Each sample was immediately placed into a 23.876 micron thick aluminum foil packet and was flash frozen in liquid nitrogen and stored at -80 degrees Celsius until metabolite extraction was performed. At 20 days, roots in the Titan treatment were dead and degraded and thus this sample was excluded. This resulted in 8 independent biological replicates (2 separate clones from each of the 4 treatment or control hydroponic systems) for each genotype and time point for a total of 11 groups and 88 samples.

5.3.2. Gas chromatography mass spectrometry

Each sample was ground into fine powder in liquid nitrogen and was stored at -80 ºC until metabolite extraction was performed. Extraction of sugars and organic acids was performed according to previous methods (Lisec et al. 2006) with modifications including the adjustment of the ribitol internal standard to a concentration of 600 ppm and the omission of alkane size standards. One hundred milligrams of the ground root tissue was used in the extraction, and a vial containing 100 uL of the polar phase was dried under vacuum in a 2 mL microcentrifuge tube. Following drying, all samples were stored at -80 ºC.

Derivitization was carried out in a two stage process starting with the addition of 40 uL of methoxyaminhydrochloride (20mg/ul in Pydidin) and shaking at 37 °C and 900 rpm for 2 hours

followed by the addition of 60 uL of N-methyl-N-trimethylsilyl-trifluoroace-tamide (MSTFA) with an additional 30 minutes of shaking at 900 rpm at 37 ºC.

Following the derivitization, metabolites were analyzed on an Agilent 7890A GC/5975C MS (Agilent Technology, Palo Alto, CA, USA). A 1 uL sample was injected at 230 ºC in splitless mode with helium carrier gas flow at 1 mL/min and the flow rate was constant with electronic pressure control enabled. Chromatography was carried out with a 20 m by 0.18 mm x 0.18 μm DB-5MS capillary column with a 5 m Duraguard front end column.

The GC/MS program used was as follows: 1.) warm up to 70 ºC; 2.) hold at 70 ºC for 2.471 minutes; 3.) ramp temperature by 5 ºC per minute up to 330 ºC; 4.) hold at 330 ºC for 5 minutes; 5.) cool the instrument and equilibrate at 70 ºC for 5 minutes before the next injection. Collection of mass spectral data was at 5.6 scans per second with a 50–600 m/z scanning range. The temperature of the transfer line was set to 250 ºC and the ion source was set to 230 ºC.

Identification of sugars was performed using the Agilent Chem Station (Agilent Technology, Palo Alto, CA, USA) and the GC–MS Golm Metabolome Database (http://csbdb.mpimp- golm.mpg.de/csbdb/gmd/msri/gmd_msri.html). Autointegration of corresponding peaks was performed using the Agilent Chem Station and quantification was performed using standard curves generated with standards from Sigma–Aldrich (St. Louis, MO, USA). Statistical analysis was performed in R using ANOVA and Tukey's HSD.

5.4. Results and Discussion

Sucrose, glucose, citrate, and malate were quantified in root samples from controls and at 2 days, 5 days, and 20 days post inoculation with *P. rubi*. Metabolite levels were analyzed for differences between treatments and controls at specific times and were also analyzed between

genotypes. There were no statistically significant differences between treatments and controls at specific time points for either genotype; however, there were strong trends in metabolite changes over time and statistically significant differences between genotypes.

In both genotypes, for treatments and controls, both sucrose (figure 5.1) and glucose (figure 5.2) declined over time and citrate (figure 5.3) and malate (figure 5.4) peaked at five days. These trends in metabolite levels likely parallel normal growth of roots in this hydroponic system, where plants reached the growth limits imposed by the hydroponic system and cease to be significant sinks for sugars between 2 and 20 days during this experiment.

In addition to these temporal trends there was a statistically significant difference in citrate concentration between the 'Latham' control and the 'Titan' control (p adjusted = 0.0008) and between the 'Latham' treatment and 'Titan' treatment (p adjusted = 0.0164). These genotypic differences are consistent with changes in expression of key enzymes of the TCA, such as ATP citrate synthase (figure 5.5). ATP citrate synthase was among the most highly expressed genes in the resistant treatment and was significantly differentially expressed between the treatment and control with an FPKM of 510.06 in the 'Latham' control and an FPKM of 1007.01 in the 'Latham' treatment. The same gene had lower expression and was significantly down regulated in the 'Titan' treatment (FPKM=212.69) compared to the 'Titan' control (FPKM = 386.37).

Although few differences were observed in metabolite levels between the treatments and controls, the difference seen in citrate concentration between the resistant genotype 'Latham' and the susceptible genotype 'Titan' is consistent with gene expression changes observed in the RNAseq experiment and may be an indication of increased flux through the TCA. These observations are consistent with previous reports that increased flux through the TCA may play a

role in disease resistance (Bolton 2009). This finding agrees with a recent report suggesting that root vigor may play a significant role in resistance to PRR (Graham et al. 2011). Further work is needed to pair these findings with previous mapping studies in order to confirm results and identify markers linked to resistance.

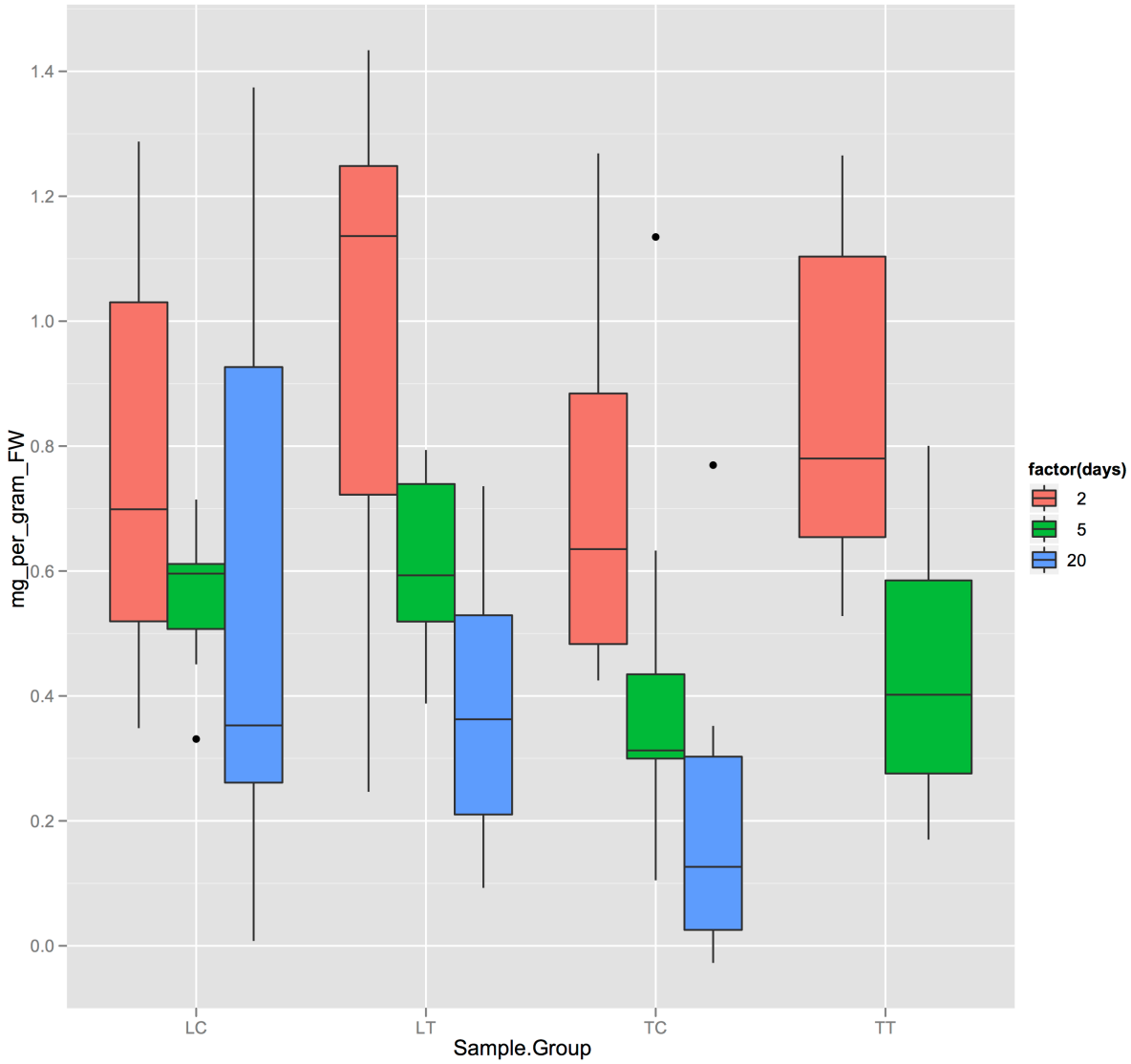

Figure 5.1. Sucrose concentration in eight biological replicates (the central horizontal line is the median, the box represents the interquartile range, vertical lines are the minimum and maximum values excluding additional plotted points which were considered outliers) for each time point (red at 2 days, green at five days, and blue at 20 days), treatment ('Latham Treatment' = LT 'Titan Treatment'=TT), control ('Latham

Control = LC ‘Titan Control=TC), and genotype showing a consistent trend toward a peak in malate concentration at five days.

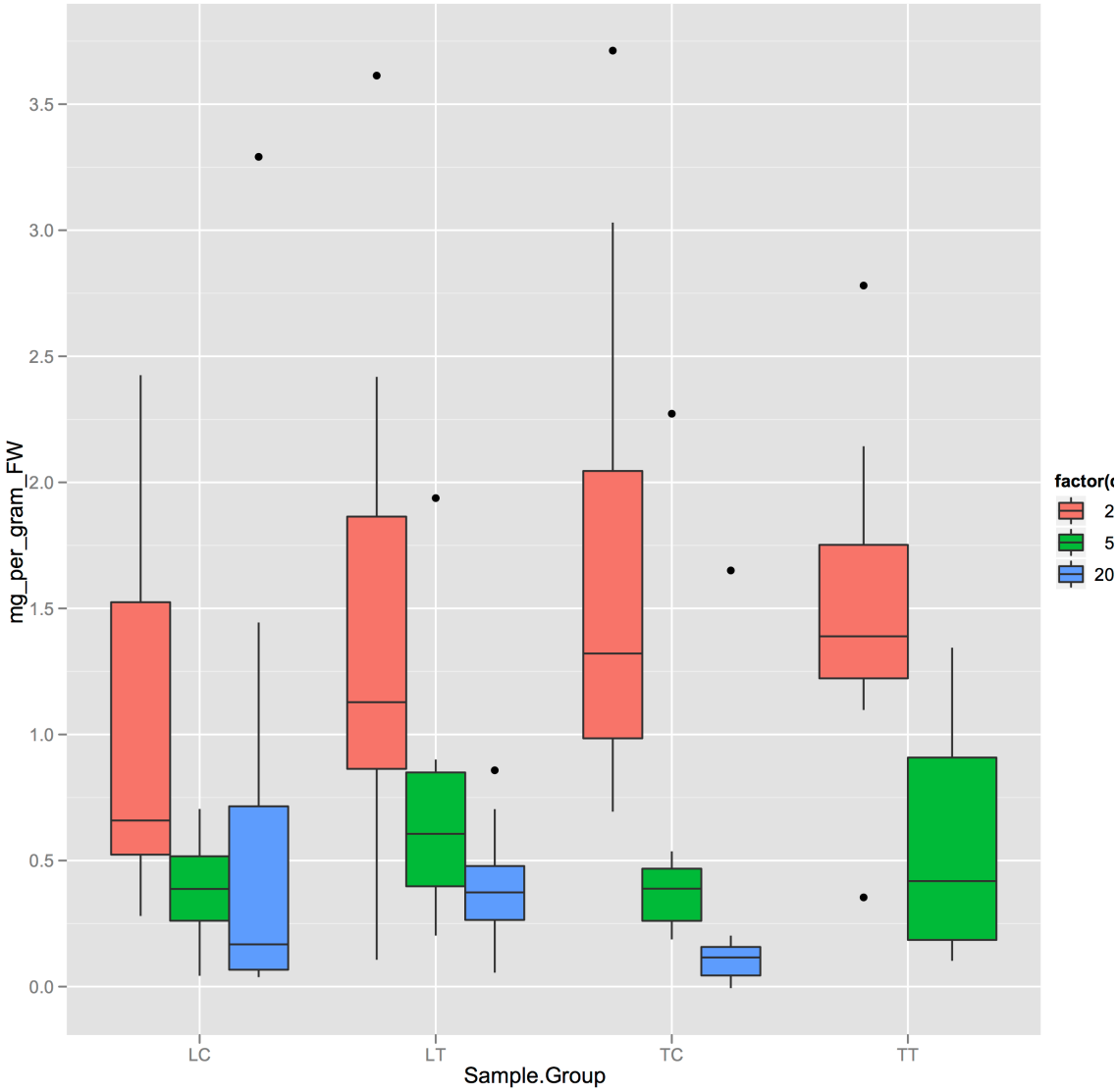

Figure 5.2. Glucose concentration in eight biological replicates (the central horizontal line is the median, the box represents the interquartile range, vertical lines are the minimum and maximum values excluding additional plotted points which were considered outliers) for each time point (red at 2 days, green at five days, and blue at 20 days), treatment (‘Latham Treatment’ = LT, ‘Titan Treatment’=TT), control (‘Latham Control = LC, ‘Titan Control=TC), and genotype showing a consistent trend toward a decline in glucose levels over time and a trend toward slightly increased levels of glucose in all ‘Latham’ Treatments compared to the ‘Latham’ controls.

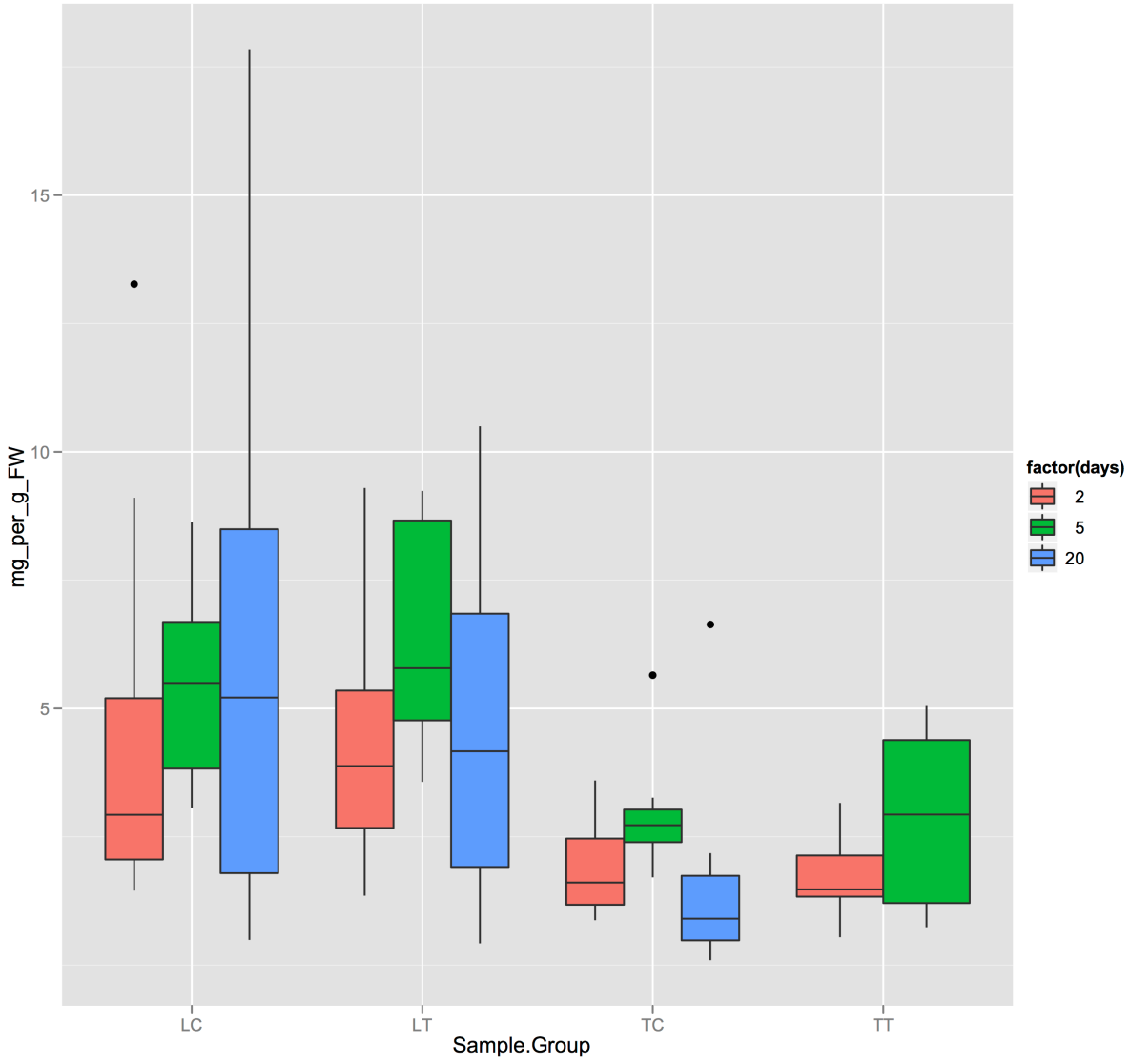

Figure 5.3. Citrate concentration in eight biological replicates (the central horizontal line is the median, the box represents the interquartile range, vertical lines are the minimum and maximum values excluding additional plotted points which were considered outliers) for each time point (red at 2 days green at five days, and blue at 20 days), treatment ('Latham Treatment' = LT 'Titan Treatment'=TT), and control ('Latham Control = LC 'Titan Control=TC), showing a consistent trend toward a peak in citrate concentration at five days. There are statistically significant differences between the 'Latham' Control and the 'Titan' Control (p adjusted = 0.0008021) and between the 'Latham' Treatment and 'Titan' Treatment (p adjusted = 0.0163706) when sampling date is not considered.

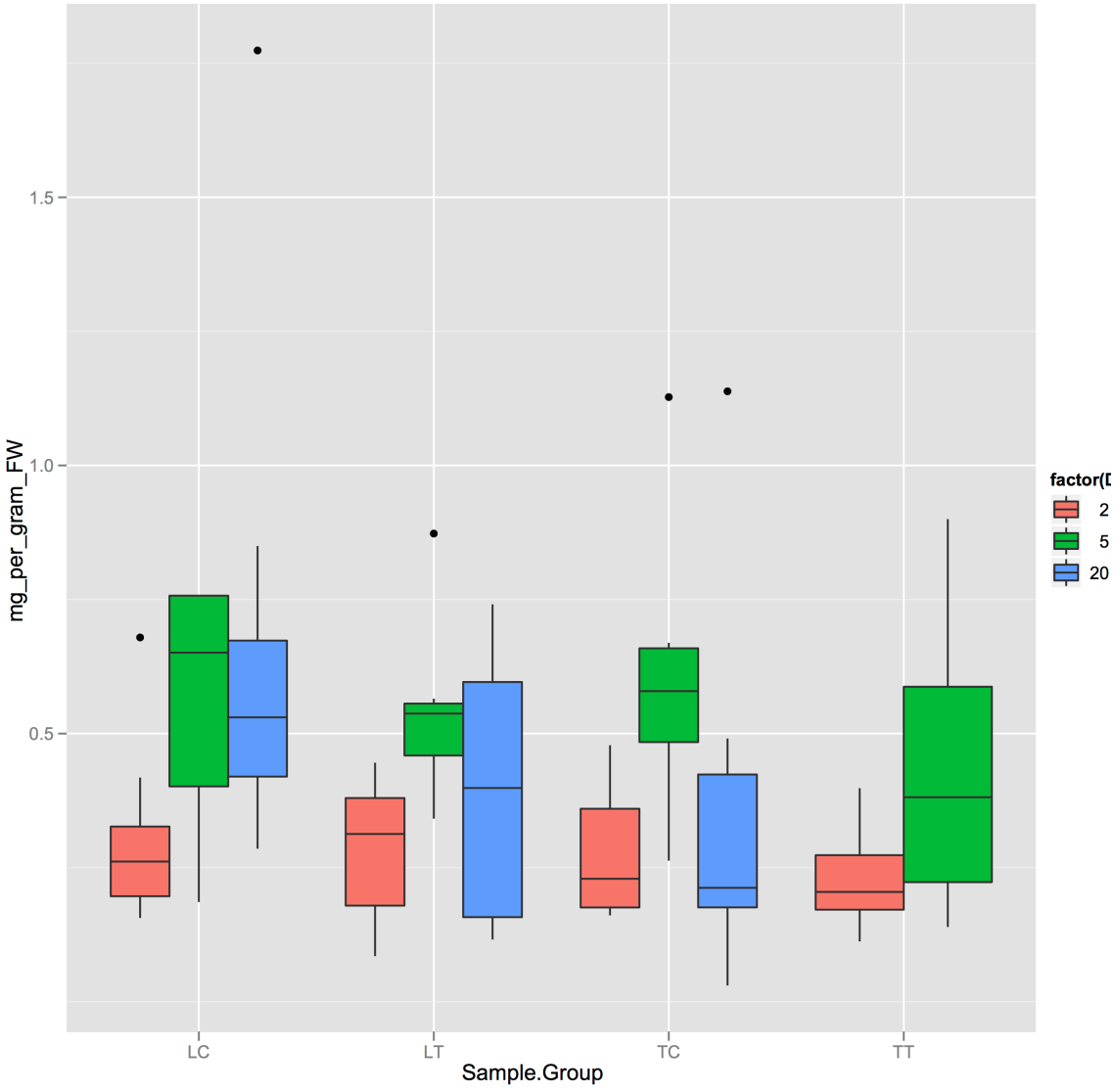

Figure 5.4. Malate concentration in eight biological replicates (the central horizontal line is the median, the box represents the interquartile range, vertical lines are the minimum and maximum values excluding additional plotted points which were considered outliers) for each time point (red at 2 days, green at five days, and blue at 20 days), treatment (‘Latham Treatment’ = LT ‘Titan Treatment’=TT), control (‘Latham Control = LC ‘Titan Control=TC), and genotype showing a consistent trend toward a peak in malate concentration at five days.

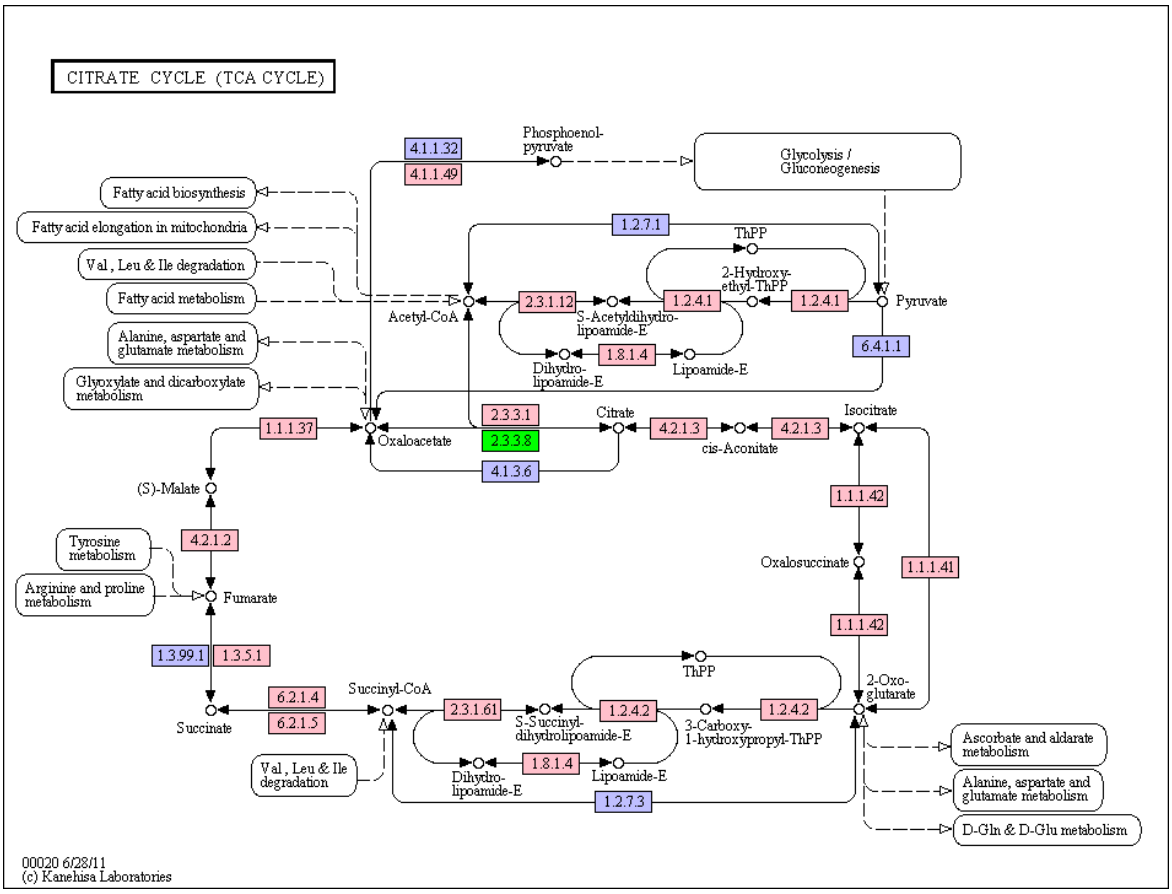

Figure 5.5. KEGG map showing all TCA-related enzymes detected in the RNAseq experiment. Both pink and green were detected in the experiment. Purple enzymes were not. Of significant note is enzyme 2.3.3.8 in green, which corresponds to ATP citrate synthase (gene04096). ATP citrate synthase was among the most highly expressed genes in the resistant treatment and was significantly differentially expressed between the treatment and control with an FPKM of 510.058 in the 'Latham' control and an FPKM of 1007 in the 'Latham' treatment. The same gene had lower expression and was significantly down regulated between the 'Titan' control (FPKM = 386.369) and the 'Titan' treatment (FPKM=212.685).

5.5. References

Bolton, M.D., 2009. Primary Metabolism and Plant Defense—Fuel for the Fire. *Molecular Plant-Microbe Interactions*, 22(5), pp.487–497.

Graham, J. et al., 2011. Towards an understanding of the nature of resistance to Phytophthora root rot in red raspberry. *Theoretical And Applied Genetics*, 123(4), pp.585–601.

Heidel, A.J. et al., 2004. Fitness costs of mutations affecting the systemic acquired resistance pathway in Arabidopsis thaliana. *Genetics*, 168(4), pp.2197–2206.

Heil, M., Hilpert, A. & Kaiser, W., 2000. Reduced growth and seed set following chemical induction of pathogen defence: does systemic acquired resistance (SAR) incur allocation costs? *Journal of Ecology*, 88, pp.645–654.

Lisec, J. et al., 2006. Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nature Protocols*, 1(1), pp.387–396.

Niemann, G.J. et al., 1992. Differences in relative growth rate in 11 grasses correlate with differences in chemical composition as determined by pyrolysis mass spectrometry. *Oecologia*, 89(4), pp.567–573.

Pattison, J.A., Samuelian, S.K. & Weber, C.A., 2007. Inheritance of Phytophthora root rot resistance in red raspberry determined by generation means and molecular linkage analysis. *Theoretical And Applied Genetics*, 115(2), pp.225–236.

Pattison, J.A., Wilcox, W. & Weber, C.A., 2004. Assessing the resistance of red raspberry (Rubus idaeus L.) genotypes to Phytophthora fragariae var. rubi in hydroponic culture. *Hortscience*, 39(7), pp.1553–1556.

Smedegaard-Petersen, V. & Tolstrup, K., 1985. The Limiting Effect of Disease Resistance on Yield. *Annual Review of Phytopathology*, 23(1), pp.475–490.

Tian, D. et al., 2003. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature*, 423(6935), pp.74–77.

Wilcox, W., 1989. Identity, virulence, and isolation frequency of seven Phytophthora spp. causing root rot of raspberry in New York. *Phytopathology*, 79(1), pp.93–101.

Wilcox, W. et al., 1993. Identity of a Phytophthora species attacking raspberry in Europe and North America. *Mycological Research*, 97, pp.817–831.

Zangerl, A.R., Arntz, A.M. & Berenbaum, M.R., 1997. Physiological price of an induced chemical defense: photosynthesis, respiration, biosynthesis, and growth. *Oecologia*, 109(3), pp.433–441.

APPENDIX 1

Significantly differentially expressed genes in 'Latham' with GO Terms associated with defense responses and their corresponding expression levels in 'Titan'.

GeneID	FPKM Latham Control	FPKM 3 Latham Treatment	Significantly up or down regulated in Latham?	FPKM Titan Control	FPKM Titan Treatment	Significantly up or down regulated in Titan?	gene id	GO DATA
gene17426	455.58	556.26	up	407.58	621.27	up	lipid binding	P:systemic acquired resistance, salicylic acid mediated signaling pathway; F:zinc ion binding; F:fatty acid binding
gene04076	397.83	543.48	up	313.32	98.12	down	mlp-like protein 28	P:defense response; P:response to biotic stimulus; F:protein binding; C:chloroplast
gene00517	333.91	387.72	up	209.12	195.09	-	nadp-specific isocitrate dehydro genase	P:glyoxylate cycle; C:cytosol; F:NAD or NADH binding; F:copper ion binding; P:NADP metabolic process; P:response to salt stress; P:defense response to bacterium; C:plasmodesma; P:tricarboxylic acid cycle; F:isocitrate dehydrogenase (NADP+) activity; F:protein binding; F:magnesium ion binding; P:isocitrate metabolic process; P:response to zinc ion; P:response to cadmium ion; C:plasma membrane; C:apoplast
gene01663	190.97	249.72	up	165.18	234.79	up	arginine decarbox ylase	P:response to oxidative stress; P:response to jasmonic acid stimulus; P:putrescine biosynthetic process; P:seed development; P:spermidine biosynthetic process; P:arginine catabolic process;

								P:response to salt stress; P:response to cold; P:response to wounding; F:arginine decarboxylase activity; P:response to abscisic acid stimulus;
gene26770	142.93	234.23	up	65.6	125.85	up	phospho enolpyru vate carboxy kinase	C:cytosol; C:nucleolus; F:phosphoenolpyruvate carboxykinase (ATP) activity; P:leaf morphogenesis; F:DNA binding; F:ATP binding; F:protein binding; P:phosphorylation; F:kinase activity; P:defense response to fungus, incompatible interaction; P:gluconeogenesis; P:response to cadmium ion; P:cellular response to phosphate starvation; C:membrane
gene12042	354.05	230.85	down	162.46	216.47	up	14-3-3-like protein gf14 kappa	C:cytosol; F:protein phosphorylated amino acid binding; P:defense response to bacterium; C:chloroplast envelope; C:plant-type cell wall; F:protein domain specific binding; P:leaf development; P:brassinosteroid mediated signaling pathway; P:response to cadmium ion; C:nucleus; C:plasma membrane
gene32372	125.43	183.93	up	56.41	135.22	up	gamma-glutamyl cysteine syntheta se	P:flower development; P:defense response to bacterium, incompatible interaction; P:callose deposition in cell wall during defense response; P:response to jasmonic acid stimulus; P:defense response to fungus; P:indole phytoalexin biosynthetic process; C:chloroplast stroma; P:glucosinolate biosynthetic process; P:defense response to insect; P:response to cadmium ion; P:response to heat; P:response to ozone; P:glutathione

								biosynthetic process; F:ATP binding; F:glutamate-cysteine ligase activity
gene17454	115.28	168.37	up	71.69	50.37	down	atp binding	P:negative regulation of defense response; C:plasmodesma; C:cytoplasmic membrane-bounded vesicle; C:plasma membrane; F:protein kinase activity; P:protein amino acid phosphorylation; F:ATP binding; F:receptor serine/threonine kinase binding; C:integral to membrane
gene28681	112.6	157.35	up	77.98	140.33	up	glutathione s-transferase	C:cytosol; F:copper ion binding; P:defense response to bacterium; F:glutathione peroxidase activity; C:plasmodesma; F:glutathione transferase activity; C:thylakoid; C:chloroplast stroma; F:glutathione binding; P:response to zinc ion; C:vacuole; P:response to cadmium ion; P:toxin catabolic process; C:plasma membrane; C:apoplast
gene00682	89.03	154.63	up	45.56	119.29	up	protein	P:response to jasmonic acid stimulus; F:(E)-beta-ocimene synthase activity; P:monoterpenoid biosynthetic process; P:response to wounding; F:alpha,alpha-trehalose-phosphate synthase (UDP-forming) activity; F:myrcene synthase activity; P:response to herbivore; F:trehalose-phosphatase activity; P:trehalose biosynthetic process
gene15951	100.66	144.01	up	68.73	63.7	-	protein topless	C:cytosol; P:jasmonic acid mediated signaling pathway; F:Hsp90 protein binding; P:primary shoot apical meristem specification; P:xylem and phloem pattern formation; P:response to auxin

								stimulus; F:protein homodimerization activity; P:negative regulation of transcription, DNA-dependent; C:nucleus
gene32492	166.03	130.81	down	68.45	46.39	down	auxin-responsive protein iaa7	P:lateral root morphogenesis; P:response to jasmonic acid stimulus; P:response to water deprivation; P:response to wounding; F:protein dimerization activity; F:transcription factor activity; P:gravitropism; P:negative regulation of transcription, DNA-dependent; P:auxin mediated signaling pathway; C:nucleus
gene08833	87.63	129.1	up	54.86	83	up	peroxisomal acyl-oxidase 1a	C:cytosol; F:acyl-CoA dehydrogenase activity; F:cytochrome-c oxidase activity; C:peroxisome; P:response to fungus; F:acyl-CoA oxidase activity; C:plasmodesma; P:response to wounding; P:jasmonic acid biosynthetic process; P:fatty acid beta-oxidation; P:pollen development; F:FAD binding; F:nucleotide binding; P:defense response to insect; P:long-chain fatty acid metabolic process; P:response to cadmium ion
gene07034	10.74	128.9	up	1.47	34.74	up	allergen pru p	P:response to biotic stimulus; P:defense response
gene31799	98.81	124.18	up	57.09	241.8	up	serine threonine-protein kinase	C:cytosol; P:protein amino acid phosphorylation; P:response to water deprivation; P:response to salt stress; P:defense response to bacterium; P:growth; F:calcium-dependent protein serine/threonine kinase activity; P:sucrose metabolic process; P:regulation of stomatal movement; F:ATP

								binding; F:protein binding; P:unsaturated fatty acid biosynthetic process; P:triglyceride biosynthetic process; P:leaf development; P:abscisic acid mediated signaling pathway; ; C:nucleus
gene27687	91.21	123.16	up	59.37	282.38	up	protein	F:methyl salicylate esterase activity; P:positive regulation of gibberellic acid mediated signaling pathway; P:floral organ morphogenesis; F:receptor activity; F:methyl jasmonate esterase activity; F:carboxylesterase activity; F:protein binding; F:methyl indole-3-acetate esterase activity; P:raffinose family oligosaccharide biosynthetic process
gene11058	51.55	116.54	up	28.6	28.34	-	major allergen pru	P:response to biotic stimulus; P:defense response
gene25666	20.27	103.87	up	14.92	24.26	up	atp-binding cassette (PEN3)	P:cellular response to indolebutyric acid stimulus; C:vacuolar membrane; P:negative regulation of defense response; F:phosphoprotein phosphatase activity; P:cadmium ion transport; C:chloroplast envelope; F:lupeol synthase activity; P:indole glucosinolate catabolic process; F:cadmium ion transmembrane transporter activity; P:callose deposition in cell wall during defense response; P:defense response to bacterium; P:drug transmembrane transport; C:cytosol; P:defense response to fungus, incompatible interaction; P:dephosphorylation; P:response to abscisic acid

								stimulus; C:plasmodesma; F:phosphonate transmembrane-transporting ATPase activity; P:ATP catabolic process; C:plasma membrane; P:pentacyclic triterpenoid biosynthetic process; F:ATP binding; C:mitochondrion; P:systemic acquired resistance
gene23316	62.23	101.95	up	63.74	178.96	up	disease resistance-dirigent domain-containing protein	C:cytoplasmic membrane-bounded vesicle; P:defense response; P:lignan biosynthetic process
gene29352	55.83	101.85	up	1.86	-	down	protein	P:systemic acquired resistance; F:methyl indole-3-acetate esterase activity; P:defense response to fungus, incompatible interaction; F:polyneuridine-aldehyde esterase activity; P:salicylic acid metabolic process; F:methyl salicylate esterase activity; F:methyl jasmonate esterase activity
gene30873	142.18	94.76	down	75.22	89.4	-	histone 2	C:nucleosome; P:defense response to bacterium; P:regulation of flower development; F:DNA binding; P:detection of temperature stimulus; F:protein binding; P:nucleosome assembly; C:vacuole; C:nucleus
gene07610	55.81	79.42	up	35.23	30.84	-	argonaute protein group	P:translational initiation; P:chromatin silencing by small RNA; C:Cajal body; F:protein binding; P:histone H3-K9 methylation; P:defense response to bacterium, incompatible interaction; C:nuclear euchromatin; F:siRNA binding; F:translation initiation factor activity;

								P:production of siRNA involved in RNA interference; C:nucleolus; C:cytosol; P:long-distance posttranscriptional gene silencing; P:DNA methylation
gene30476	25.26	79.26	up	4.54	4.53	-	wound-induced protein win2	P:defense response to fungus; C:extracellular region; F:chitin binding; P:defense response to bacterium; C:cell wall
gene29486	58.83	74.98	up	33.27	83.45	up	splicing factor u2af large subunit a	P:defense response to bacterium; F:RNA binding; C:nucleus; P:nuclear mRNA splicing, via spliceosome; F:nucleotide binding
gene20057	51.1	73.99	up	25.43	120.11	up	gibberellin receptor	F:methyl salicylate esterase activity; P:positive regulation of gibberellic acid mediated signaling pathway; P:floral organ morphogenesis; F:receptor activity; F:methyl jasmonate esterase activity; F:carboxylesterase activity; F:protein binding; F:methyl indole-3-acetate esterase activity; P:raffinose family oligosaccharide biosynthetic process
gene03083	55.96	72.71	up	31.63	51.56	up	protein	P:vesicle-mediated transport; C:mitochondrion; C:plasmodesma; C:vacuolar membrane; P:GTP catabolic process; F:GTPase activity; P:response to cadmium ion; P:defense response to fungus; F:GTP binding; C:plasma membrane
gene04073	23.58	71.99	up	28.37	9.56	down	mlp-like protein 28	P:defense response; P:response to biotic stimulus; F:protein binding; C:chloroplast
gene11527	51.42	70.86	up	31.84	51.28	up	pre-mrna-processing factor 19-2	C:nucleolus; P:defense response to bacterium; P:response to cadmium ion; F:protein binding; P:protein ubiquitination;

								C:cell wall; C:CUL4 RING ubiquitin ligase complex; F:ubiquitin-protein ligase activity; C:chloroplast; F:nucleotide binding
gene26291	44.95	64.46	up	25.81	21.09	-	phosphatidylinositol- -trisphosphate 3-phosphatase	P:cellular response to indolebutyric acid stimulus; F:phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase activity; C:vacuolar membrane; P:negative regulation of defense response; F:protein tyrosine/serine/threonine phosphatase activity; P:cadmium ion transport; C:chloroplast envelope; P:tetracyclic triterpenoid biosynthetic process; F:lupeol synthase activity; F:protein tyrosine phosphatase activity; P:indole glucosinolate catabolic process; F:cadmium ion transmembrane transporter activity; F:baruol synthase activity; P:callose deposition in cell wall during defense response; P:defense response to bacterium; P:drug transmembrane transport; C:peroxisome; C:cytosol; F:thioglucosidase activity; P:peptidyl-tyrosine dephosphorylation; P:defense response to fungus, incompatible interaction; F:ATPase activity, coupled to transmembrane movement of substances; P:response to abscisic acid stimulus; C:plasmodesma; P:pentacyclic triterpenoid biosynthetic process; C:plasma membrane; P:systemic acquired resistance; C:mitochondrion
gene28816	42.06	55.2	up	19.31	13.82	-	amp	P:auxin biosynthetic

							dependent	process; F:fatty-acyl-CoA synthase activity; P:jasmonic acid biosynthetic process; C:peroxisome; F:Photinus-luciferin 4-monooxygenase (ATP-hydrolyzing) activity; F:4-coumarate-CoA ligase activity; P:oxidation reduction
gene30835	38.22	55.11	up	23.65	29.54	up	aim1 protein	C:cell wall; P:seed germination; C:peroxisome; C:plasmodesma; F:3-hydroxybutyryl-CoA epimerase activity; P:jasmonic acid biosynthetic process; P:flower development; F:dodecenoyl-CoA delta-isomerase activity; F:3-hydroxyacyl-CoA dehydrogenase activity; P:fatty acid beta-oxidation; F:coenzyme binding; C:chloroplast; F:enoyl-CoA hydratase activity; F:3-hydroxyacyl-CoA dehydratase activity
gene12546	35.4	52.33	up	23.34	26.25	-	protein topless	C:cytosol; P:jasmonic acid mediated signaling pathway; P:primary shoot apical meristem specification; P:xylem and phloem pattern formation; P:response to auxin stimulus; F:protein homodimerization activity; P:negative regulation of transcription, DNA-dependent; C:nucleus
gene24788	32.58	50.68	up	19.34	21.31	-	cullin 1	C:cytosol; P:ubiquitin-dependent protein catabolic process; P:cell cycle; P:jasmonic acid mediated signaling pathway; P:regulation of circadian rhythm; C:condensed nuclear chromosome; P:SCF complex assembly; C:cullin-RING ubiquitin ligase complex; F:ubiquitin protein ligase

								binding; P:response to auxin stimulus; C:spindle; C:phragmoplast; P:embryonic development ending in seed dormancy
gene27117	27.83	50.65	up	28.5	36.54	-	protein	P:defense response; C:cytosol; F:transferase activity, transferring phosphorus-containing groups; F:RNA binding; P:ethylene biosynthetic process; P:cell death; C:nucleolus; F:protein binding; P:leaf senescence; F:nucleotide binding
gene12639	31.37	49.19	up	22.26	21.43	-	mitogen-activated protein kinase kinase 2	F:MAP kinase kinase activity; P:protein amino acid phosphorylation; F:protein serine/threonine kinase activity; P:response to water deprivation; P:response to salt stress; P:defense response to bacterium; P:MAPKKK cascade; P:response to wounding; F:ATP binding; P:cold acclimation; F:protein binding; P:response to molecule of bacterial origin; C:cytoplasm; P:response to hydrogen peroxide; P:defense response, incompatible interaction; C:plasma membrane
gene15248	25.49	48.75	up	22.33	62.74	up	late embryogenesis abundant hydroxyproline-rich glycoprotein	P:defense response to virus; C:plasmodesma
gene11666	31.48	47.26	up	26.79	17.48	-	protein phosphatase 2c	C:protein serine/threonine phosphatase complex; F:metal ion binding; F:protein serine/threonine phosphatase activity; P:regulation of defense response to virus; P:protein amino acid dephosphorylation

gene17117	32.44	46.72	up	27.82	46.55	up	histone deacetylase	F:histone deacetylase activity (H3-K16 specific); ; P:DNA mediated transformation; P:histone deacetylation; F:NAD-dependent histone deacetylase activity (H3-K14 specific); P:jasmonic acid and ethylene-dependent systemic resistance; F:protein binding; F:NAD-dependent histone deacetylase activity (H4-K16 specific); F:NAD-dependent histone deacetylase activity (H3-K9 specific); P:histone acetylation; P:negative regulation of transcription, DNA-dependent; C:nucleus
gene09784	4.29	46.37	up	-	48.02	up	thaumatin-like protein	P:defense response; C:extracellular region; F:IgE binding; P:metabolic process; F:glucan endo-1,3-beta-D-glucosidase activity
gene00479	24.42	45.03	up	7.21	8.99	-	late embryogenesis abundant hydroxyproline-rich glycoprotein	P:salicylic acid mediated signaling pathway; P:defense response, incompatible interaction
gene09719	24.29	43.5	up	21.48	33.31	up	ein2 - like nramp transporter	P:negative regulation of defense response; F:protein binding; P:regulation of stomatal movement; P:response to salt stress; P:callose deposition in cell wall during defense response; P:response to jasmonic acid stimulus; P:defense response to fungus; P:defense response to bacterium; P:establishment of planar polarity; P:sugar mediated signaling pathway; P:root hair cell differentiation; P:leaf senescence; P:auxin polar transport;

								P:response to heat; P:response to molecule of bacterial origin; P:jasmonic acid and ethylene-dependent systemic resistance, ethylene mediated signaling pathway; P:positive regulation of abscisic acid mediated signaling pathway; P:response to oxidative stress; P:cellular response to iron ion
gene31001	27.68	43.23	up	18.2	26.04	up	protein topless	C:cytosol; P:jasmonic acid mediated signaling pathway; P:primary shoot apical meristem specification; P:xylem and phloem pattern formation; P:response to auxin stimulus; F:protein homodimerization activity; P:negative regulation of transcription, DNA-dependent; C:nucleus
gene16676	25.07	41.16	up	18.7	21.96	-	protein ilityhia	C:membrane; C:cytosol; P:defense response to bacterium; F:binding; C:nucleus
gene00851	27.12	39.66	up	23.33	64.75	up	rrm-containing rna-binding protein	F:RNA binding; P:defense response to bacterium; C:nucleus; F:nucleotide binding; F:zinc ion binding
gene12974	23.86	36.43	up	9.31	6.86	-	mlo-like protein 10	C:endomembrane system; P:defense response; F:calmodulin binding; C:integral to membrane; P:cell death; C:cytoplasm; P:pollen tube reception; C:plasma membrane
gene14177	16.66	36.4	up	12.17	23.06	up	at5g13530-like protein	P:negative regulation of abscisic acid mediated signaling pathway; P:defense response; P:protein amino acid phosphorylation; F:zinc ion binding; F:protein kinase activity; C:plastid; F:protein self-association; C:early endosome; F:ubiquitin-protein ligase activity; F:ATP binding;

								P:protein ubiquitination; P:developmental growth; C:trans-Golgi network; F:phospholipase A2 activity
gene05590	23.17	36.13	up	12.5	44.84	up	phospho enolpyru vate carboxy kinase	C:cytosol; C:nucleolus; F:phosphoenolpyruvate carboxykinase (ATP) activity; P:leaf morphogenesis; F:DNA binding; F:ATP binding; F:protein binding; P:phosphorylation; F:kinase activity; P:defense response to fungus, incompatible interaction; P:gluconeogenesis; P:response to cadmium ion; P:cellular response to phosphate starvation; C:membrane
gene12924	21.78	28.74	up	13.66	7.94	down	guanine nucleoti de-exchang e factor	P:defense response to bacterium; C:early endosome; P:regulation of ARF protein signal transduction; F:ARF guanyl-nucleotide exchange factor activity; F:protein binding; P:vesicle-mediated transport; P:growth; P:regulation of catalytic activity; C:cytosol; C:trans-Golgi network
gene17682	20.25	27.39	up	13.83	11.97	-	glucan synthase compone nt	P:1,3-beta-glucan biosynthetic process; P:defense response to bacterium; C:plasmodesma; C:1,3-beta-glucan synthase complex; P:leaf morphogenesis; P:reproduction; P:defense response signaling pathway, resistance gene-dependent; P:circadian rhythm; C:cytoplasmic membrane-bounded vesicle; P:callose deposition in cell wall during defense response; F:1,3-beta-glucan synthase activity; P:pollen development; P:salicylic acid mediated signaling

								pathway; P:leaf senescence; P:defense response to fungus
gene29192	19.04	26.34	up	14.52	11.88	-	protein	P:negative regulation of flower development; F:porin activity; C:nuclear membrane; C:nuclear pore; P:defense response signaling pathway, resistance gene-dependent; P:response to auxin stimulus; F:protein binding; P:mRNA export from nucleus
gene00245	13.65	25.35	up	4.83	5.16	-	rna binding protein	F:nutrient reservoir activity; C:ribosome; P:seed germination; C:peroxisome; P:seed maturation; P:cellular response to abscisic acid stimulus; P:defense response to bacterium; C:plasmodesma; P:rRNA processing; P:response to cold; C:chloroplast envelope; P:circadian rhythm; P:chloroplast organization; C:plastoglobule; C:endomembrane system; F:coenzyme binding; F:catalytic activity; C:vacuole; C:membrane; C:stromule; C:apoplast
gene29068	15	25.06	up	9.03	14.83	up	spl1-related2 protein	P:defense response to bacterium; F:DNA binding; F:transcription factor activity; P:positive regulation of transcription, DNA-dependent; C:nucleus; C:plasma membrane
gene17518	16.92	24.19	up	7.39	9.97	-	3-ketoacyl-thiolase	P:positive regulation of abscisic acid mediated signaling pathway; C:cytosol; P:glyoxysome organization; C:peroxisome; C:mitochondrion; C:nucleolus; F:acetyl-CoA C-acyltransferase activity; P:response to wounding; P:jasmonic acid biosynthetic process; C:vacuolar membrane;

								P:fatty acid beta-oxidation; C:chloroplast
gene11014	17.66	24.14	up	9.83	3.64	down	udp-glucose: glycoprotein glucosyltransferase	P:carbohydrate transport; P:anthocyanin metabolic process; F:RNA-directed DNA polymerase activity; F:pyrimidine nucleotide sugar transmembrane transporter activity; P:response to salicylic acid stimulus; P:protein amino acid glycosylation; P:carbohydrate biosynthetic process; P:defense response signaling pathway, resistance gene-independent; F:UDP-glucose:glycoprotein glucosyltransferase activity; P:RNA-dependent DNA replication; F:RNA binding; P:plant-type hypersensitive response; C:endoplasmic reticulum
gene14977	14.02	24.01	up	8.24	36.13	up	cbl-interacting protein kinase	F:calmodulin-dependent protein kinase activity; F:protein binding; P:defense response to fungus; P:protein amino acid phosphorylation; F:ATP binding; P:signal transduction
gene32316	1.52	21.6	up	-	0.97	up	thaumatin-like protein	C:extracellular region; P:defense response to fungus; P:killing of cells of another organism
gene07425	12.65	21.49	up	7.64	9	-	protein	P:arsenite transport; F:cofactor binding; F:zinc ion binding; C:membrane; P:indole glucosinolate catabolic process; P:immune response; ; P:phytochelatin biosynthetic process; P:callose deposition in cell wall during defense response; P:lignin biosynthetic process; P:response to arsenic; P:defense response to bacterium; C:cytosol; F:oxidoreductase activity, acting on the CH-OH

								group of donors, NAD or NADP as acceptor; F:cadmium ion binding; P:cell death; P:response to cadmium ion; P:oligopeptide transport; F:arsenite transmembrane-transporting ATPase activity; F:copper ion binding; F:glutathione gamma-glutamylcysteinyltransferase activity
gene19169	8.28	19.94	up	8.77	4.86	-	beclin 1 protein	C:cytosol; C:peroxisome; P:protein targeting to vacuole; F:acyl-CoA oxidase activity; P:short-chain fatty acid metabolic process; C:pre-autophagosomal structure; P:autophagic vacuole assembly; P:fatty acid beta-oxidation; P:pollen germination; P:defense response to fungus; P:embryonic development ending in seed dormancy
gene26386	11.37	19.67	up	8.3	6.41	-	exostosin family protein	C:cytosol; F:adenylosuccinate synthase activity; C:peroxisome; P:purine nucleotide biosynthetic process; P:response to fungus; F:acyl-CoA oxidase activity; C:plasmodesma; P:response to wounding; P:jasmonic acid biosynthetic process; C:cytoplasmic membrane-bounded vesicle; P:fatty acid beta-oxidation; P:pollen development; F:magnesium ion binding; C:endomembrane system; P:defense response to insect; P:long-chain fatty acid metabolic process; P:response to cadmium ion; C:membrane; F:GTP binding
gene04426	11.88	19.4	up	7.89	6.32	-	glycogen synthase	P:response to jasmonic acid stimulus; F:starch synthase activity; C:plasmodesma; P:response to wounding;

								F:strictosidine synthase activity; C:vacuolar membrane; C:chloroplast; C:plasma membrane; P:glycogen biosynthetic process
gene03843	6.95	18.91	up	4.96	6.8	-	phytochrome and flowering time regulatory protein 1	P:red, far-red light phototransduction; P:response to red light; P:jasmonic acid mediated signaling pathway; P:regulation of flower development; P:positive regulation of defense response; C:mediator complex; F:transcription coactivator activity; P:response to far red light; P:positive regulation of transcription, DNA-dependent; P:defense response to fungus
gene32055	10.43	18.76	up	6.68	9.71	-	protein	P:response to jasmonic acid stimulus; F:(E)-beta-ocimene synthase activity; P:monoterpenoid biosynthetic process; P:response to wounding; F:alpha,alpha-trehalose-phosphate synthase (UDP-forming) activity; F:myrcene synthase activity; P:response to herbivore; F:trehalose-phosphatase activity; P:trehalose biosynthetic process
gene23483	11.84	18.69	up	8.41	7.26	-	kelch repeat-containing protein	P:response to jasmonic acid stimulus; F:acyl-CoA binding; F:protein binding; P:lipid transport; P:response to ethylene stimulus; C:nucleus; C:cytosol; P:response to light stimulus
gene05701	9.47	16.83	up	7.51	4.42	-	btb and taz domain protein 4	P:response to jasmonic acid stimulus; F:calmodulin binding; F:transcription cofactor activity; P:response to gibberellin stimulus; F:zinc ion binding; P:response to salt stress; P:response to cold; P:regulation of

								transcription, DNA-dependent; P:response to wounding; P:response to salicylic acid stimulus; C:vacuolar membrane; P:response to auxin stimulus; P:response to chitin; F:histone acetyltransferase activity; P:histone acetylation; P:response to hydrogen peroxide; C:nucleus
gene05033	6.77	15.53	up	6.57	15.48	up	protein	P:defense response to bacterium; F:sequence-specific DNA binding; P:response to chitin; F:transcription factor activity; P:positive regulation of transcription, DNA-dependent; C:nucleus
gene18938	8.59	15.45	up	4.85	8.13	up	phospholipase d	P:response to cadmium ion; P:lipid catabolic process; C:plasmodesma; F:protein binding; F:NAPE-specific phospholipase D activity; F:calcium ion binding; F:phospholipase D activity; C:membrane; P:phosphatidylcholine metabolic process; P:defense response to bacterium, incompatible interaction; F:phosphatidylinositol-4,5-bisphosphate binding
gene32258	8.19	15.41	up	3.64	8.41	up	phospholipase d	P:response to cadmium ion; P:lipid catabolic process; C:plasmodesma; F:protein binding; F:NAPE-specific phospholipase D activity; F:calcium ion binding; F:phospholipase D activity; C:membrane; P:phosphatidylcholine metabolic process; P:defense response to bacterium, incompatible interaction; F:phosphatidylinositol-4,5-bisphosphate binding
gene07895	6.57	14.41	up	4.08	5.77	-	protein phosphat	C:protein serine/threonine phosphatase complex;

							ase 2c	F:metal ion binding; P:response to wounding; F:protein serine/threonine phosphatase activity; P:defense response to fungus; P:protein amino acid dephosphorylation; C:plastid; P:abscisic acid mediated signaling pathway
gene06346	3.7	13.56	up	3.88	7.5	-	acyl:coa ligase	P:auxin biosynthetic process; F:fatty-acyl-CoA synthase activity; P:jasmonic acid biosynthetic process; F:oxidoreductase activity; C:peroxisome; F:4-coumarate-CoA ligase activity
gene21970	0.91	13.4	up	1.85	0.77	-	wrky transcrip tion	P:jasmonic acid mediated signaling pathway; P:defense response to bacterium; P:regulation of transcription, DNA-dependent; C:plastid; F:sequence-specific DNA binding; F:transcription factor activity; P:defense response to fungus
gene21319	5.32	13.34	up	2.08	1.73	-	wrky transcrip tion	P:induced systemic resistance, jasmonic acid mediated signaling pathway; P:systemic acquired resistance, salicylic acid mediated signaling pathway; P:response to chitin; F:protein binding; F:transcription factor activity; P:negative regulation of transcription, DNA-dependent; P:defense response to fungus
gene09313	7.47	12.84	up	4.29	7.51	up	nuclear transcrip tion x-box	P:response to light intensity; F:zinc ion binding; F:nucleic acid binding; C:mitochondrion; P:response to salt stress; P:defense response to bacterium; P:regulation of transcription, DNA-dependent; F:potassium ion transmembrane transporter activity;

								P:response to microbial phytotoxin; ; F:transcription factor activity; F:nucleotide binding; P:regulation of hydrogen peroxide metabolic process; C:membrane; C:nucleus; P:salicylic acid biosynthetic process
gene13759	7.44	12.6	up	1.98	4.94	up	nbs-lrr resistance protein	P:defense response; P:apoptosis; F:ATP binding; F:hydrolase activity
gene09621	4.46	12.56	up	3.48	3.25	-	protein	P:negative regulation of defense response; P:leaf senescence; P:defense response to fungus, incompatible interaction; C:plasmodesma; C:plasma membrane
gene13022	1.03	12.36	up	-	2.91	up	class i chitinase	F:chitinase activity; P:defense response; F:chitin binding; P:cell wall macromolecule catabolic process; P:chitin catabolic process
gene07235	6.4	12.12	up	5.51	1.66	down	protein	P:negative regulation of abscisic acid mediated signaling pathway; P:trichome differentiation; P:response to jasmonic acid stimulus; P:response to salt stress; P:response to salicylic acid stimulus; F:protein C-terminus binding; P:vegetative to reproductive phase transition of meristem; P:root hair cell differentiation; F:phosphoprotein phosphatase activity; C:nucleus
gene20031	5.36	11.49	up	3.33	5.39	-	eukaryotic translation initiation factor	P:defense response to virus; F:siRNA binding; P:translational initiation; F:translation initiation factor activity
gene16054	2.74	11.28	up	0.54	-	down	protein	F:L-aspartate:2-oxoglutarate aminotransferase activity; F:zinc ion binding; P:systemic acquired

								resistance, salicylic acid mediated signaling pathway; F:L,L-diaminopimelate aminotransferase activity; C:cytosol; P:regulation of ARF GTPase activity; C:chloroplast stroma; F:pyridoxal phosphate binding; P:lysine biosynthetic process via diaminopimelate, diaminopimelate-aminotransferase pathway; F:copper ion binding; F:ARF GTPase activator activity; C:mitochondrion
gene04859	6	11.25	up	5.25	7.97	-	protein	P:negative regulation of programmed cell death; P:defense response to bacterium; C:plastid
gene05279	5.89	11.02	up	3.05	6.57	up	lipoxygenase	P:defense response; P:response to jasmonic acid stimulus; P:response to ozone; P:oxidation reduction; P:response to fungus; P:response to wounding; P:growth; P:stamen filament development; P:jasmonic acid biosynthetic process; P:response to bacterium; P:pollen development; F:lipoxygenase activity; C:chloroplast; P:response to high light intensity; P:anther dehiscence; F:iron ion binding
gene07367	5.24	10.98	up	6.26	2.23	down	transcription initiation factor tfiid subunit a	P:regulation of ethylene mediated signaling pathway; P:jasmonic acid mediated signaling pathway; P:transcription initiation; F:DNA binding; C:transcription factor TFIID complex; P:cytokinin mediated signaling pathway
gene04020	37.69	10.93	down	11.35	5.66	-	uncharacterized protein	P:defense response to fungus
gene05115	19.27	9.42	down	1.78	16.4	up	beta- -glucanase	P:carbohydrate metabolic process; P:defense response; C:vacuole;

								F:glucan endo-1,3-beta-D-glucosidase activity; F:cation binding
gene25058	2.79	9.17	up	1.67	2.37	-	calmodulin-binding protein	P:regulation of salicylic acid biosynthetic process; F:transcription factor activity; F:sequence-specific DNA binding; P:regulation of systemic acquired resistance; C:nucleus; F:calmodulin binding
gene13138	4.28	9.07	up	1.07	2.02	-	heat shock protein 81-1	P:protein folding; C:plasma membrane; C:cell wall; P:response to arsenic; F:ATP binding; F:unfolded protein binding; C:cytosol; P:response to heat; P:defense response to bacterium, incompatible interaction
gene30533	0	8.68	up	-	-	-	protein tify 5a	P:response to chitin; F:protein binding; P:response to jasmonic acid stimulus
gene00713	1.14	7.79	up	0.26	-	down	udp-glucosyl transferase 74b1	P:defense response to bacterium; P:glucosinolate biosynthetic process; F:thiohydroximate beta-D-glucosyltransferase activity; P:callose deposition in cell wall during defense response; F:indole-3-acetate beta-glucosyltransferase activity
gene27528	3.42	7.48	up	2.19	3.46	-	heat shock protein	P:protein folding; C:nucleomorph; C:plasma membrane; C:cell wall; P:response to arsenic; F:ATP binding; F:unfolded protein binding; C:cytosol; P:response to heat; P:defense response to bacterium, incompatible interaction
gene00434	2.46	7.07	up	2.1	2.09	-	myb domain protein 33	P:positive regulation of abscisic acid mediated signaling pathway; P:negative regulation of growth; P:pollen sperm cell differentiation; P:positive regulation of

								programmed cell death; P:response to salicylic acid stimulus; F:DNA binding; P:response to ethylene stimulus; P:positive regulation of transcription, DNA-dependent; C:nucleus
gene25116	1.68	6.76	up	0.72	0.18	-	benzoate carboxyl	P:methylation; F:jasmonate O-methyltransferase activity
gene18982	15.32	6.74	down	1.95	2.73	-	group 4 late embryogenesis-abundant protein	P:response to desiccation; P:response to high light intensity; P:response to wounding; P:embryonic development ending in seed dormancy; P:defense response to fungus; C:cytosol
gene12877	1.35	6.36	up	0.92	0.86	-	flavin-containing monooxygenase 1	P:cellular response to hypoxia; P:oxidation reduction; P:plant-type hypersensitive response; P:defense response to bacterium; C:intrinsic to endoplasmic reticulum membrane; F:flavin-containing monooxygenase activity; P:defense response signaling pathway, resistance gene-dependent; P:systemic acquired resistance; P:defense response signaling pathway, resistance gene-independent; F:FAD binding; F:NADP or NADPH binding; P:defense response to fungus
gene03983	1.43	6.1	up	1.56	3.16	-	wrky transcription	P:regulation of transcription, DNA-dependent; P:transcription, DNA-dependent; F:sequence-specific DNA binding; F:transcription factor activity; F:DNA binding; P:defense response
gene15007	1.68	5.37	up	1.29	2.69	-	nbs-lrr resistance protein	P:defense response; C:apoplast; F:nucleotide binding
gene11179	0	3.88	up	-	-	-	tir-nbs-	P:defense response;

							lrr resistance protein	P:cellular process; F:nucleotide binding
gene27441	0	3.4	up	0.94	2.93	-	atp synthase delta chain	C:proton-transporting ATP synthase complex, catalytic core F(1); F:hydrogen ion transporting ATP synthase activity, rotational mechanism; C:mitochondrion; P:defense response to bacterium; P:ATP synthesis coupled proton transport; P:photosynthetic electron transport in photosystem I; P:response to cold; C:chloroplast envelope; P:photosynthetic electron transport in photosystem II; C:chloroplast thylakoid membrane; C:plastoglobule; C:plasma membrane; C:stromule
gene10679	4.4	0	down	-	13.72	up	lsd one like 1 protein	P:induction of apoptosis; ; P:positive regulation of plant-type hypersensitive response; F:DNA binding; C:endomembrane system; P:defense response, incompatible interaction