UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato Scienze Agrarie Forestali e Ambientali

Dipartimento Scienze Agrarie, Alimentari e Forestali

Settore Scientifico Disciplinare AGR/03

INVESTIGATION OF BIOTIC STRESS RESPONSES IN FRUIT TREE CROPS USING META-ANALYTICAL TECHNIQUES

IL DOTTORE

BIPIN BALAN

IL COORDINATORE

CH.MO PROF. VINCENZO BAGARELLO

IL TUTOR

CH.MO PROF. TIZIANO CARUSO

CO TUTOR

CH.MO PROF. FEDERICO MARTINELLI

CICLO XXXII

ANNO CONSEGUIMENTO TITOLO 2020

Table of Contents

State of art …......……………………………………………….....………….6

Experiment 1…….………………….…...………………………......……..…10

Transcriptomic responses to biotic stresses in *Malus x domestica*: a meta-analysis study. …………………………………………...........................……..…10

1. Introduction ……………………………………………...............………….…10

2. Materials and methods ………………………………..…………...…..……21

 2.1 Search strategy of published study identification and selection for meta-analysis ……………………………………………………......…..……21

 2.2 Extraction and annotation of differentially expressed genes ……….…....23

 2.3 Gene enrichment analysis ………………………………………......……23

 2.4 Functional analysis…………………………………..………......………23

 2.5 Protein-protein interaction network……………………..……………..…25

3. Results…….……………………………............................……...…27

 3.1 Meta-analysis of transcriptome data …………………….......……..…27

 3.2 Gene ontology analysis ………………………………………………..…28

 3.3 Gene set enrichment analysis ……………………………...……..…32

 3.4 Metabolism overview ……………………………………………….…34

 3.5 Hormone-related pathways ……………………………...……..…35

 3.6 Detoxifying pathways and secondary metabolism ……………..…..……36

 3.7 Transcription factors and defense stress-related genes ……………...……38

 3.8 Commonly regulated genes among biotic stresses ………………..…..…41

3.9 Inferred protein-protein interaction network analysis ……………..…….42

4. Discussion ……………………………………………..........................……….…44

5. Conclusions ……………………………………………..........................….…….51

6. References ……………………………………………..........................….…..….51

Experiment 2…………………………………………….........................……….61

Gaining Insight into Exclusive and Common Transcriptomic Features Linked with Biotic Stress Responses in *Malus x domestica* ……………………………....…61

1. Introduction ……………………………………………..........................……..…61

2. Materials and methods ………………………………………..........………..…63

2.1 Search strategy of published study identification for meta-analysis …..….63

2.2 Differentially expressed gene selection and annotation ……………………64

2.3 Gene enrichment analysis ……………………………………………….....…67

2.4 Functional analysis ………………………………………………............…68

2.5 Protein-protein interaction network ………………………………..……..…70

3. Results …………………………………………..............................……..…….72

3.1 Gene set enrichment analysis ………………………………………..….…73

3.2 Biological process enrichment analysis ……………………………..……75

3.3 Hormone-related pathways …………………………………....………76

3.4 Secondary metabolism ……………………………………………………77

3.5 Protein targeting and transcription factors ……………………………..……78

3.6 Protein-protein interaction network analysis …………………………..…80

4. Discussion ……………………………………………………………81

5. Conclusions ……………………………………………………………85

6. References ……………………………………………………………86

Experiment 3……………………………………………………………92

Identifying Host Molecular Features Strongly Linked With Responses to Huanglongbing Disease in *Citrus sinensis* Leaves.……………………………………92

1. Introduction ……………………………………………………………92

2. Materials and methods ……………………………………………………95

2.1 Search Strategy to Identify Published Studies for Bioinformatic Analysis……………………………………………………………95

2.2 Bioinformatic Analysis of Raw Data ………………………………………96

2.3 Differentially Expressed Gene Selection ……………………………………97

2.4 Splice Analysis ……………………………………………………………97

2.5 Gene Enrichment and Functional Analysis …………………………………97

2.6 Protein–Protein Interaction Network ………………………………………98

3. Results ……………………………………………………………………98

3.1 Workflow, Bioinformatics Analysis, and Venn Diagrams ……………………98

3.2 Gene Set- and Pathway-Enrichment Analysis ………………………………101

3.3 Molecular Responses to Huanglongbing Disease ……………………………103

3.3.1 Metabolism Overview ……………………………………………103

3.3.2 Hormone Overview ………………………………………………104

3.3.3 Transcription Factors ……………………………………………105

3.3.4 Biotic Stress Responses 106

3.3.5 Genes Commonly Involved in HLB Response Between Datasets..107

3.4 Molecular Mechanisms of HLB Tolerance 108

3.4.1 Hormone Overview 109

3.4.2 Transcription Factors 110

3.4.3 Biotic Stress Responses 110

3.4.4 Large Enzyme Families 111

3.5 Protein–Protein Network Analysis 112

3.6 Splice Analysis 114

4. Discussion 118

5. Conclusions 122

6. References 123

General Conclusion..128

Appendix..131

1. TrimSeq.pl 131

2. CheckAfterAdapterTrimming.pl 133

3. ExtractDEGs.pl 136

Acknowledgements...147

State of art

In the past decade, Next Generation Sequencing (NGS) methods have been widely adopted over Sanger sequencing referred to as "first-generation" sequencing due to their dropping costs and ability to sequence DNA at an unprecedented speed. The huge amounts of data generated by NGS have extended our understanding of structural and functional genomics through the concepts of "omics" providing new insight into the workings and meaning of genetic conservation and diversity of living things. NGS technologies can be applied for multiple applications such as Sequencing the Whole-Exome (WES) to identify the genetic variants, whole transcriptome sequencing (RNA-seq) which helps to understand the expression of transcripts , Targeted (TS) or candidate gene sequencing to sequence only the genomic regions of interest to identify variants, and Methylation Sequencing (MeS) or Bisulfite Sequencing to investigate epigenetic modification, which plays a pivotal role in regulating the gene expression. In the area of plant research, NGS technologies have become crucial tools for assembly of crop reference genomes, transcriptome sequencing for the study of gene expression, whole-genome molecular marker development, and identification of markers in known-function genes.

RNA sequencing (RNA-Seq) uses the capabilities of high-throughput sequencing methods to provide higher coverage and greater resolution of the dynamic nature of the transcriptome and the opportunity to elucidate different physiological and pathological conditions. This technology consists of converting RNA molecules to a library of cDNA fragments with adaptors, these fragments are sequenced, and the resulting reads are either aligned to a reference genome (if available), or assembled de novo followed by transcript quantification.

Global gene expression profiling using RNA-Seq technologies has been widely used to study biological and cellular responses due to oxidative stress responses in plants. Since the number of such transcriptome studies is growing, it is very significant to have a comprehensive analysis by integrating multiple studies to identify robust gene expression signatures that would be subtle in individual studies.

Initially developed by medical researchers to synthesize data from multiple clinical trials, systematic literature review and meta-analysis are increasingly popular in the area of agricultural sciences. During the 1920s and 1930s, British statistician Ronald Fisher worked

at the Agricultural Research Station in Rothamstead and in his 1935 textbook; he gives an example of the appropriate analysis of multiple studies in agriculture. Meta-analysis technique has been applied in numerous fields for example, psychology, law, management, education, medicine, and even policy formulation. Across various fields, meta-analysis has been used to examine (a) the strength of relationship between two variables (b) the effectiveness of treatments or interventions (c) the accuracy of theories (d) the validity of measuring instruments (e) the validity of procedures and (f) the presence of moderation effects. Meta-analysis facilitates to derogate or decimate potential biases associated with individual studies and to improve statistical power to enable detection of subtle but biologically meaningful variations through increased sample sizes.

The main objective my PhD projects was to perform a comprehensive study of the application of meta-analytical techniques to analyze gene expression data pertaining to biotic stress in different fruit tree crops in order to detect the strongly associated genes, pathways and gene set categories. Identifying key information in transcriptomic data is very important, especially when the "omic" study deals with plant responses to stresses in field conditions where a high number of variables and disturbing factors may affect the analysis. A wide range of biotic stress due to fungi, bacteria, and virus adversely affect plant growth and productivity worldwide. There were individual transcriptome studies based on individual pathogen attack on different crops, which lacks the significance of identifying the potential genes, which are vulnerable for any biotic stress. The proper understanding of plant stress response mechanisms under various stresses can draw a better view for improving worldwide food production.

In my first meta-analysis study, the objective was to identify specific and common molecular responses between different transcriptomic data related to fungi, virus and bacteria attacks in *Malus x domestica*. In this study, the transcriptomic datasets in *Malus x domestica* were collected from published literatures and divided into three groups, according to the pathogen type as a) responses to fungal pathogens, b) virus and c) bacteria (*Erwinia amylovora*). Data were dissected using an integrated approach of pathway- and gene- set enrichment analysis, Mapman visualization tool, gene ontology analysis and inferred protein-protein interaction network. In summary, my meta-analysis study provides a better understanding of the *Malus x domestica* transcriptome responses to different biotic stress conditions; and I anticipate that these insights will assist in the development of genetic resistance and acute therapeutic

strategies. This work would be an example for next meta-analysis works aiming at identifying specific common molecular features linked with biotic stress responses in other speciality crops.

In my second project, my focus was to obtain normalized differentially expressed genes in *Malus* x *domestica* and report only the key genes, which are only regulated by biotic stress. To achieve my aim, I considered the following steps which helped me to increase the specificity of study, which were a) download row data from the literature for analysis b) use single bioinformatics pipeline for data analysis, c) use reference genome downloaded from single source (NCBI), and d) remove the genes which plays role in tree development and also affected by biotic stress. So, my main focus was to conduct raw data analysis by developing a bioinformatics pipeline by using the reference genome from a single source. Thus, I searched in literature, curated and manually collected 12 transcriptomic works in *Malus x domestica* in order to identify which key genes, proteins, gene categories are involved in general plant pathological conditions and those features linked with exclusive biotic stress responses. Those genes that are only related to molecular responses to pathogen attacks and linked with other plant physiological processes were identified. A pipeline composed by pathway and gene set enrichment analysis, protein-protein interaction networks and gene visualization tools were employed. This study represents a preliminary curated meta-analysis of apple transcriptomic responses to biotic stresses.

After my second project, I got an opportunity to spend one year in abroad university in Prof.Abhaya M Dandekar's lab at University of California, Davis and continued my research. Prof. Dandekar's research was mainly focused on understanding the effects of Huanglongbing (HLB) disease in *Citrus sinensis* and published several articles related to this topic. So, in my third project, I decided to investigate the biotic stress response in *Citrus sinensis* by using the meta-analysis pipeline, which was developed in my second study. In this study, my main attention was to identify genes commonly modulated between studies and genes, pathways and gene set categories strongly associated with the Huanglongbing (HLB) disease in *Citrus sinensis*. Bioinformatic analysis of previously published RNA-Seq studies on HLB response and tolerance in *Citrus sinensis* leaf tissues was performed. The expression data of four datasets present in NCBI were analyzed using a single transcriptome analysis pipeline, following with Gene set enrichment analysis and protein–protein interaction (PPI) to identify the different gene categories affected by HLB disease. In addition,

I updated my pipeline to report the alternative splicing events like exon skipping, intron retention, alternative donor and acceptor splice sites, which aid the investigation on correlations between differences in AS patterns and functional/structural features of genes due to the pathogen attack.

Experiment 1

Transcriptomic responses to biotic stresses in *Malus x domestica*: a meta-analysis study.

DOI : https://doi.org/10.1038/s41598-018-19348-4

1. Introduction

Apple (*Malus x domestica Borkh*) is one of the most important cultivated tree fruit crops in temperate climates. It is an important source of energy, vitamins and minerals in human diet. Unfortunately, this crop is severely affected by diseases mainly caused by fungi (Yin et al., 2016; Zhu et al., 2017; Shin et al., 2016; Wang et al., 2019), bacteria (Kamber et al., 2016; Silva et al., 2019; Singh et al., 2019) and viruses (Chen et al., 2014) with a consequent drastic reduction in fruit quantity and quality that threatens grower's profit (Fig. 1.1).

The fire blight disease caused by *Erwinia amylovora* (*E. amylovora*) is a global invasive threat for apple and pear production which affects blossoms, fruits, shoots, and branches and under optimal conditions, it can destroy an entire orchard in a single growing season. Alternaria blotch disease of apple, caused by the *Alternaria alternata* apple pathotype (AAAP), is one of the most serious fungal diseases affecting apples globally, especially in East Asia (Zhu et al., 2017). This disease affects apple tree growth and production via the infection of leaves, young shoots, and fruits and leads to marked declines in tree vigor. Another major constraint of apple cultivation is the apple scab, a fungal disease caused by *Venturia inaequalis*, which can lead to important crop losses if not properly controlled. In East Asia and China, Marssonina apple blotch caused by the fungus *Marssonina coronaria* is one of the most prevalent apple diseases. Valsa canker caused by the necrotrophic ascomycete *Valsa mali* is a destructive disease on apple in eastern Asia. The pathogen causes extensive necrotic lesions on apple trunks, and even death to the whole tree. Apple replant disease (ARD) is caused by a complex of soilborne necrotrophic fungi (*Cylindrocarpon* and *Rhizoctonia*) and oomycetes (*Phytophthora* and *Pythium*), and at times it can be aggravated by the lesion nematode *Pratylenchus penetrans*. The virus infections such as Apple stem grooving virus (ASGV), Apple chlorotic leaf spot virus (ACLSV) and Apple stem pitting virus (ASPV) usually do not induce visible disease symptoms in the infected trees and fruits, although the infection eventually does lead to significant reduction in fruit yield and quality (Chen et al., 2014). Soil metagenomic study explored the associations of nematodes and microbes in Apple replant disease (ARD) (Kanfra et al., 2018). Many studies have recently

investigated host global gene expression changes in plant-pathogen interactions to understand the molecular basis of various apple diseases. In response to stress, plant physiology and transcriptomes undergo changes in alarm, resistance, exhaustion, and regeneration phases. Since different tissues and developmental stages present different resistances to stress, transcriptome profiling of different tissues, strains, and developmental stages under various environmental stress conditions could provide insights into the molecular mechanisms as how plants respond to stress (Li et al., 2014).

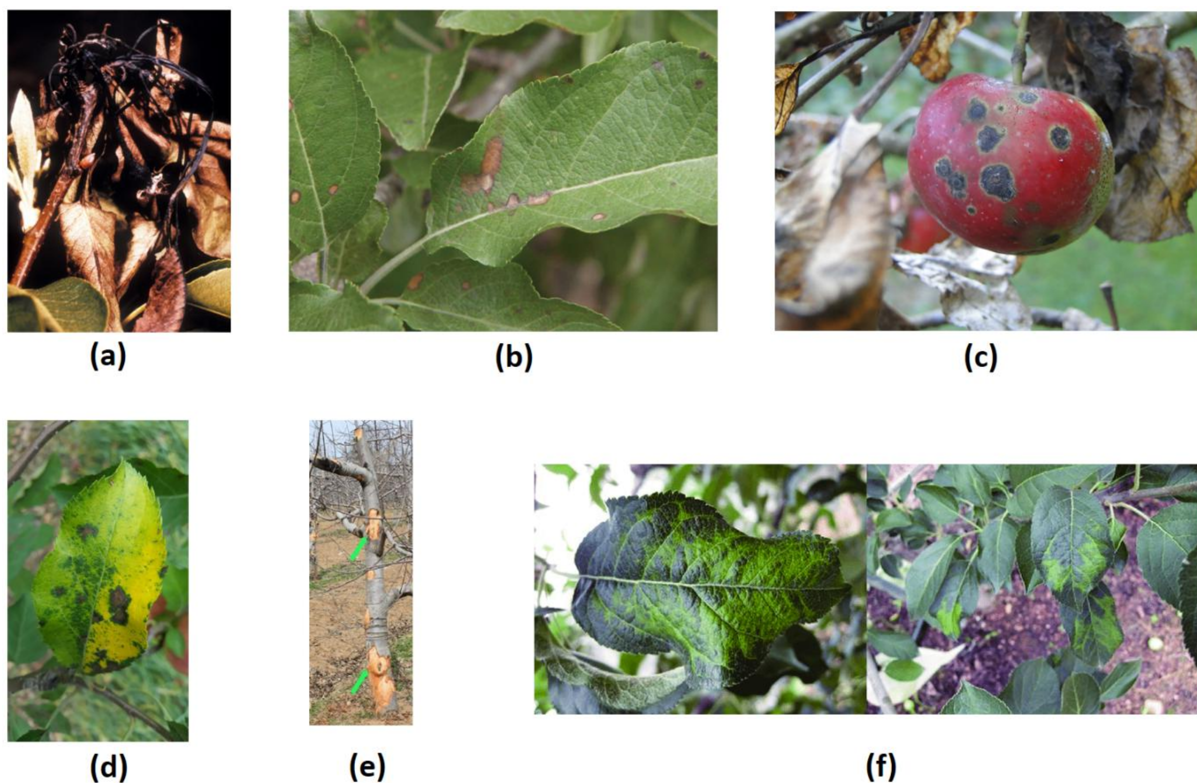


Figure 1.1– Apple diseases caused by fungi, virus and bacteria. (a) Apple leaves infected with fire blight, (b) Lesions caused by alternaria fungus on diseased apple leaf, (c) Apple scab disease on apples fruits, (d) *Marssonina* leaf blotch spots and leaf yellowing on apple leaf, (e) Damage to apple caused by *V. mali*. Green arrows indicate the position of canker lesions that are stripped artificially, (f) Apple cultivar 'Hongro' mixed infection with *Apple stem pitting virus*, *Apple chlorotic leafspot virus*, and *Apple stem grooving virus*, showing chlorosis along the leaf veins.

Plant diseases enforce significant crop losses in agriculture, horticulture and forestry. Genetic resistances to pathogens represent a large proportion of traits required by breeding programs. Plants have evolved sophisticated resistance mechanisms to pathogens (fungal, bacterial, viral or nematodes) which can be responsible for heavy crop losses (Gallois et al., 2018). Once pathogens overcome mechanical barriers to infection, plant receptors initiate signaling pathways driving the expression of defense response genes. Plant immune systems rely on their ability to recognize enemy molecules, execute cell signaling, and respond defensively through pathways involving many genes and their products. In summary, plant resistant mechanism can be classified as resistance associated with pathogen recognition, followed by defense induction and resistance by the loss-of-susceptibility affecting plant factors on which the pathogen relies for infection. Identification of resistance genes is useful in gene transformation as well as benefit in marker-assisted selection for introgression in conventional plant breeding. Quantitative trait locus (QTL) mapping is a highly effective approach for analyzing genetically complex forms of plant disease resistance (Brekketet et al., 2019; Shen et al., 2019). QTL can decrease pathogen traits related to infection efficiency or can modulate the efficiency of major-effect resistance genes (Pilet-Nayel et al., 2017; Nelson et al., 2018).

Plant resistance genes (R) defend against an invading pathogen by detecting the corresponding pathogen avirulence factors (Avr), which are often secreted effector proteins. These mechanisms are known as Effector Triggered Immunity or ETI (Jones and Dangl 2006) which functions most often in the plant cell cytoplasm, either, acting directly by detecting pathogen virulence-factors called effectors, or acting indirectly by monitoring host proteins that have been altered by effector activity. The resistance process is mediated by diverse group of mostly intracellular R proteins that are encoded by a few to hundreds of R genes that are present typically in clusters in every plant genome. The Nucleotide Binding Domains and Leucine-Rich Repeats (NB-LRR) factors are encoded by large families of genes and are associated with resistance to all kinds of pathogens and pests affecting plants (oomycetes, fungi, bacteria, insects, and nematodes, etc.) (Jones and Dangl 2006; de Ronde et al., 2014).

Modern biotechnology tools, such as tissue culture and genetic engineering, offer an alternative to conventional breeding in order to generate new cultivars with enhanced agronomic and nutritional characteristics (Sabbadini et al., 2019). In recent years, sequence-specific genome editing technologies were found to be useful tools for crop improvement and

clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein9 (Cas9) (Bhaya et al., 2011) is the newest and most widely used genome editing technology for the study of the function of genes and for the development of mutant lines with enhanced tolerance to biotic and abiotic stresses, herbicide resistance or improved yield. In the last decades, transgenic crops have been developed and genetic modifcation has been performed to confer resistance against insects, bacteria, virus and fungi diseases (Limera et al., 2017; Moradpour et al., 2020). The CRISPR/Cas9-mediated genome editing technology has opened a new opportunity for rapid development of disease resistant crop varieties by either stacking of disease resistant (R) gene(s) or disruption/deletion of susceptibility genes (Rojas-Vásquez et al., 2020; Haque et al., 2018;).

In plant-pathogen interactions, epigenetic mechanisms has gained interest during the last years. Recent studies link DNA methylation and demethylation as well as chromatin remodeling by posttranslational histone modifications, including acetylation, methylation, and ubiquitination, to changes in the expression levels of defense genes upon pathogen challenge (Marone et al., 2013; Zhu et al., 2016). Genome-wide analysis (GWA) of changes in host plant DNA methylation and histone modification linked with alterations in gene expression during several plant-microbe interactions is likely to provide a better understanding of epigenetic regulation in plant defense. Similar analyses of pathogen genomes and transcriptomes will illuminate epigenetic modifications that assist pathogen virulence and environmental adaptation (Zhu et al., 2016). Recent advanements in high-throughput sequencing technologies make such analyses now possible.

RNA sequencing (RNA-Seq) is a revolutionary tool that has been used extensively for the discovery of plant biotic stress genes, molecular patterns that are consistently associated with pathogen infection, molecular signal that if early detected can help to speed the diagnosis. . RNA-Seq can quantify gene/isoform expression levels at a higher resolution than microarray technology and provide coding-transcript profiling as well as long noncoding RNA (lncRNA) profiling. Initial gene expression studies relied on low-throughput methods, such as northern blots and quantitative polymerase chain reaction (qPCR) that are limited to measuring single transcripts. Over the last two decades, methods have evolved to enable genome-wide quantification of gene expression, or better known as transcriptomics. The first transcriptomics studies were performed using hybridization-based microarray technologies, which provide a high throughput option at relatively low cost. However, these methods have

several limitations: the requirement for a priori knowledge of the sequences being interrogated; problematic cross-hybridization artifacts in the analysis of highly similar sequences; and limited ability to accurately quantify lowly expressed and very highly expressed genes (Kukurba et al. 2015). The development of high-throughput next-generation sequencing (NGS) has revolutionized transcriptomics by enabling RNA analysis through the sequencing of complementary DNA (cDNA). This method, termed RNA sequencing (RNA-Seq), has distinct advantages over previous approaches and has revolutionized our understanding of the complex and dynamic nature of the transcriptome. RNA-Seq provides a more detailed and quantitative view of gene expression, alternative splicing, and allele-specific expression. Recent advances in the RNA-Seq workflow, from sample preparation to sequencing platforms to bioinformatic data analysis, has enabled deep profiling of the transcriptome and the opportunity to elucidate different physiological and pathological conditions (Kukurba et al. 2015; Griffith et al., 2015). In a typical RNA-Seq experiment, a sample of RNA is converted to a library of cDNA fragments and then sequenced on a high-throughput commercially available platform, such as Illumina' s Genome Analyzer, Helicos BioSciences ' HeliScope, Applied Biosystems ' SOLiD, Pacifi c Biosciences ' SMRT or Roche ' s 454 Life Sciences sequencing systems. The RNA-seq method typically consists of identification of suitable biological samples (and replicates), isolation of total RNA, enrichment of nonribosomal RNAs, conversion of RNA to cDNA, construction of a fragment library, sequencing on a high-throughput sequencing platform, generation of single or paired-end reads of 30–300 base pairs in length, alignment or assembly of these reads, and downstream analysis.

The first step in transcriptome sequencing is the isolation of RNA from a biological sample. To ensure a successful RNA-Seq experiment, the RNA should be of sufficient quality to produce a library for sequencing where the quality of RNA is typically measured using an Agilent Bioanalyzer, which produces an RNA Integrity Number (RIN) between 1 and 10 with 10 being the highest quality samples showing the least degradation (Kukurba et al. 2015). Following RNA isolation, the next step in transcriptome sequencing is the creation of an RNA-Seq library, which can vary by the selection of RNA species and between Next Generation Sequencing (NGS) platforms. The construction of sequencing libraries principally involves isolating the desired RNA molecules, reverse-transcribing the RNA to cDNA, fragmenting or amplifying randomly primed cDNA molecules, and ligating sequencing

adaptors. The efficient removal of rRNA is critical for successful transcriptome profiling, many protocols focus on enriching for mRNA molecules before library construction by selecting for polyadenylated (poly-A) RNAs. In this approach, the 3′ poly-A tail of mRNA molecules is targeted using poly-T oligos that are covalently attached to a given substrate (e.g., magnetic beads). Alternatively, researchers can selectively deplete rRNA using commercially available kits, such as RiboMinus (Life Technologies) or RiboZero (Epicentre). Universal to all RNA-Seq preparation methods is the conversion of RNA into cDNA because most sequencing technologies require DNA libraries. Most protocols for cDNA synthesis create libraries that were uniformly derived from each cDNA strand, thus representing the parent mRNA strand and its complement. In this conventional approach, the strand orientation of the original RNA is lost as the sequencing reads derived from each cDNA strand are indistinguishable in an effort to maximize efficiency of reverse transcription. Another consideration for constructing cost-effective RNA-Seq libraries is assaying multiple indexed samples in a single sequencing lane. The large number of reads that can be generated per sequencing run (e.g., a single lane of an Illumina HiSeq 2500 generates up to 750 million paired-end reads) permits the analysis of increasingly complex samples. The introduction of unique 6-bp indices, also known as "barcodes," to each RNA-Seq library enables the pooling and sequencing of multiple samples in the same sequencing reaction because the barcodes identify which sample the read originated from. The selection of a sequencing platform is important and dependent on the experimental goals. Currently, several NGS platforms are commercially available and other platforms are under active technological development. The majority of high-throughput sequencing platforms use a sequencing-by-synthesis method to sequence tens of millions of sequence clusters in parallel. In recent years, the sequencing industry has been dominated by Illumina, which applies an ensemble-based (i.e. sequencing many identical copies of a DNA molecule) sequencing-by-synthesis approach.

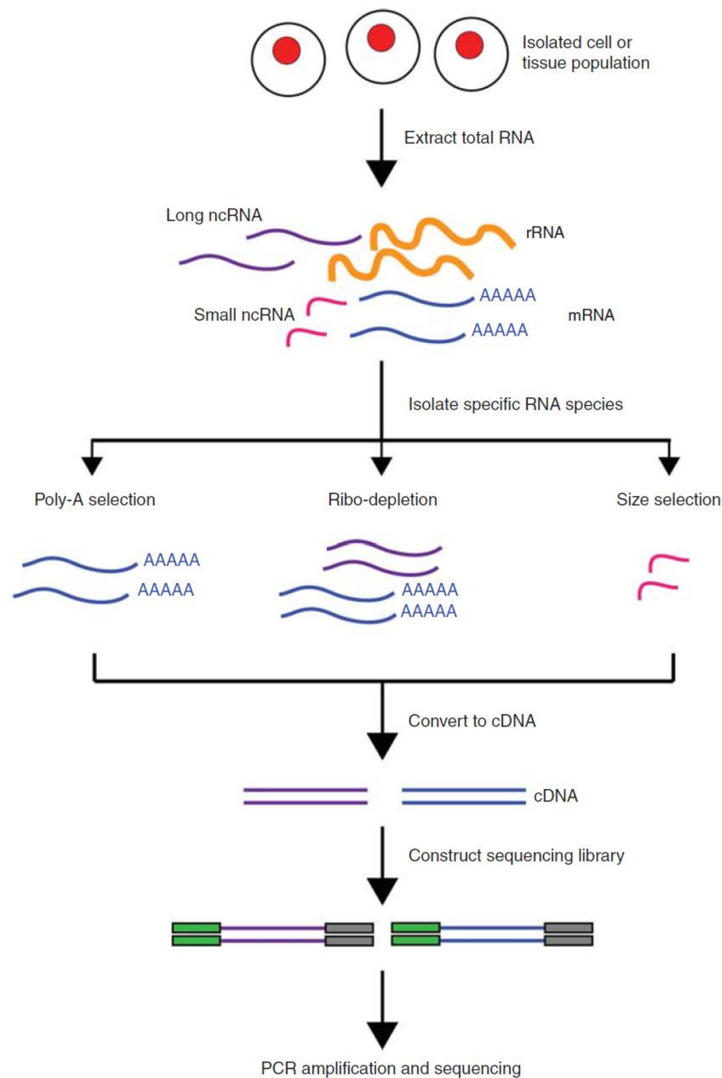


Figure 1.2– Overview of RNASeq library preparation (Kukurba et al. 2015)

The conventional bioinformatics pipeline for RNA-Seq data includes generating FASTQ-format files contains reads sequenced from an NGS platform, aligning these reads to an annotated reference genome, and quantifying expression of genes. The initial step of RNA-Seq data analysis pipeline is the quality check of the raw sequence data output from sequencing. FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) is a tool, which can process FASTQ files and summarize the Quality reports of the reads in figures and tables. The main parameters need to be checked during the Qualify check are a) Base quality score distribution b) Sequence quality score distribution c) Average base content per read d) GC distribution in the reads e) PCR amplification issue and f) Over-represented sequences.

According to the quality report from FASTQC, the low quality bases from the reads will be removed. Average Q30 (Phred value) score was used as a cutoff to remove low quality bases. Also, if there any specific bias observed in base composition, those bases can also be trimmed. Another recurring problem in Illumina sequencing is adapter contamination in reads, were adapters can be present partially or completely within the read. There are a plenty of bioinformatics tools available for the adapter removal from fastq reads where the popular tools are Cutadapt (Martin M., 2011) and Trimmomatic (Bolger et al., 2014). The pre-processed reads with read length >=30 bases can be considered for the mapping to reference genome, since the reads less than 30 bases doesn't make any influence in alignment process.

Mapping RNA-Seq reads to the reference genome is a very challenging task because many reads may map across the splice junctions and only a "splicing-aware" aligner can only recognize it and map properly. The more commonly used RNA-Seq alignment tools include GSNAP (http://research-pub.gene.com/gmap/), MapSplice (http://www.netlab.uky.edu/p/bioinfo/MapSplice), RUM (http://www.cbil.upenn.edu/RUM/), STAR (https://github.com/alexdobin/STAR), TopHat (https://ccb.jhu.edu/software/tophat/index.shtml) and HISAT (https://ccb.jhu.edu/software/hisat2/index.shtml). Each aligner has different advantages in terms of performance, speed, and memory utilization. Selecting the best aligner to use depends on these metrics and the overall objectives of the RNA-Seq study. A reference genome with sequences derived from exon–exon splice junctions acquired from known gene annotations is required for the mapping programs, which can be downloaded from various sources like National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/), Ensembl (http://ensemblgenomes.org/), Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) etc. Some tools require to generate index files using the fasta file of the genome, which helps to speed up the read mapping. After alignment, the SAM (Sequence Alignment/Map format file is a TAB-delimited text format consisting of a header section, which is optional, and an alignment section) files of each samples will be generated. The SAM files can be converted to BAM (binary format of SAM, and were using for most of the downstream analysis) files using the tool samtools (http://samtools.sourceforge.net/). The R package CummeRbund (https://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html) has the possibility to represent the transcripts (RNA-Seq reads) mapped to the reference genome

(together with transcript abundances). The Integrative Genomics Viewer (IGV; https://www.broadinstitute.org/igv/) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets, which helps to zoom in to the region of interest in the gene/transctipt. The complete bioinformatics workflow for the RNA-Seq analysis is given in Fig. 1.3.

After RNA-Seq reads are aligned, the mapped reads can be assembled into transcripts. The majority of computational programs infer transcript models from the accumulation of read alignments to the reference genome. Computational tools such as Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/), FluxCapacitor (https://omictools.com/the-flux-capacitor-tool), and MISO (https://miso.readthedocs.io/en/fastmiso/), quantify expression by counting the number of reads that map to full-length transcripts. Alternative approaches, such as HTSeq (https://htseq.readthedocs.io/en/release_0.11.1/count.html), can quantify expression without assembling transcripts by counting the number of reads that map to an exon. To accurately estimate gene expression, read counts must be normalized to correct for systematic variability, such as library fragment size, sequence composition bias, and read depth. To account for these sources of variability, the reads per kilobase of transcripts per million mapped reads (RPKM) metric normalizes a transcript's read count by both the gene length and the total number of mapped reads in the sample. For paired end-reads, a metric that normalizes for sources of variances in transcript quantification is the paired fragments per kilobase of transcript per million mapped reads (FPKM) metric, which accounts for the dependency between paired-end reads in the RPKM estimate.

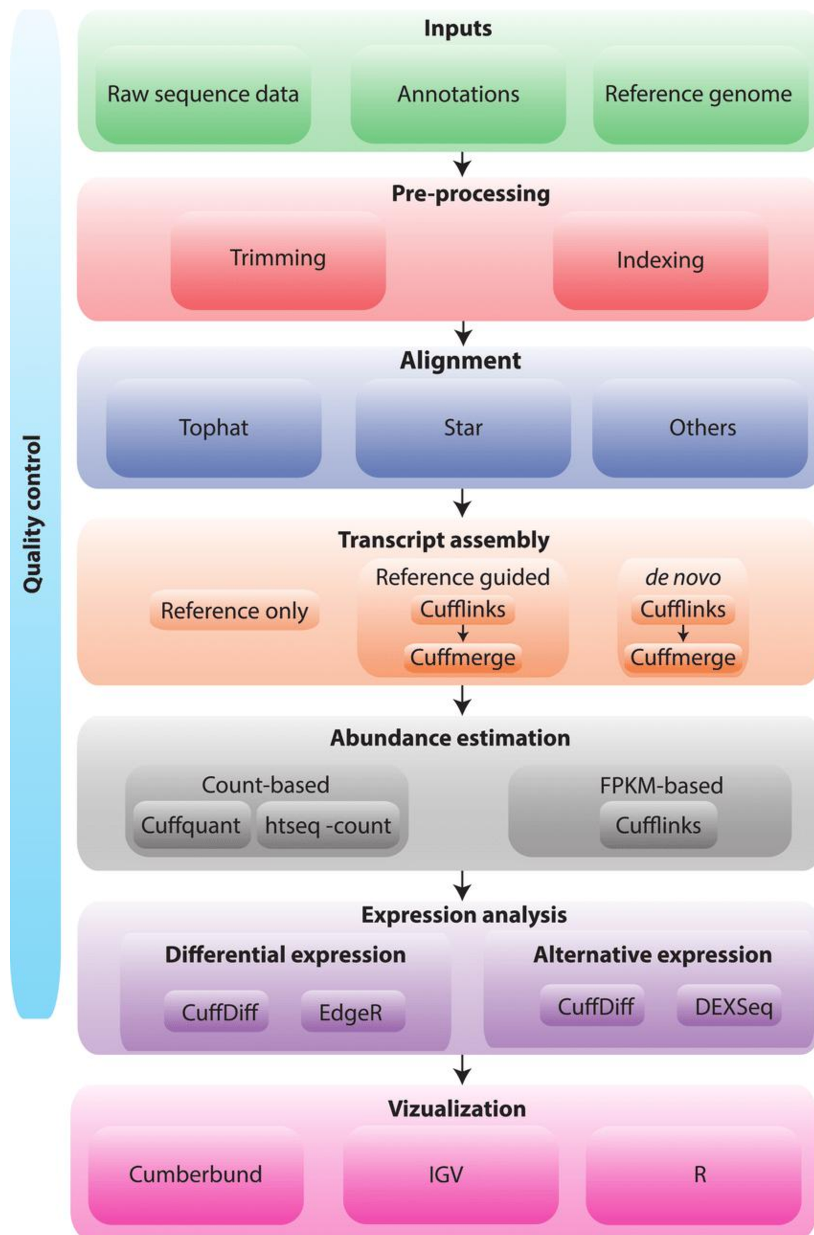


Figure 1.3– RNA-Seq analysis flow chart (Griffith et al., 2015).

To detect differential expression, a variety of statistical methods have been designed specifically for RNA-Seq data. A popular tool to detect differential expression is Cuffdiff (http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/), which is part of the Tuxedo suite of tools (Bowtie, Tophat, and Cufflinks) developed to analyze RNA-Seq data. In addition to Cuffdiff, several other packages support testing differential expression, including baySeq (https://omictools.com/bayseq-tool), DESeq (https://bioconductor.org/packages/release/bioc/html/DESeq.html), DEGseq

(https://bioconductor.org/packages/release/bioc/html/DEGseq.html), and edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html). Replicates in RNA-Seq experiments are crucial for measuring variability and improving estimations for the model parameters.

In recent years, RNA sequencing and analysis using Next Generation Sequencing (NGS) methods have enabled to understand the gene expression pertaining to plant biotic and abiotic stress conditions in both quantitative and qualitative manner (Martinelli et al., 2015; Muleo et al., 2016). In each study, the large quantity of obtained data makes very difficult the analysis and the identification of the role of each gene in the molecular networks. False positive results often occur due to the RNA-Seq method that needs validation with other quantitative gene expression methods. In addition, there are some genes, which can be expressed in any physiological condition, which makes the conclusions often weak. The large number of transcriptomic works published in plants requires more meta-analysis studies that would identify common and specific features in relation of the high number of objective studies performed at different developmental and environmental conditions. This is due to several reasons. First, transcript amounts are highly affected by changing environmental conditions and a high number of variables such as timing, environmental factors and experimental conditions, tissues and their developmental stages, genotypes, finely modulates gene expression. Secondly, transcriptomic studies are often performed only one time with no repetition. Field studies are usually conducted only in one season leading to unreliable results affected by a high number of environmental disturbing factors. Third, few replicates (frequently only three) are usually considered due to the high costs of "omic" analysis. More biological replicates would be really useful to reduce environmental confounding variability. Finally, transcriptomic studies should be integrated with proteomics and metabolomics performed on the same samples of the same study in order to clarify post-transcriptional and post-transductional regulation mechanisms.

Transcriptomic studies are usually conducted in a singular time, they do not provide any repetition across different seasons and frequently they are performed in field conditions where environmental variability is high and disturbing factors are frequently present. The identification of up- or down-regulated genes is often not enough to draw meaningful biological conclusions because it is hard to identify which gene plays a key role in specific signaling networks in host responses (Yin et al., 2016). This issue leads to high difficulties in

deriving conclusive models for understanding disease symptomatology. For these reasons, more meta-analysis is needed in order to validate singular transcriptomic works with other similar studies performed with the same research purposes. A Meta-analysis of transcriptomic data will identify commonalities and differences between differentially regulated gene lists and will allow screen which genes are key players in gene-gene and protein-protein interaction networks. These analyses will allow delivering important information on how a specific environmental factor affects plant molecular responses and how plants activate general stress responses to environmental stresses (Rest et al., 2016; Cohen et al., 2019). An early "stress condition" in plants is similar to the "inflammatory response" occurring in animals in response to pathogen-associated factors. The identification of common genes between different biotic stress will allow to gain insight into these general responses and help the diagnosis of an early "stress state" of the plants. These analyses help in monitoring stressed plants to start early specific management procedures for each disease or disorder. The activation of common responses to different biotic stresses may precede the onset of symptoms, where more physiological changes lead to specific phenotypic changes and peculiar metabolic dysfunctions (Gambino et al., 2012; Dandekar et al., 2010). Indeed, there is a strong need for compelling cases in order to generalize results across studies performed in the same crop and determine the most reliable and meaningful information linked with agronomic factors such as biotic stress responses.

In this meta-analysis study, I considered all transcriptomic data related to biotic stresses in *Malus x domestica*, which are already published. The aim to determine which genes, pathways, gene set categories and predicted protein-protein interaction networks may play key roles in specific responses to pathogen infections.

2. Materials and methods

2.1 Search strategy of published study identification and selection for meta-analysis

As a first step, all published transcriptomic studies in *Malus x domestica* were searched and collected from Scopus (https://www.scopus.com/search/form.uri?display=basic) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) before March 2017, using the combination of keywords 'Transcriptomics" and "malus" or 'Transcriptomics" and "apple" in computer-

based searches. The RNA-Seq studies pertaining to biotic stress on *Malus x domestica* were selected and classified into three groups a) Fungal pathogens (Yin et al., 2016; Zhu et al., 2017; Shin et al., 2016; Gusberti et al., 2013; Xu et al., 2015), b) ASGV (Chen et al., 2014) and c) *E. amylovora* (Kamber et al., 2016) based on the pathogen type. The list of differentially regulated genes, obtained from the selected seven published transcriptomic articles in *Malus x domestica* were given in Table 2.1. Only genes reported in the main text and supplementary files of these articles were considered in this meta-analysis.

Article	Objective	Pathogen Species	Pathogen	Tissue	DEGs		Group
					Up	Down	
Gusberti et al., 2013	Resistance to Venturia	*Venturia inaequalis*	Fungi	Leaf	112	100	Fungal Pathogen
Yin et al., 2016	Resistance to *Valsa mali*	*Valsa mali*	Fungi	Twig	14	15	
Xu et al., 2015	Response to *Marssonina coronaria* inoculation	*Marssonina coronaria*	Fungi	Leaf	58	32	
Zhu et al., 2017	Response to *Alternaria alternata*	*Alternaria alternate*	Fungi	Leaf	2,108	1,746	
Shin et al., 2016	Response to *Pythium ultimum*	*Pythium ultimum*	Fungi	Root	63	19	
Chen et al., 2014	*Apple stem grooving virus*	*Apple stem grooving virus*	Virus	Shoot	184	136	ASGV
Kamber et al., 2016	Responses to *Erwinia amylovora*	*Erwinia amylovora*	Bacteria	Flower	640	183	*E. amylovora*

Table2.1 - Transcriptomic studies dealing with biotic stress responses in *Malus x domestica* used for meta-analysis. Number of up-regulated and down-regulated genes was indicated for each study.

2.2 Extraction and annotation of differentially expressed genes

The up- and down-regulated genes and the fold change information were extracted from the supplementary tables of the articles. The genes with fold change and p-value cutoffs (log2 FC > 1 or log2 FC < −1; p-value < 0.05) were only selected in order to strengthen the accuracy of the analysis and normalization; except one article where the fold change is not given (Yin et al., 2016). All the *Malus x domestica* gene ids were based on the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) and were mapped to the corresponding *Arabidopsis* id, using the annotation file downloaded from Phytozome. The data extraction and mapping were done by in-house Perl scripts. During the analysis, the 5 fungi datasets were merged in to one single file in order to analyze the entire list of fungal pathogen-regulated genes in *Malus x domestica*. This operation was not needed for viral and bacterial responses since only one transcriptomic study was available for both types of pathogens.

2.3 Gene enrichment analysis

The metabolic overview, hormone regulation, large enzyme families, transcription factors and biotic stress gene categories of the three groups were visualized using MapMan (Thimm et al., 2004) with the *Malus x domestica* mapping file downloaded from MapMan web site (http://mapman.gabipd.org/). The PageMan (Usadel et al., 2006) analysis plugin of MapMan was used to visualize differences among metabolic pathways using Wilcoxon tests, no correction, and an over-representation analysis (ORA) cutoff value of 3.

2.4 Functional analysis

Gene ontology analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Web server (https://david.ncifcrf.gov/) (Huang et al., 2009), based on the homologous TAIR IDs. The gene ontology information of each group was extracted from the DAVID results using in-house Perl script. The top biological process, cellular component and molecular function in each biotic stresses were given in Table 2.2-2.8.

GO ID	GO Term	Count	pval	Fold Enrichment	FDR
GO:0046686	response to cadmium ion	48	1.46E-06	2.121201697	2.39E-03
GO:0009651	response to salt stress	61	1.86E-06	1.904808445	3.03E-03
GO:0009414	response to water deprivation	47	7.98E-09	2.546012252	1.30E-05
GO:0006979	response to oxidative stress	42	3.61E-06	2.181338859	5.89E-03
GO:0009611	response to wounding	35	2.95E-07	2.658164187	4.83E-04
GO:0009738	abscisic acid-activated signaling pathway	31	1.28E-05	2.402668845	2.09E-02
GO:0010200	response to chitin	25	6.06E-06	2.84089513	0.009902214

Table 2.2–Top up-regulated GO-terms involved in biological process due to fungal pathogens.

GO ID	GO Term	Count	pval	Fold Enrichment	FDR
GO:0055114	oxidation-reduction process	130	1.40E-12	1.891394281	2.22E-09
GO:0015979	Photosynthesis	56	7.69E-35	7.852343703	1.22E-31
GO:0009409	response to cold	43	3.46E-09	2.782836092	5.50E-06
GO:0009735	response to cytokinin	42	8.22E-16	4.441079635	1.23E-12
GO:0009658	chloroplast organization	36	5.97E-17	5.572920502	1.78E-13
GO:0009416	response to light stimulus	30	1.22E-07	3.087832725	1.93E-04
GO:0045454	cell redox homeostasis	28	2.48E-08	3.495559455	3.95E-05
GO:0015995	chlorophyll biosynthetic process	25	1.36E-20	11.5181062	2.16E-17
GO:0006633	fatty acid biosynthetic process	23	6.12E-07	3.477028308	9.73E-04
GO:0034599	cellular response to oxidative stress	17	1.06E-07	5.13995489	1.68E-04

Table 2.3–Top down-regulated GO-terms involved in biological process due to fungal pathogens.

Group	Up/Down	GO ID	GO Term	Count	pval	Fold Enrichment	FDR
ASGV	Down	GO:0009611	response to wounding	12	1.39E-09	13.27997128	1.74E-06
	Up	GO:0010200	response to chitin	9	5.71E-07	12.5181203	7.15E-04
F. amylovora	Up	GO:0009611	response to wounding	17	7.30E-07	4.703323163	0.00107614
		GO:0009753	response to jasmonic acid	13	2.79E-05	4.588045635	0.041175655

Table 2.4–The GO-terms involved in cellular component due to ASGV and *E. amylovora*.

2.5 Protein-protein interaction network

The protein-protein interaction network (PPI) information based on both experiment (ppi(exp).v5.03) and integrated prediction based (ppi(pred).v5.03) were downloaded from AtPID (*Arabidopsis thaliana* Protein Interactome Database; http://www.megabionet.org/atpid/webfile/) (Li et al., 2011). The top 100 genes for each group were selected based on the PPI count and were considered for the PPI network analysis. PPI network was constructed based on the protein interaction information retrieved from STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, http://string-db.org/) (Szklarczyk et al., 2015), an online protein-protein interaction database curated from literature and predicted associations from systemic genome comparisons.

GO ID	GO Term	Count	Pval	Fold Enrichment	FDR
GO:0005737	Cytoplasm	406	3.92E-22	1.552745723	5.18E-19
GO:0005886	plasma membrane	307	2.59E-10	1.397717233	3.43E-07
GO:0005829	Cytosol	278	1.12E-31	2.029262462	1.47E-28
GO:0005774	vacuolar membrane	62	8.34E-07	1.949594354	1.10E-03
GO:0005730	Nucleolus	51	3.82E-06	2.008370186	5.04E-03
GO:0022626	cytosolic ribosome	29	3.01E-05	2.395991037	3.97E-02

Table 2.5–Top up-regulated GO-terms involved in cellular component due to fungal pathogens.

GO ID	GO Term	Count	Pval	Fold Enrichment	FDR
GO:0009507	chloroplast	601	2.70E-187	3.283461623	3.38E-184
GO:0009570	chloroplast stroma	213	3.84E-120	6.869866176	4.81E-117
GO:0009941	chloroplast envelope	196	1.56E-119	7.602179097	1.95E-116
GO:0009535	chloroplast thylakoid membrane	174	2.54E-120	9.004025039	3.18E-117
GO:0016020	membrane	153	1.47E-12	1.79318546	1.85E-09
GO:0009534	chloroplast thylakoid	123	2.95E-107	12.57534029	3.70E-104
GO:0009579	thylakoid	111	8.55E-90	11.34847783	1.07E-86
GO:0048046	apoplast	69	2.71E-15	2.941738042	3.34E-12
GO:0009536	plastid	44	1.38E-10	3.058383225	1.73E-07
GO:0010287	plastoglobule	39	1.69E-30	10.95179229	2.12E-27

Table 2.6– Top down-regulated GO-terms involved in cellular component due to fungal pathogens.

Group	Up/Down	GO ID	GO Term	Count	pval	Fold Enrichment	FDR
ASGV	Up	GO:0005634	nucleus	77	1.59E-05	1.475106349	0.0156192
E. amylovora	Down	GO:0048046	apoplast	17	3.00E-09	6.923060729	3.20E-06
		GO:0005618	cell wall	15	1.29E-06	5.132040816	1.37E-03
		GO:0009535	chloroplast thylakoid membrane	12	5.74E-06	5.931479115	6.14E-03
		GO:0009579	thylakoid	10	8.78E-07	9.765825243	9.38E-04
		GO:0009523	photosystem II	5	4.49E-05	25.147	4.79E-02
		GO:0009522	photosystem I	5	2.33E-05	29.58470588	2.49E-02

Table 2.7– Top regulated GO-terms involved in cellular component due to ASGV and *E. amylovora*.

Group	Up/Down	GO ID	GO Term	Count	pval	Fold Enrichment	FDR
Fungal Pathogens	Up	GO:0030170	pyridoxal phosphate binding	22	2.07E-05	2.869194496	0.032063107
		GO:0005515	protein binding	188	6.38E-08	1.457442683	9.86E-05
		GO:0046872	metal ion binding	125	9.22E-07	1.540638048	0.00140215
	Down	GO:0031409	pigment binding	11	2.16E-08	10.36834734	3.29E-05
		GO:0019843	rRNA binding	27	1.13E-08	3.737350884	1.72E-05
		GO:0016671	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	12	8.42E-06	5.398395722	0.012801046
		GO:0016491	oxidoreductase activity	46	2.17E-07	2.328719723	3.29E-04
		GO:0016168	chlorophyll binding	15	1.30E-09	8.024642289	1.97E-06
		GO:0008266	poly(U) RNA binding	8	2.40E-05	8.334365325	0.036556976
ASGV	Up	GO:0003700	transcription factor activity, sequence-specific DNA binding	25	2.94E-06	2.85486699	0.003286404
F. amylovora	Down	GO:0016168	chlorophyll binding	5	3.89E-05	25.84779516	0.045307927

Table 2.8– Top regulated GO-terms involved in cellular component due to the pathogens.

3. Results

3.1 Meta-analysis of transcriptome data

The list of the up- and down-regulated genes were finalized and compared these lists in order to identify common and different regulated genes between the 7 studied research works dealing with biotic stress responses. The data normalization was done using the same log Fold change and p-values (log2 FC > 1 or log2 FC < −1; p-value < 0.05).Venn diagrams showed the numbers of specific and commonly regulated genes between the three types of biotic stresses in *Malus x domestica* (Fig. 3.1.1). I observerd that 16 genes were commonly

regulated in responses to fungal pathogens, virus (*Apple Stem Grooving Virus*) and bacteria (*Erwinia amylovora*).

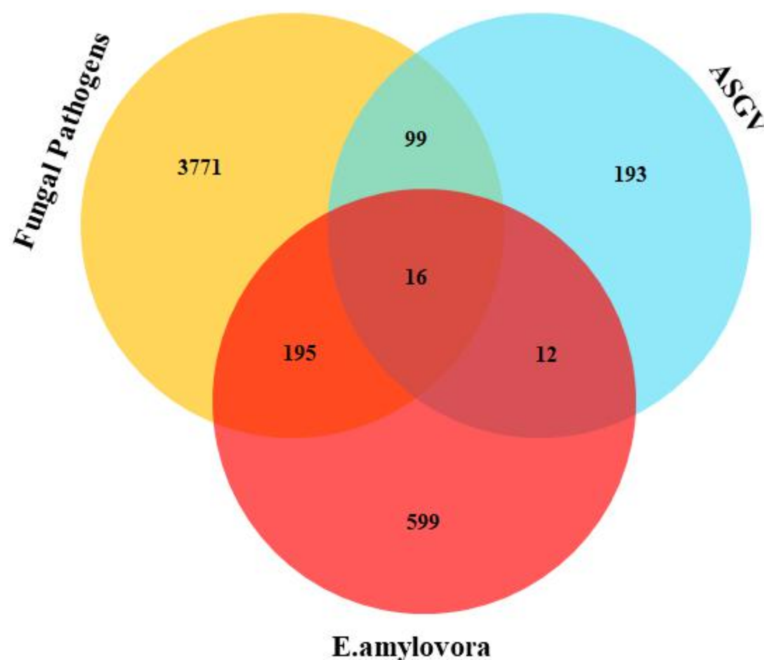


Figure 3.2.1–Distribution of up-regulated biological process terms in fungal pathogens

3.2 Gene ontology analysis

Gene Ontology enrichment analysis was conducted to explore other possible functions of the differentially expressed genes in different biotic stress conditions. Pie charts showing the distribution of up regulated and down regulated GO-terms in biological processes for each of the three types of biotic stresses were generated (Fig. 3.2.1 - 3.2.6). It is clear that *Apple Stem Grooving Virus* (ASGV) upregulated a higher percentage of GO-terms related to transcription regulation, response to chitin and phosphorylation. Percentage of up regulated GO-terms related to ethylene and jasmonic acid defense responses were higher in response to *Erwinia amylovora* (*E. amylovora*) than to fungal pathogens. Oxidation- reduction pathways were strongly repressed in response to fungal pathogens. In the enrichment analysis, strong differences were observed between the three types of stresses in relation to repressed GO-terms. While fungal pathogens inhibited hormone-related genes, photosynthesis, responses to abiotic stresses, *E. amylovora* reduced responses to xyloglucan metabolic process, actin,

cellular responses to gravity and lipid transport. Virus infection specifically repressed jasmonic acid-related GO terms, response to the bacterium and fungal pathogens.

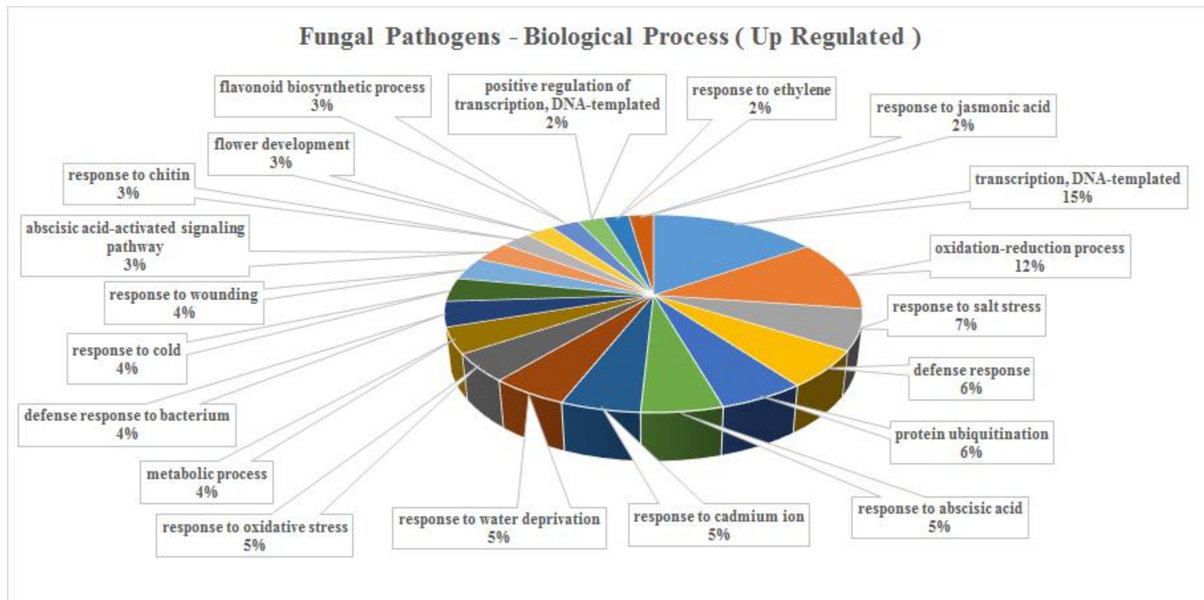


Figure 3.2.1–Distribution of up-regulated biological process terms in fungal pathogens

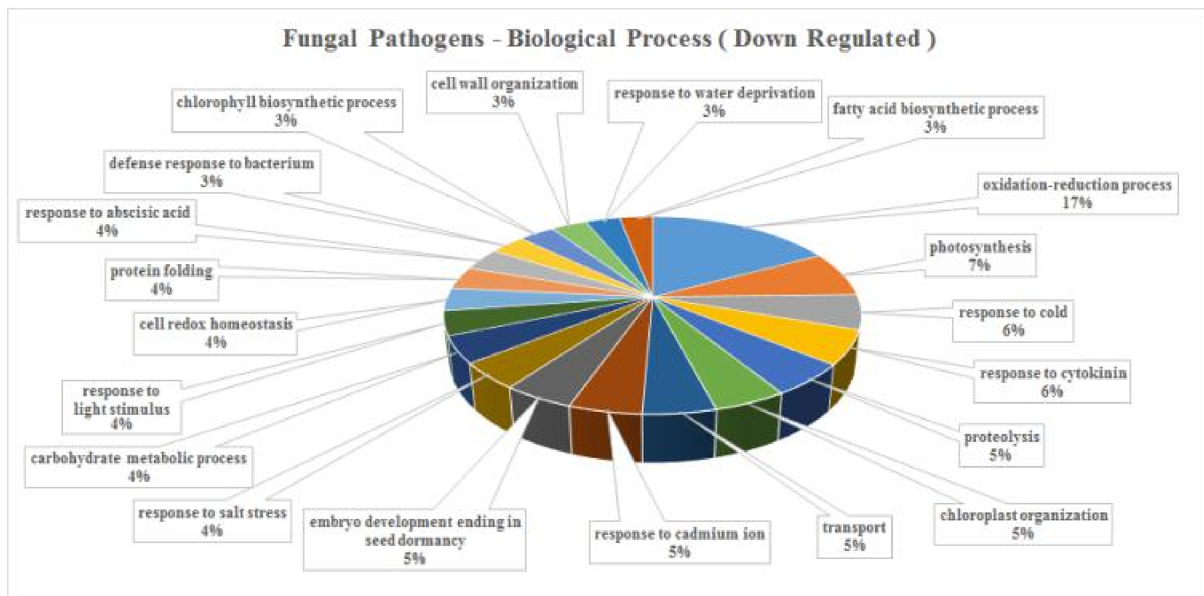


Figure 3.2.2–Distribution of down-regulated biological process terms in fungal pathogens

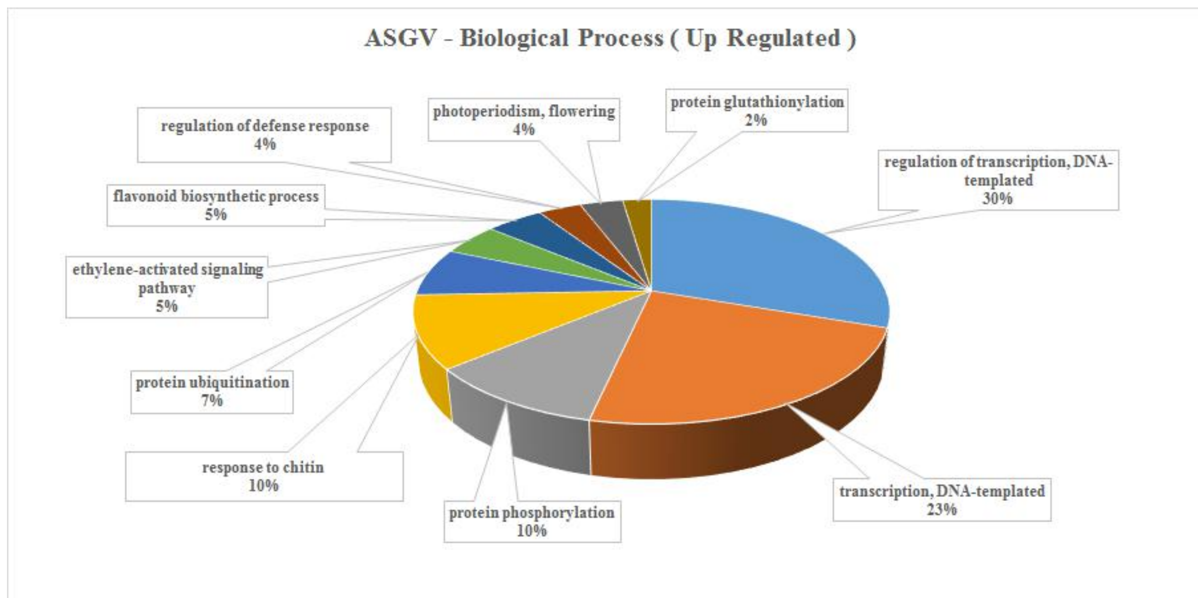


Figure 3.2.3–Distribution of up-regulated biological process terms in ASGV

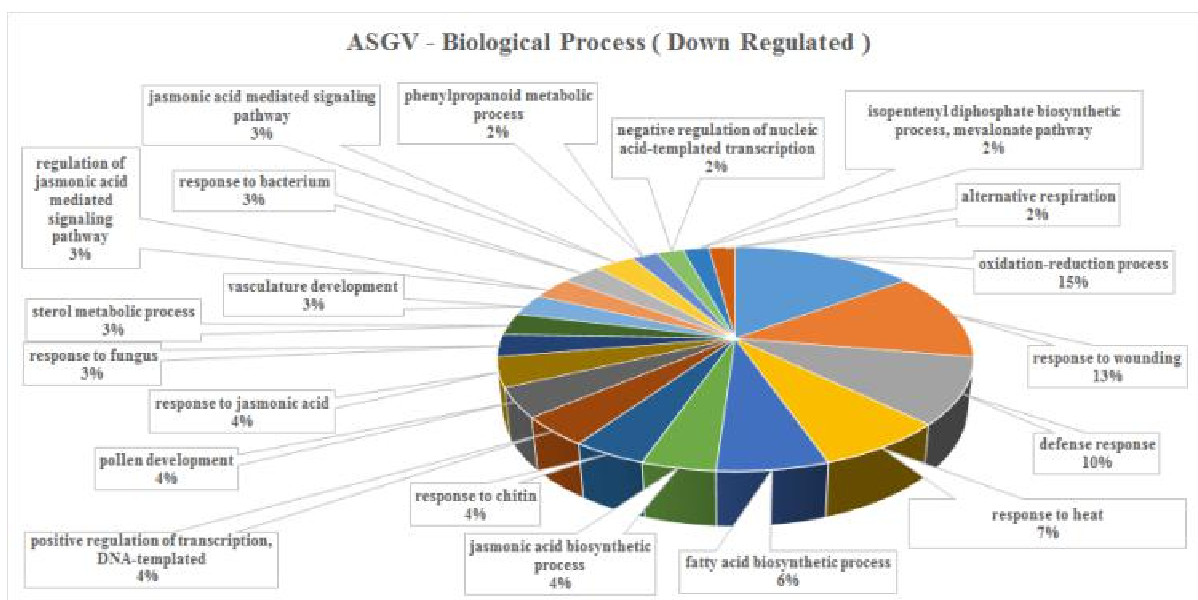


Figure 3.2.4–Distribution of down-regulated biological process terms in ASGV

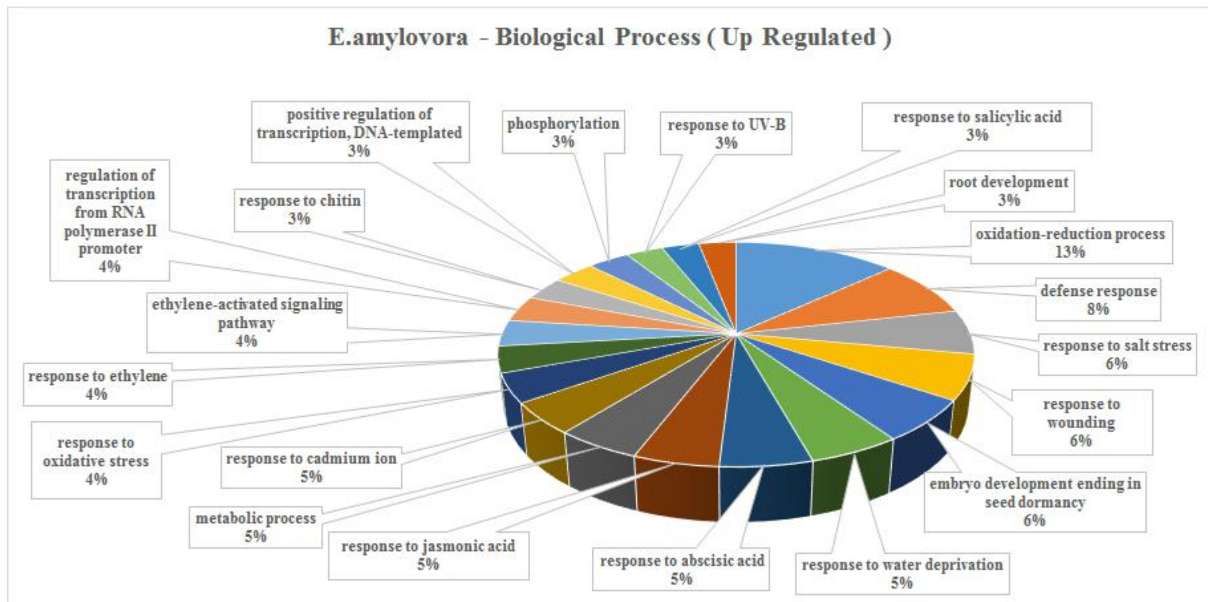


Figure 3.2.5–Distribution of up-regulated biological process terms in *E. amylovora*

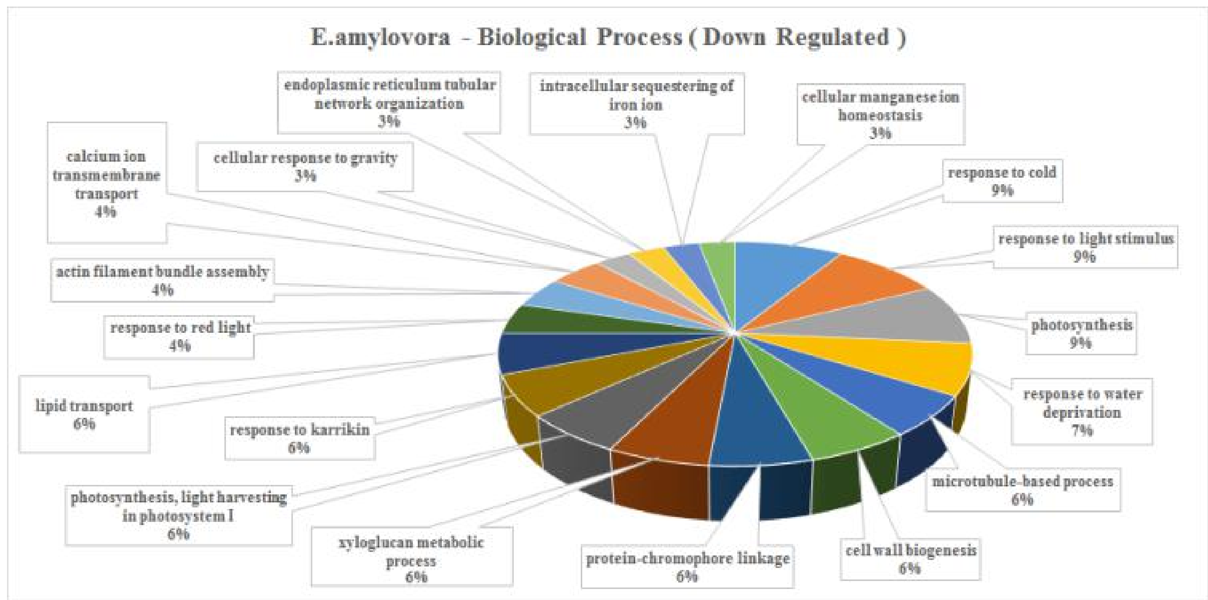


Figure 3.2.6–Distribution of down-regulated biological process terms in *E. amylovora*

3.3 Gene set enrichment analysis

Gene enrichment analysis was carried out using pageman (Usadel et al., 2006) to identify any relationship between the expression and function of differentially expressed genes in different biotic stress conditions (Fig. 3.3.1). As expected, *E. amylovora* and fungal pathogens repressed photosynthesis-related genes such as those involved in photosystem II. Adenylpyrophosphatase (ATPase), photorespiration, calvin cycle and major CHO metabolism were significantly inhibited by fungal pathogens while genes encoding electron carriers were repressed by *E. amylovora*. Cell wall genes were down-regulated in all the three datasets. Fungal pathogens up-regulated several gene set categories involved in both primary and secondary metabolism including amino acids (glutamate, aromatic ones), flavonoids and isoprenoid mevalonate pathway. PR-proteins and other stress-related proteins were up regulated by fungal pathogens.

Relating to hormones, brassinosteroids were induced by fungal pathogens while *E. amylovora* enhanced ethylene gene set category. Jasmonate was repressed by fungal and viral infections. Most of the transcription factors were induced by fungal pathogens such as *AP2-EREBP*, *WRKY*s and *ARR*s. Ubiquitin-mediated degradation was mainly up regulated by fungal pathogens. Different gene transport-related categories were repressed by different types of pathogens: transporters in envelope membrane by fungal pathogens while metal transporters were down-regulated by *E. amylovora*. Overall, the gene set enrichment analysis identified additional responses due to various biotic stresses in *Malus x domestica*.

E.amylovora
ASGV
Fungal Pathogens

-3
-2
-1
0
1
2
3

PS
cell wall
lipid metabolism
amino acid metabolism
metal handling
secondary metabolism
tetrapyrrole synthesis
stress
redox
polyamine metabolism
misc
RNA
DNA
signalling
transport

PS
PS.lightreaction
PS.lightreaction.photosystem II
PS.lightreaction.photosystem II.LHC-II
PS.lightreaction.photosystem II.PSII polypeptide subunits
PS.lightreaction.photosystem I
PS.lightreaction.photosystem I.PSI polypeptide subunits
PS.lightreaction.ATP synthase
PS.lightreaction.other electron carrier (ox/red)
PS.lightreaction.other electron carrier (ox/red).ferredoxin
PS.lightreaction.cyclic electron flow-chlororespiration
PS.photorespiration
PS.calvin cycle
PS.calvin cycle.GAP
PS.calvin cycle.FBPase
PS.calvin cycle.seduheptulose bisphosphatase
PS.calvin cycle.rubisco interacting
major CHO metabolism.synthesis
TCA / org transformation.carbonic anhydrases
cell wall
cell wall.degradation.pectate lyases and polygalacturonases
cell wall.modification
lipid metabolism
lipid metabolism.FA synthesis and FA elongation
lipid metabolism.FA synthesis and FA elongation.Acetyl CoA Carboxylation
lipid metabolism.lipid transfer proteins etc
amino acid metabolism
amino acid metabolism.synthesis
amino acid metabolism.synthesis.aromatic aa
amino acid metabolism.degradation
amino acid metabolism.degradation.glutamate family
amino acid metabolism.degradation.branched chain group
amino acid metabolism.degradation.serine-glycine-cysteine group
amino acid metabolism.degradation.serine-glycine-cysteine group.glycine
amino acid metabolism.degradation.aromatic aa
amino acid metabolism.degradation.aromatic aa.tyrosine
metal handling
secondary metabolism
secondary metabolism.isoprenoids.non-mevalonate pathway
secondary metabolism.isoprenoids.mevalonate pathway
secondary metabolism.isoprenoids.carotenoids
secondary metabolism.flavonoids.dihydroflavonols
secondary metabolism.flavonoids.dihydroflavonols.dihydroflavonol 4-reductase
hormone metabolism.brassinosteroid.synthesis-degradation.sterols.other
hormone metabolism.ethylene
hormone metabolism.gibberelin.induced-regulated-responsive-activated
hormone metabolism.jasmonate
hormone metabolism.jasmonate.synthesis-degradation
tetrapyrrole synthesis
tetrapyrrole synthesis.magnesium chelatase
stress
stress.biotic
stress.biotic.PR-proteins
stress.abiotic
stress.abiotic.heat
stress.abiotic.unspecified
redox
redox.thioredoxin
polyamine metabolism
polyamine metabolism.synthesis
polyamine metabolism.synthesis.SAM decarboxylase
misc
misc.UDP glucosyl and glucoronyl transferases
misc.gluco-, galacto- and mannosidases
misc.oxidases - copper, flavone etc
misc.myrosinases-lectin-jacalin
misc.plastocyanin-like
misc.protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
misc.rhodanese
misc. other Ferredoxins and Rieske domain
RNA
RNA.processing
RNA.processing.splicing
RNA.regulation of transcription
RNA.regulation of transcription.AP2/EREBP, APETALA2/Ethylene-responsive element binding protein family
RNA.regulation of transcription.bHLH,Basic Helix-Loop-Helix family
RNA.regulation of transcription.C2C2(Zn) CO-like, Constans-like zinc finger family
RNA.regulation of transcription.WRKY domain transcription factor family
RNA.regulation of transcription.Global transcription factor group
RNA.regulation of transcription.Psudo ARR transcription factor family
RNA.RNA binding
DNA
DNA.synthesis/chromatin structure.histone
DNA.unspecified
signalling
signalling.receptor kinases
signalling.receptor kinases.leucine rich repeat XI
signalling.G-proteins
cell.vesicle transport
transport
transport.metabolite transporters at the envelope membrane
transport.metabolite transporters at the mitochondrial membrane
transport.metal
transport.misc
not assigned.no ontology.pentatricopeptide (PPR) repeat-containing protein

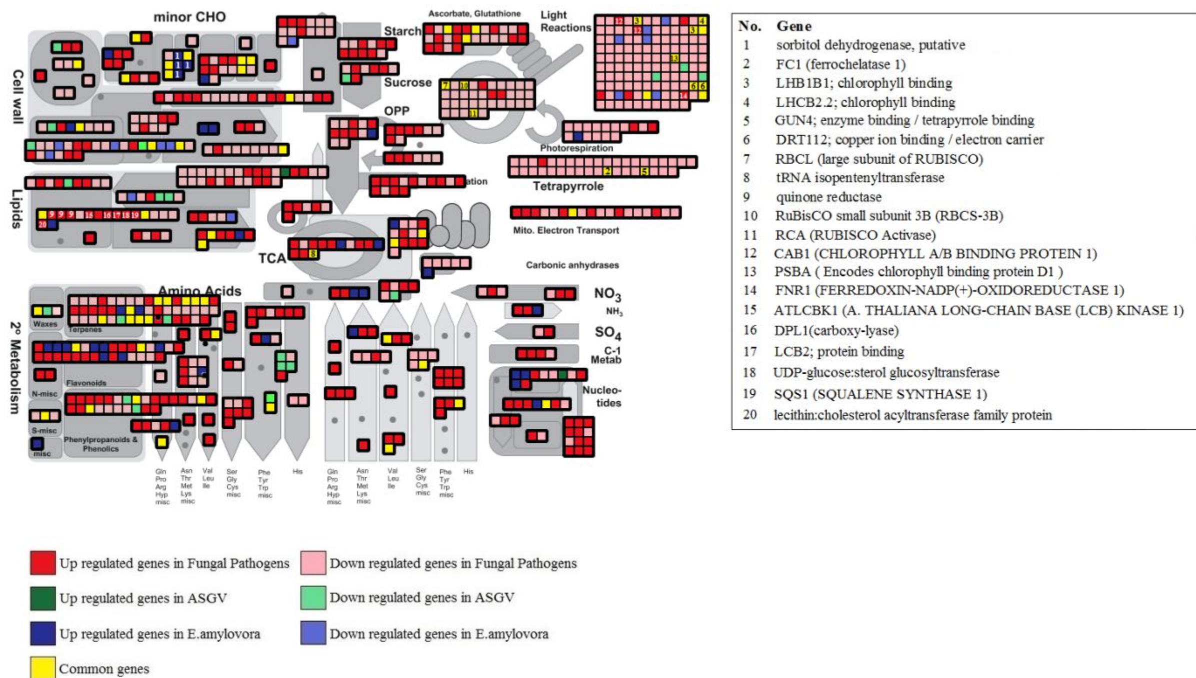
Figure 3.3.1–Gene set enrichment analysis of the differentially expressed genes of the three type of stresses: a) Fungal Pathogens b) ASGV and c) *E. amylovora*. The red color indicates the up-regulated categories and green indicates down-regulated.

3.4 Metabolism overview

I used the Mapman (Thimm et al., 2004) web-tool to visualize the metabolome changes in *Malus x domestica* due to biotic stress by using the transcriptomic data of the seven datasets. Metabolism overview clearly showed the high number of down-regulated genes by fungal pathogens involved in light reactions, photorespiration, Calvin cycle, photorespiration and tetrapyrrole pathways (Fig. 3.4.1). There were several genes commonly related between at least two of the three types of pathogens such as those genes related to phosynthesis: cholorophyll binding (*LHB1B1*, *LHCB2.2*), photorespiration-related genes (*FC1*, *GUN4*), large and small subunits of Rubisco Protein (*RBCL*, *RBCS*), rubisco activase. Several genes were induced by fungal pathogens involved in TCA cycle, detoxifying mechanisms (ascorbate and glutathione), gluconeogenesis, starch and fermentation, lipid metabolism. n contrast, *E. amylovora* and viruses repressed genes involved in cell wall modifications. Although three genes (MDP0000188052, MDP0000873573, MDP0000515106

) were commonly regulated by more than one stress, *E. amylovora* seems to induce specific expression changes in sugar alcohol metabolism (upregulation of genes MDP0000707567, MDP0000149907, MDP0000167088 and MDP0000638442

). As far the secondary metabolism concerns, an induction of genes involved in phenylpropanoids and phenolics were mostly induced by fungal pathogens. Ten genes involved in terpene pathways were commonly regulated. Degradation of nucleotides was mostly enhanced by fungal pathogens. In contrast, Mapman displays large gene expression datasets from different studies in a single metabolic pathway diagram, which help us to easily identify the key genes and its details in different functional categories (Fig. 3.4.1).



No.	Gene
1	sorbitol dehydrogenase, putative
2	FC1 (ferrochelatase 1)
3	LHB1B1; chlorophyll binding
4	LHCB2.2; chlorophyll binding
5	GUN4; enzyme binding / tetrapyrrole binding
6	DRT112; copper ion binding / electron carrier
7	RBCL (large subunit of RUBISCO)
8	tRNA isopentenyltransferase
9	quinone reductase
10	RuBisCO small subunit 3B (RBCS-3B)
11	RCA (RUBISCO Activase)
12	CAB1 (CHLOROPHYLL A/B BINDING PROTEIN 1)
13	PSBA (Encodes chlorophyll binding protein D1)
14	FNR1 (FERREDOXIN-NADP(+)-OXIDOREDUCTASE 1)
15	ATLCBK1 (A. THALIANA LONG-CHAIN BASE (LCB) KINASE 1)
16	DPL1(carboxy-lyase)
17	LCB2; protein binding
18	UDP-glucose:sterol glucosyltransferase
19	SQS1 (SQUALENE SYNTHASE 1)
20	lecithin:cholesterol acyltransferase family protein

Figure 3.4.1–Mapman metabolism overview of differentially expressed genes divided in 7 categories based on their pattern of expression. The important genes were indicated in the figure.

3.5 Hormone-related pathways

It is very important to study the plant hormonal responses because the signaling pathways of different hormones regulate biotic stress responses antagonistically. The Abscisic acid (ABA) related genes *CCD1* (Carotenoid cleavage dioxygenase 1), *NCED4* (Nine-cis-epoxycarotenoid dioxygenase 4), abscisic acid-responsive HVA22 family protein and HVA22-like protein were up-regulated by *Erwinia amylovora*. In contrast, the genes ABF3 (Abscisic acid responsive elements-binding factor 3), GRAM domain-containing protein and *GEM* (gl2-expression modulator) in ABA were downregulated by bacteria. Three genes (MDP0000837051, MDP0000746652, MDP0000130173) in saliciylic acid were affected by all types of pathogenes. I observed that several genes involved in auxin (IAA), benzyl-adenine (BA), ethylene, jasmonate, saliciylic acid (SA) were commonly affected by all types of pathogens. Fungal pathogens up-regulated ethylene, benzyl-adenine, salicylic acid while mainly repressed jasmonate-related genes. *Erwinia amylovora* up-regulated ethylene and gibberellin-related (GA) genes while viruses affected some key genes involved in ethylene

and auxins. The results suggest that the hormone-related pathways and consequently their crosstalk were profoundly affected by all *Malus x domestica* pathogens (Fig. 3.5.1).

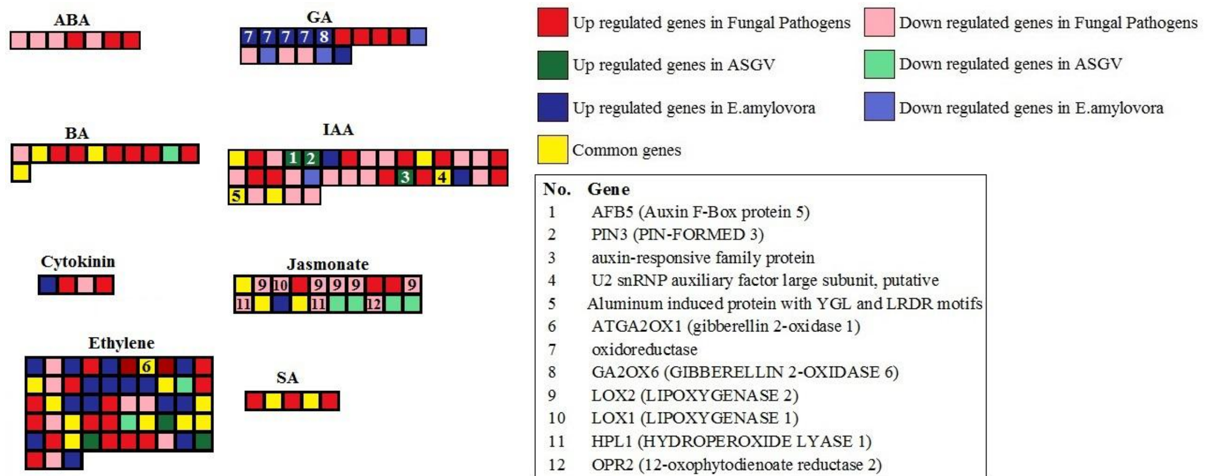
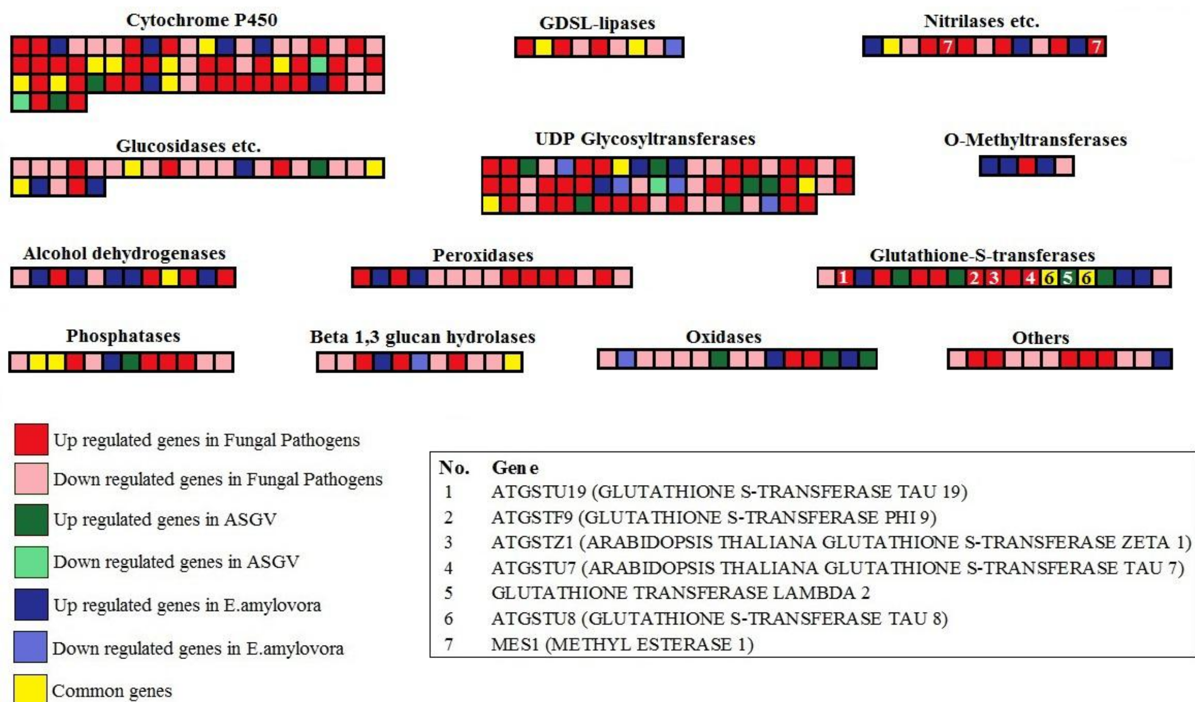


Figure 3.5.1–Hormone-related genes affected by biotic stresses in *Malus x domestica*. The important genes were indicated.

3.6 Detoxifying pathways and secondary metabolism

Genes encoding Cytochrome P450 were commonly affected by all three types of pathogens (Fig.3.6.1). Fungal pathogens induced UDP Glycosyltransferases, Phosphatases, Nitrilases, Glutathione-S-transferases. In addition to UDP-Glycosyltransferases, viruses mainly enhanced the expressions of Oxidases and Glutathione-S-transferases. Alcohol dehydrogenases, Nitrilases, O-Methyltransferases and Peroxidases genes were up-regulated by *Erwinia amylovora*. Two genes encoding GDSL-lipases were commonly regulated by all the three types of pathogens.



No.	Gene
1	ATGSTU19 (GLUTATHIONE S-TRANSFERASE TAU 19)
2	ATGSTF9 (GLUTATHIONE S-TRANSFERASE PHI 9)
3	ATGSTZ1 (ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE ZETA 1)
4	ATGSTU7 (ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE TAU 7)
5	GLUTATHIONE TRANSFERASE LAMBDA 2
6	ATGSTU8 (GLUTATHIONE S-TRANSFERASE TAU 8)
7	MES1 (METHYL ESTERASE 1)

Figure 3.6.1–Hormone-related genes affected by biotic stresses in *Malus x domestica*. The important genes were indicated.

Clear differences in pattern regulation were observed for fungal pathogens in relation to secondary metabolism gene categories (Fig.3.6.2). While Phenlypropanoids, Shikimate pathway, Dihydroflavonols, MVA pathway, Simple phenols were mostly induced, Non-MVA pathway, Carotenoids were repressed. On the other hand, *Erwinia amylovora* mostly up-regulated the genes in Chalcones metabolism and MVA pathway. Overall, I observed that the biotic stress influenced the secondary metabolites in *Malus x domestica*, since these are the compounds which are important for the plant to interact with the environment for adaptation and defense.

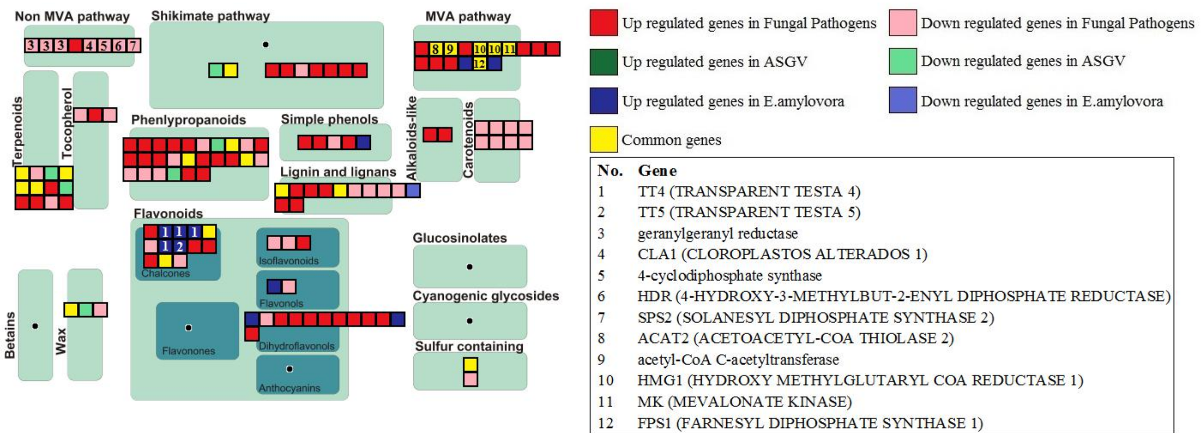


Figure 3.6.2–Secondary metabolism genes affected by the three different biotic stresses. The key genes were indicated in the figure.

3.7 Transcription factors and defense stress-related genes

The Mapman software (https://mapman.gabipd.org/) were used to understand the influence of biotic stress in metabolism, hormone regulation, large enzyme families, secondary metabolism and transcription factors. The transcription factors were drastically affected by the three types of stresses (Fig.3.7.1). RAP2.3, ERF110, CRF4, four RAP2.4, six AP2 domain-containing transcription factor family proteins, two CEJ1 and one unknown protein in AP2-EREBP were up-regulated by fungal pathogens. GRAS factors (SCL1, SCL3, RGA1, SCL13), MADS box (AGL8, AGL24, AGL20, AGL42), C2H2 (MGP, STZ,SEU,STOP1), Psudo ARR (PRR7, PRR5, PRR1) were enhanced by fungal necrotrophic pathogens. Viruses specifically induced some key genes encoding two AP2-EREBPs (AP2, TEM1), two C2H2 (RHL41, zinc ion binding), two HB (GL2, BLH3), two trihelix factors (GT2, trihelix DNA-binding protein). *Erwinia amylovora* up-regulated specifically MYB factors (MYB42, MYB15, MYB14, AtMYB111, DNA binding, AtMYB74, MYB33, MYB62) and JUMONJI (cyclin-like F box, jmjC).

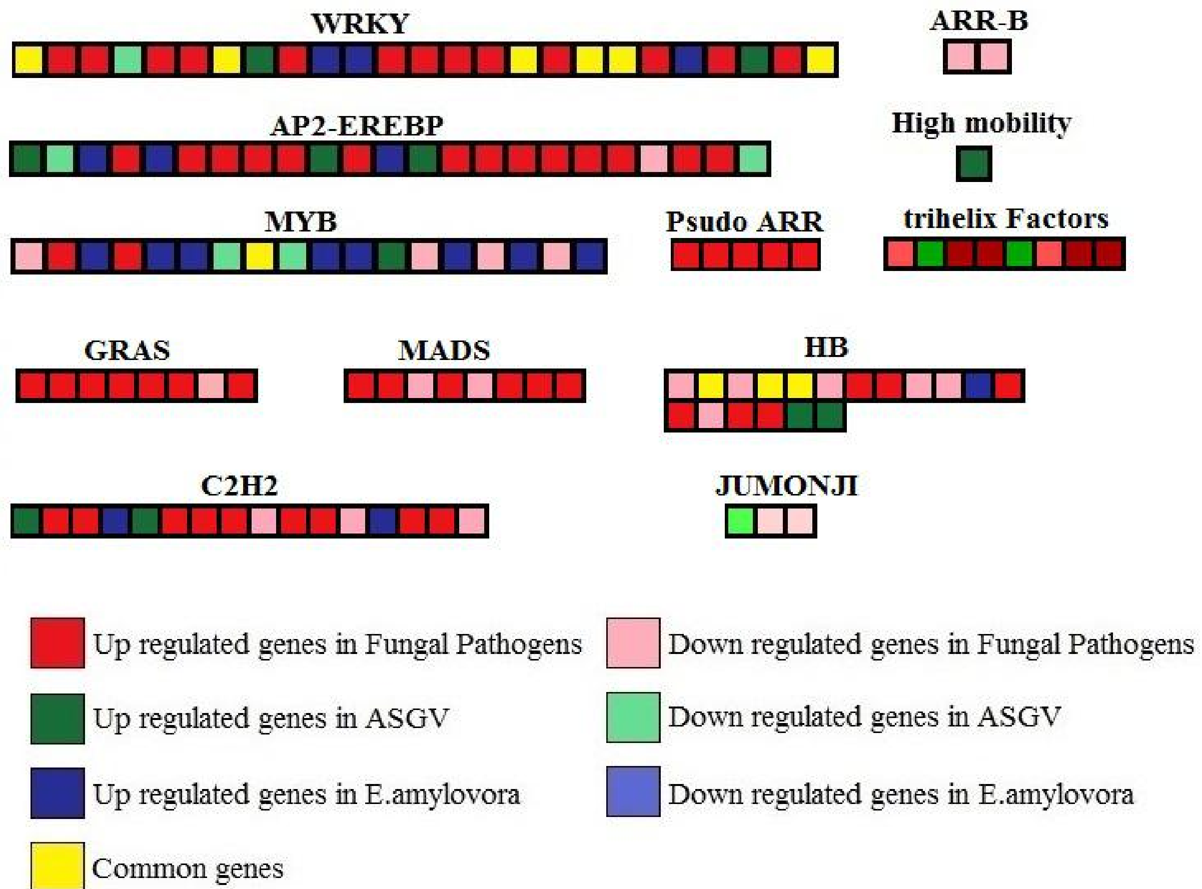


Figure 3.7.1–Genes encoding transcription factors and affected by the different categories of biotic stress. The important genes were indicated.

As expected, *WRKY*s were mostly induced by all three kinds of biotic attacks (Table 3.1). *WRKY11*, *WRKY32*, *WRKY33*, *WRKY35*, *WRKY40*, *WRKY6*, *WRKY65*, *WRKY69*, *WRKY70*, *WRKY72*, *WRKY75* and *TTG2* were induced by at least one of the 5 fungal pathogens. *WRKY53*, *WRKY70* and *WRKY35* were enhanced by ASGV. *WRKY75*, *WRKY33* were specifically induced by *E. amylovora*.

Malus Gene ID	TAIR ID	*WRKY* Gene	Fungal Pathogens	ASGV	*E. amylovora*
MDP0000794439	AT1G80840	*WRKY40*	Up		Up
MDP0000175240	AT3G56400	*WRKY70*	Up		
MDP0000304113	AT1G29280	*WRKY65*	Up		
MDP0000118810	AT1G29860	*WRKY71*		Down	
MDP0000307516	AT1G80840	*WRKY40*	Up		
MDP0000293456	AT4G30935	*WRKY32*	Up		
MDP0000123467	AT5G13080	*WRKY75*	Up		Up
MDP0000191017	AT4G23810	*WRKY53*		Up	
MDP0000676216	AT3G58710	*WRKY69*	Up		
MDP0000154734	AT5G13080	*WRKY75*			Up
MDP0000792088	AT5G13080	*WRKY75*			Up
MDP0000273851	AT5G15130	*WRKY72*	Up		
MDP0000133918	AT1G29280	*WRKY65*	Up		
MDP0000935652	AT1G62300	*WRKY6*	Up		
MDP0000301666	AT1G62300	*WRKY6*	Up		
MDP0000708692	AT2G38470	*WRKY33*	Up		Up
MDP0000935996	AT2G38470	*WRKY33*	Up		
MDP0000177906	AT1G80840	*WRKY40*	Up		Up
MDP0000228304	AT3G56400	*WRKY70*	Down	Up	Up
MDP0000514115	AT2G38470	*WRKY33*	Up		
MDP0000507805	AT2G38470	*WRKY33*			Up
MDP0000272940	AT4G31550	*WRKY11*	Up		
MDP0000202292	AT2G34830	*WRKY35*		Up	
MDP0000169621	AT2G37260	*TTG2*	Up		
MDP0000294489	AT2G34830	*WRKY35*	Up	Up	

Table 3.1–List of *WRKY*s affected by the types of biotic stresses and their pattern of regulation.

Other genes involved in biotic stress responses were drastically affected by all the three types of stresses. One gene involved in respiratory burst was commonly regulated (Fig.3.7.2). Four signaling MLO-like genes were up-regulated by fungal pathogens while one was induced by *E. amylovora* and one was commonly regulated between stresses. In general, it is clear that pathogenesis-related proteins were more induced by fungal pathogens than viruses and *E. amylovora*. Fungi-driven up-regulated genes belonged to TIR-NBS-LRR, ATP binding, CC-NBS-LRR, ADR1-L1, RPP1. Four PR-related genes (MDP0000287351, MDP0000685425,

MDP0000171644, MDP0000635659) were induced only by *E. amylovora* while only one disease resistance gene (MDP0000222184) was commonly regulated.

The results demonstrated that most of the transcription factors and defense stress-related genes were influenced by all types of biotic stresses and also identified the crucial genes response to each type of biotic stress conditions.

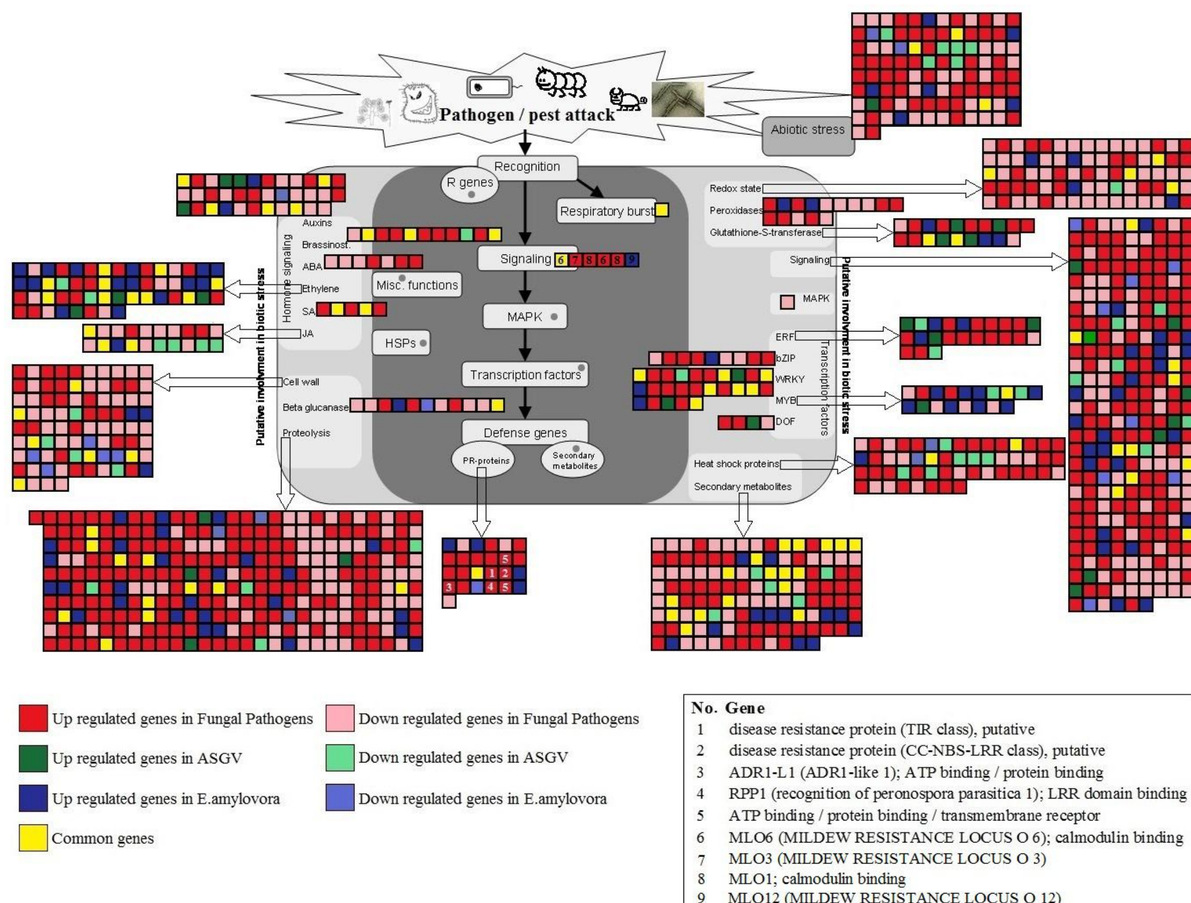


Figure 3.7.2–Biotic stress mapman overview showing genes differentially expressed in the 7 transcriptomic articles and divided in three types of stresses and 7 categories depending on their trend of expression. The key genes were indicated.

3.8 Commonly regulated genes among biotic stresses

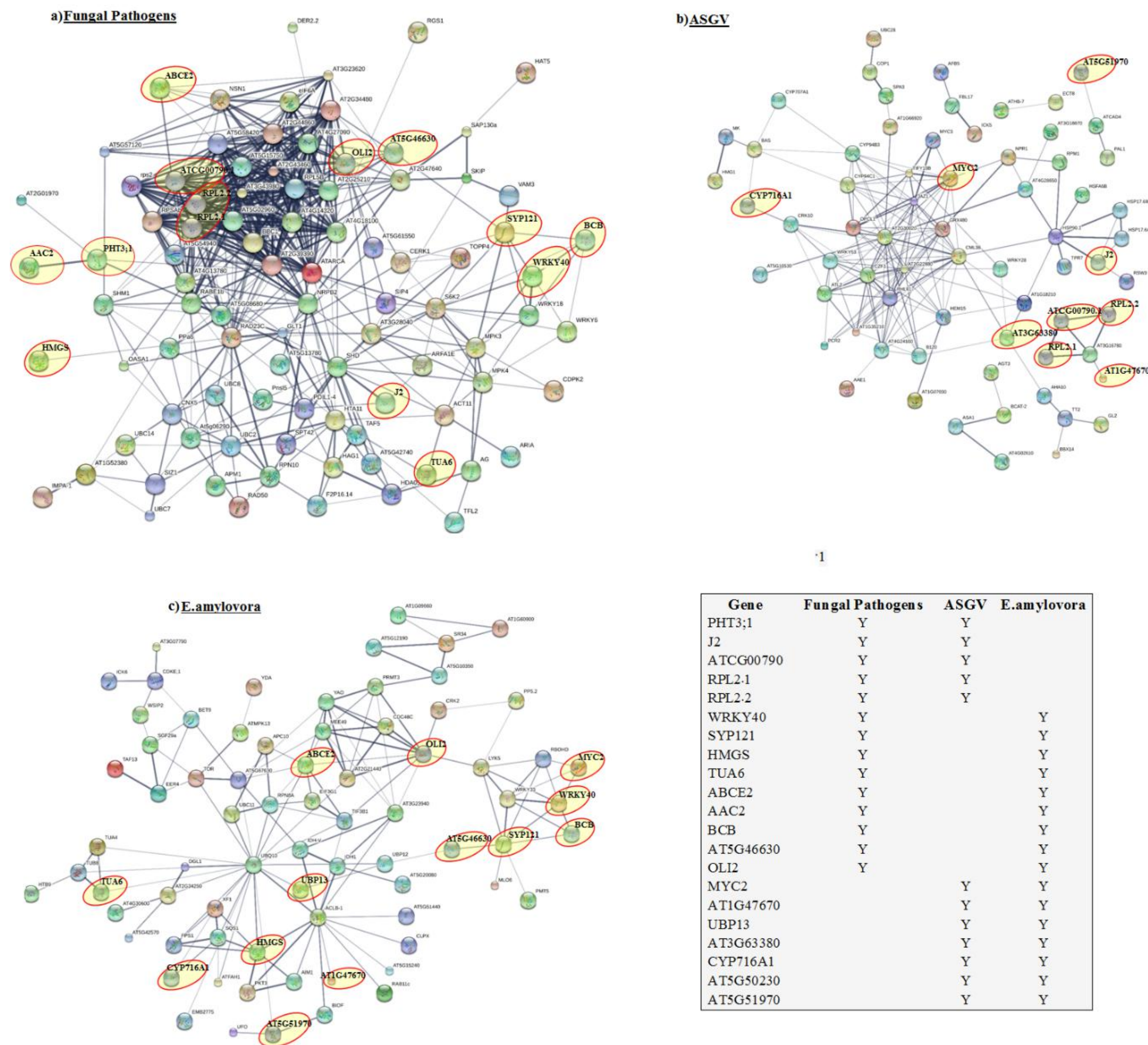
It is very important to find the genes, which are regulated unique to each type of biotic stress and also commonly regulated by all types of biotic stresses. I found that a total of 322 genes

were commonly affected by at least 2 of the 3 types of biotic stresses. These genes represent common responses to stresses and might be helpful to characterize general stress responses in *Malus x domestica*. A great number of these genes were linked with the repression of photosynthesis. Eight genes involved in minor CHO metabolism were affected. Terpenes were affected by all three stresses such as acetyl-coa thiolase2, hydroxyl methylglutaryl coa reductase 1 (HMG1), farnesyl diphosphate synthase1, lyase – magnesium ion binding, beta-amyrin. Also MVA pathway was affected by the three stresses as shown by the differential expression of *HMG1*, *MK*, *FPS1*, *ACAT2*, Acetyl-CoA (Fig. 3.4.1). Four *WRKY*s were commonly regulated by at least 2 of 3 types of pathogens: *WRKY40*, *WRKY75*, *WRKY33*, *WRKY35* and *WRKY70* (Table 3.1). This is very significant information to identify the targets for genetic modification to improve plant resistance to multiple biotic stresses.

3.9 Inferred protein-protein interaction network analysis

To understand the degree of conservation in the protein-protein interaction in *Malus x domestica* in different biotic stress conditions, I visualized the network of the 100 top highly interactive proteins for each of the three types of biotic stresses. *Arabidopsis* orthologs of the *Malus x domestica* pathogen-regulated genes were mapped and the protein-protein interactions were determined basing a combined file of inferred and validated interactions (Li et al., 2011). The network was visualized using STRING (Szklarczyk et al., 2015) software (Version 10.0). A highly dense core of 22–23 highly interactive proteins was observed on the top of the network of fungal pathogens (Fig. 3.9.1 a). Some well-known proteins players in biotic stress responses were noticed such as WRKY40, WRKY18 and WRKY6. These proteins were connected with MPK3 and MPK4. Relating to ASGV infection, MYC2, GRX480, JAZ1, PCL1, RHL41, WRKY53 were hub proteins of this network. Of them only MYC2 was also present in other biotic stresses (*E. amylovora*). Four HSPs were significantly regulated by virus infections and strictly connected each other (Fig. 3.9.1 b). A small network composed of four interactive proteins such as RPL2.1, RPL2.2, ATCG00790.1 and AT1G47670 in ASGV was overlapping with fungal pathogen network. NPR1 was present in the virus-affected Protein-Protein Interaction (PPI) network together with the well-known interactive protein GRX480. *E. amylovora* affected a network connected with five highly interactive genes such as MYC2, WRKY40, WRKY33, BCB, SYP121 and AT5G46630 (Fig.

3.9.1 c). Some key highly interactive ubiquitin proteins were observed in the network such as UBQ10. Sixteen *E. amylovora* proteins regulated at transcriptional level were commonly present also in at least one of the other stresses. The protein-protein interaction network analysis helped to minimize the complexity in understanding physical interaction between proteins due to different biotic stresses.



Gene	Fungal Pathogens	ASGV	E.amylovora
PHT3;1	Y	Y	
J2	Y	Y	
ATCG00790	Y	Y	
RPL2-1	Y	Y	
RPL2-2	Y	Y	
WRKY40	Y		Y
SYP121	Y		Y
HMGS	Y		Y
TUA6	Y		Y
ABCE2	Y		Y
AAC2	Y		Y
BCB	Y		Y
AT5G46630	Y		Y
OLI2	Y		Y
MYC2		Y	Y
AT1G47670		Y	Y
UBP13		Y	Y
AT3G63380		Y	Y
CYP716A1		Y	Y
AT5G50230		Y	Y
AT5G51970		Y	Y

Figure 3.9.1–Inferred protein-protein interaction network based on *Arabidopsis* knowledgebase for the pathogen-regulated genes encoding highly interactive proteins. The common genes (present in more than one in the three groups (1) Fungal Pathogens, (2) ASGV and (3) *E. amylovora*) are highlighted in the red oval shape and also indicated in the given table. Y-indicates that the gene is present in the group.

4. Discussion

My study highlighted the need of more meta-analysis of transcriptomic studies due to several reasons. First, transcript amounts are highly affected by changing environmental and developmental conditions. Secondly, field studies are usually conducted only in one season leading to unreliable results affected by a high number of environmental disturbing factors. Third, few replicates (frequently only three) are usually considered due to the high costs of "omic" analysis. Fourthly, transcriptomic studies should be integrated with proteomics and metabolomics in order to clarify post-transcriptional and post-transductional regulation mechanisms. Finally the identification of commonalities between similar independent studies will identify which gene are more strongly associated with the subject of the study and focus the functional analysis only on those common findings (Sweeney et al., 2017).

Here I showed data of a meta-analysis of 7 published transcriptomic articles dealing with biotic stress responses in *Malus x domestica*. At the moment, in Scopus database, there are 5 articles related to fungal pathogens, one related to virus and one to bacteria (*E. amylovora*).

The significant downregulation of light reactions in response to both fungal pathogens and *E. amylovora* was expected due to the symptomatic stages reported in 6 analyzed articles (Yin et al., 2016; Zhu et al., 2017; Shin et al., 2016; Kamber et al., 2016; Gusberti et al., 2013; Xu et al., 2015). These evidences have been previously reported as a typical response clearly shown not only at phenotypic but also evident in previous gene set enrichment analysis (Martinelli et al., 2013; Punelli et al., 2016). Carbohydrate metabolism has been frequently shown as a key pathway affected by biotic stresses responses in plants (Martinelli et al., 2012). Growing tissues may be seen as a collection of sinks of carbohydrate attracting photosynthates produced by leaves. A correct mechanism of source-sink relationship allows carbon allocation during abiotic and biotic stress consequently improves plant performance in harsh environments (Lo Bianco et al., 2011). The source-sink disruption has been linked with the early pathogenic mechanisms of diseases in plants (Martinelli et al., 2013; Punelli et al., 2016). Indeed, I believe that the dysregulation of this pathway at transcriptomic level may be associated with a general plant stress state. This altered transcript condition may be seen by growers as a sort of "alarm bell" to help further monitoring actions and the beginning of management procedures. Sugar alcohols are acyclic polyols produced outside the chloroplast and they are directly linked with stress responses. Although their role in tolerance to stress

has been more linked with abiotic than biotic stresses, it has been hypothesized that they may play a key role also in a beneficial modulation of biotic stress responses (Moing et al., 1997). My meta-analysis pointed out how sugar alcohols may be more involved in responses to bacterial pathogens.

The repression of detoxifying genes such as those involved in ascorbate and glutathione-S-transferases mainly observed in response to fungal pathogens is a clear evidence of pathological status. Recently the over-expression of these genes have been linked with an increased tolerance to Huanglongbing disease in *Citrus sinensis* (Martinelli et al., 2016). Glutathione S-transferases (GSTs) are proteins encoded by a large family of genes and involved in host defenses against environmental stresses. The transgenic overexpression of a GST in tobacco drives to an increased resistance to *Fusarium oxysporum* (Han et al., 2016) agreeing with my findings that show a significant upregulation of these genes in responses not only to fungal pathogens but also to ASGV and *E. amylovora*. Although the role of glutathione in functioning as protectors during plant abiotic stresses remains to be unclear, a recent work highlighted its importance as signal of hormones and other protecting molecules (Cheng et al., 2015).

Interestingly, I observed that polyamine metabolism was repressed by *E. amylovora* but not by the five fungal pathogens. It is well-known that these molecules are increasingly accumulated in response to stresses as well as transcript abundance of genes involved in their biosynthesis are generally up-regulated. In addition, the transgenic overexpression of these genes enhanced resistance to stresses and several studies showed their key role in the modulation of intra-cellular levels of reactive oxygen species (Liu et al., 2015). It is intriguing why polyamine metabolism resulted to be repressed by *E. amylovora*. It remains to be clarified if they might play a key role in the pathogenetic mechanisms of fire blight in *Malus x domestica*.

Interestingly *E. amylovora* significantly repressed the category of Lipid transport and this may promote the occurrence of the progression of the symptomatology. These proteins are specific pathogenesis-related proteins involved in plant defense responses (Goyal et al., 2014). These proteins are involved in the inhibition of pathogen growth (Molina et al., 1993).

A predominant number of AP2/EREBP TFs were up-regulated in comparison to the down-regulated ones. The five fungal pathogens significantly up-regulated the

APETALA2/ethylene-responsive element binding protein (AP2/EREBP) transcription factors while three of them were induced by the ASGV and three by *E. amylovora*. Only one was commonly regulated between two of the three pathogens. This evidence leads us to speculate that the three kinds of pathogens induce exclusive signaling to activate specific immune responses and the recent studies reported the involment of AP2/ERF transcription factors in plant stress response (Li et al., 2020). Their key role in signal transduction of plant hormones is well-known (Liu et al., 2017). A comprehensive analysis have been conducted in *V. fordii* and *V. montana* and showed how different members may be up- or down-regulated depending on the two species in response to *Fusarium oxysporum* (Zhang et al., 2016). RAV2 was one of these proteins specifically induced by *E. amylovora*. The constitutive overexpression of this gene in tomato enhanced *ERF5* and *PR5* genes increasing the tolerance to bacterial wilt (Li et al., 2011). Previous works suggest that *RAV1* may work as a transcriptional activator inducing resistance to bacterial infection (Sohn et al., 2006). Taken together, these findings lead to speculate that *RAV* genes may be more involved in bacterial defense than fungal and virus pathogens.

GRAS transcription factors are involved in plant disease resistance (Grimplet et al., 2016). Interestingly, I noticed five *SCL* genes that were up-regulated by fungal pathogens but not by the other two types of pathogens although *RGA1* was commonly regulated. GRAS proteins are repressors of gibberellin signaling due to the presence of the N-terminal region amino acid sequence DELLA and are considered DELLA proteins (Grimplet et al., 2016). Indeed, the downregulation of three gibberellin-responsive genes observed by the fungal pathogens agree with the upregulation of GRAS proteins.

MYB proteins present a repeated numbers of MYB domains that allow them to bind DNA. They are commonly expressed in plants and regulated by diverse environmental factors. Their role in ABA-response is well-recognized (Ambawat et al., 2013). Interestingly, 8 MYB proteins were induced only by *E. amylovora* and not by the other pathogens including MYB62 and MYB15. Only MYB6 was commonly regulated between 2 of the 3 kind of pathogens. MYB62 has been linked with phosphate starvation (Devaiah et al., 2009) while MYB15 was induced by wound and insect herbivores responses (Cheong et al., 2002). From my analysis, it seems that, at least in *Malus x domestica*, MYBs are more linked with *E. amylovora* than fungal pathogens and AGSV.

Finally, another important category of transcription factors affected by all three types of pathogens was *WRKY*s. They are well-known for their key role in response to many different environmental stresses (Wang et al., 2016). Thirteen *WRKY*s were up-regulated in response to fungal pathogens, 3 in response to *E. amylovora* and 2 to viral infection. The *WRKY*s that were commonly regulated between all the biotic stresses are interesting because of their important role in the modulation of the hormonal cross-talk in response to pathogens. Six of them were commonly regulated in at least two types of biotic stress including *WRKY70* and *WRKY40*. Since this gene was highly expressed in plants treated with ethylene (ET) and salicylic acid (SA) while it was repressed in response to methyl jasmonate (MeJA), its key role in SA-JA crosstalk has been hypothesized (Wang et al., 2016). This protein showed to have repressive effect on SA-mediated defense while it contributes to stimulate JA-mediated responses. *WRKY33* was up-regulated by all three types of biotic stress. This gene has been linked with bacterial infections (Martinelli et al., 2013; Martinelli et al., 2016; AbuQamar et al., 2016) and it is up-regulated by Trichoderma, a fungal genus that stimulates plant and root growth and nutrient uptake (Mayo et al., 2016). *WRKY53* was up-regulated only by ASGV. It has been shown that two key genes involved in biotic stress responses, a Ser/Thr receptor kinase ORK10/LRK10 and an apoplastic peroxidase were targeted by *WRKY53* (VanEck et al., 2014). *WRKY75* was enhanced by *E. amylovora*. Interestingly its transgenic overexpression allowed improving resistance to *Sclerotinia sclerotiorum* (Chen et al., 2013).

Interestingly a clear upregulation of ethylene-related genes were observed more in response to *E. amylovora* than to the other two biotic stresses. In total, nine 2OG-Fe(II) oxygenases were up-regulated by bacterial infection. Considering the total number of *E. amylovora*-regulated genes, ethylene-related category was highly represented in gene set enrichment analysis. The upregulation of *ERF1* by *E. amylovora* was expected since this gene has been linked with the enhancement of JA-responsive genes through *ORA59* (Pieterse et al., 2009). Data related to Jasmonic acid responsive genes were contrasting in response to *E. amylovora*. While an allene oxydase synthase gene was induced, two genes (an allene oxide synthase 2 and 4) were repressed. *GASA4* was repressed by *E. amylovora*. This gene is part of a family of GA-inducible and ABA-repressible genes. It is generally induced by hormones involved in growth development while it is repressed by stress-related hormones (ABA, JA, and SA), implying its key role in hormone crosstalk (AbuQamar et al., 2017). Fungal pathogens predominantly up-regulated the genes involved in ethylene, brassinosteroids and salicylic

acid while jasmonic acid responses were mostly repressed. This was expected since *Venturia inaequalis*, one of the studied fungal pathogen, is considered a hemi-biotrophic pathogen. Relating to responses to ASGV, an upregulation of two auxin responsive genes, *AFB3* and *PIN3* were observed. PIN3 is an auxin transporter that plays a key role in root growth and lateral architecture mediated in the hypocotyl (Rakusova et al., 2016). Although the role of this gene in pathogen defense responses has to be elucidated, it may be somewhat affected since exogenous SA showed that mostly repressed Pin-formed (*PIN*) genes (Armengot et al., 2016).

Different categories of genes involved in secondary metabolism were selectively regulated by *apple stem grooving virus*, fungal pathogens and *E. amylovora*. While fungal pathogens up-regulated shikimate pathways, MVA and phenylpropanoids, *E. amylovora* clearly induced genes involved in chalcones. In contrast, non-MVA was clearly repressed by fungal pathogens. MVA pathway is responsible for terpenoid biosynthesis and was commonly affected by the studied pathogens as shown by the significant regulation of 6 genes. Terpenoids comprise a series of metabolites with peculiar protection roles to biotic attacks. Several volatile sesquiterpenes are important chemical signals for the activation of plant defence mechanisms in response to biotic stresses. The wide range of different terpenoids present in plants, implied that they should have posed an important role in plant evolution in response to different ecological plant interactions with both biotic and abiotic aspects.

The expression of genes involved in phenylpropanoid metabolism was clearly induced by fungal pathogens. They have important protective roles towards both biotic and abiotic stresses and they are regulated by MYB transcription factors (Liu et al., 2015). Increased amount of phenylpropanoid transcripts were also associated with *Citrus sinensis* responses to Huanglongbing disease (Martinelli et al., 2015). Secondary metabolism genes including chalcone isomerases were up-regulated by *Marssonina coronaria*. Three genes were commonly regulated between stresses while 5 naringenin-chalcone synthase genes were up-regulated by *E. amylovora*. Both chalcones and dihydrofavonols were up-regulated by the fungal pathogens and *E. amylovora*. These compounds belonged to flavonoids, an important class of secondary metabolism compounds with protecting functions against fungal infection. They are categorized into two groups: constitutively expressed and stimulated. The first category are usually maintained in particular locations and used as signals when pathogen attacks occur (Treutter, 2006) while the second comprises genes induced during plant-

pathogen interactions. These compounds exercise a protection role thank to their antioxidant capabilities, cross-linking and inhibition of microbial proteins such as cell wall degradation enzymes, metal chelation as well as physical barrier against pathogens (Skadhauge et al., 1997; Beckman, 2000). Interestingly carotenoid genes were repressed by fungal pathogens. Their protection role in plant resistance to biotic stresses was shown in mutant experiments that demonstrated their important role in ROS detoxyfication under stress conditions (Demmig-Adams,B et al., 2014). Results of this meta-analysis suggest that the repression of these genes in response to apple fungal pathogens might provoke negative effects on the progression of the disease.

Relating to fungal pathogens, a core of 22–23 highly proteins were clearly observed by the observation of the overall network. Among these proteins, there were RPL5B, BBC1, ATARCA. Three *WRKY*s, *WRKY6*, *WRKY18* and *WRKY40* were significantly regulated by fungal pathogens in *Malus x domestica*. WRKY18 and WRKY40 proteins formed complexes and presented DNA binding properties. These WRKYs are involved in pathogen-induced HR linked with the induction of salicylic acid (SA)–mediated immune responses causing the progression of the systemic acquired resistance (SAR). WRKY18 and WRKY40 have common sequences with more than 60% identical amino acids (Xu et al., 2006). *WRKY18*, *WRKY40*, and *WRKY60* showed negatively affect resistance to hemibiotrophic pathogens (Xu et al., 2006). In addition, different *WRKY*s seems to have contrasting effects on response to *Pseudomonas syringae* and to *B. cinerea*. Indeed, these three WRKY proteins may be negative regulators of the SA-dependent pathways while they induce JA-mediated pathways. A mitogen-activated protein kinase 4 (MPK4), inhibitor of SA-dependent resistance (Petersen et al., 2000), was shown to interact with WRKY25 and WRKY33 implying that their role in response to necrotrophs might be complex (Andreasson et al., 2005). It seems that resistance to the fungal pathogens in *Malus x domestica* was associated with specific expression levels of these three *WRKY* genes. Hypersensitive reaction is important for the virulence of the fungal pathogen *B. cinerea* (Govrin and Levine, 2000). SA and ET upregulates signalling pathways antagonistic to each other, but both of them enhance pathogen-induced cell death (Shirasu et al., 1997). Therefore, it is possible that *WRKY* genes cause activation or suppression of diverse signalling mechanisms in response to necrotrophic pathogens promoting virulence. It is worthy to notice that cooperative bonds with different WRKY proteins might regulate their activity as transcription factors. Thirteen proteins interactive at

protein-protein level were commonly regulated with the other two biotic stresses and may be considered a general plant stress state.

MYC2 was differentially regulated by both ASGV and *E. amylovora*. MYC2 is considered as a key regulating protein of JA signalling in *Arabidopsis* (Fernandez-Calvo et al., 2011) since it interacts with JASMONATE ZIM-domain proteins. Transgenic increased expression of *OsMYC2* stimulated the expression of early JA-responsive genes, inducing bacterial blight resistance through JA-hypersensitive reaction (Uji et al., 2016). Interestingly, I observed a protein-protein interaction network shared with the other biotic stresses and consisting of key proteins such as MYC2, WRKY40, BCB, SYP121. The PPI network showed that WRKY33, WRKY40 and MYC2 were strictly connected and significantly regulated by *E. amylovora*. The expression of *WRKY40* was enhanced in response to wounding and infections of *Ralstonia solanacearum*. This gene was regulated by salicylic acid, methyl jasmonate, ethylene (Wang et al., 2014). MYC2 interacts with NPR1 that was also significantly affected by ASGV. This gene was up-regulated by SA (Ryals et al., 1997) and its subcellular localization was modulated by redox changes caused by salicylic acid (Mou et al., 2003). It binds directly to salicylic acid, releasing its transactivation domain and it is regulated by proteasome-mediated turnover (Spoel et al., 2009).

Interestingly, ASGV modulated the expression of heat shock proteins such as HSP90.1, HSP176B, HSP17A and HSFA6B. The PPI network showed that these proteins were strictly interacting with each other. Heat shock proteins were frequently induced in cells of all organisms in response to heat (Almoguera et al., 1995). Their function is to protect protein folding since they are able to reduce protein misfolding due to all kind of stresses. HSP90 was involved in signal transduction of plant responses through the interaction of a salicylic acid-induced protein kinase. HSP90 affects defense responses against pathogens through specific interactions with other genes, working as a scaffold in protein complexes involved in signal transduction (Schulze-Lefert, 2004). Along with transcriptomics analysis, the protein-protein interaction network analysis help us to visualize and identify the key node proteins which are affected pathogen infections. Thus, the meta-analysis plays a major role in identifying potential biomarkers for different biotic stress conditions in plants by comparing different omic data sets pertaining to a specific functional context (AbuQamar et al., 2016).

5. Conclusions

Comparisons of transcriptomic datasets obtained to study different biotic stress in the same crop allows identifying which genes are specifically involved in disease resistance and which may be associated with general plant stress conditions. Meta-analyses allow increase in reliability of transcriptomic data, reducing environmental variability due to a low number of biological replicates and repeated experiments. In this work, the meta-analysis conducted in *Malus x domestica*, highlights the role of *WRKY*s in the molecular response to biotic stresses at both transcript and protein-protein interaction levels. Although *WRKY40* was involved response to both fungal pathogens and *E. amylovora*, its interaction with other different *WRKY* may induce specific responses. In response to fungal pathogens, *WRKY* interacted with two other pathogen-regulated *WRKY6* and *WRKY18* while in response to *E. amylovora* it interacts with *WRKY33*. Specific hormones were differentially affected between the three types of stresses and drives to specific defense responses. Future studies in other crops investigating similar diseases will allow validate these findings and identify resistance mechanisms in gene regulatory networks of plant-microbe interactions.

6. References

AbuQamar S. F., Moustafa K., Tran L. S. (2016). 'Omics' and Plant Responses to *Botrytis cinerea*. Front Plant Sci. 7, 1658-1658.

AbuQamar S. F., Moustafa K., Tran L. S. (2017). Mechanisms and strategies of plant defense against *Botrytis cinerea*. Crit. Rev. Biotechnol. 37, 263–275.

Almoguera C., Coca M. A., Jordano J. (1995). Differential Accumulation of Sunflower Tetraubiquitin mRNAs during Zygotic Embryogenesis and Developmental Regulation of Their Heat-Shock Response. Plant Physiology 107, 765-773.

Ambawat S., Sharma P., Yadav N. R., Yadav R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. Physiology and Molecular Biology of Plants 19, 307-321.

Andreasson E., Jenkins T., Brodersen P., Thorgrimsen S., Petersen N. H., Zhu S., et al. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. Embo Journal 24, 2579-2589.

Armengot L., Caldarella E., Marques-Bueno M. M., Martinez M. C. (2016). The Protein Kinase CK2 Mediates Cross-Talk between Auxin- and Salicylic Acid-Signaling Pathways in the Regulation of PINOID Transcription. Plos One 11, e0157168.

Beckman C. H. (2000). Phenolic-storing cells: keys to programmed cell death and periderm formation in wilt disease resistance and in general defence responses in plants?. Physiological and Molecular Plant Patholology 57, 101-10.

Bhaya D., Davison M., Barrangou R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 45, 273-97.

Bolger A. M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Brekketet T. D., Stroud J. A., Shaw D. S., Crawford S., Steele K. A. (2019). QTL mapping in salad tomatoes. Euphytica 215, 115.

Chen S., Ye T., Hao L., Chen H., Wang S., Fan Z., et al. (2014). Infection of Apple by *Apple Stem Grooving Virus* Leads to Extensive Alterations in Gene Expression Patterns but No Disease Symptoms. Plos One 9, e95239.

Chen X., Liu J., Lin G., Wang A., Wang Z., Lu G. (2013). Overexpression of *AtWRKY28* and *AtWRKY75* in *Arabidopsis* enhances resistance to oxalic acid and Sclerotinia sclerotiorum. Plant Cell Reports 32, 1589-1599.

Cheng M. C., Ko K., Chang W. L., Kuo W. C. Chen G. H., Lin T. P. (2015). Increased glutathione contributes to stress tolerance and global translational changes in *Arabidopsis*. Plant Journal 83, 926-939.

Cheong Y. H., Chang H. S., Gupta R., Wang X., Zhu T., Luan S. (2002). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. Plant Physiol 129, 661–677.

Cohen S. P., Leach J. E. (2019). Abiotic and biotic stresses induce a core transcriptome response in rice. Sci Rep. 9, 6273.

Dandekar A. M., Martinelli F., Davis C. E., Bhushan A., Zhao W., Fiehn O., et al. (2010). Analysis of Early Host Responses for Asymptomatic Disease Detection and Management of Specialty Crops. Critical Reviews in Immunology 30, 277-289.

De Ronde D., Butterbach P., Kormelink R. (2014). Dominant resistance against plant viruses. Front. Plant Sci. 5, 307.

Demmig-Adams B., Stewart J. J., Adams W. W. (2014). Multiple feedbacks between chloroplast and whole plant in the context of plant adaptation and acclimation to the environment. Philosophical Transactions of the Royal Society B-Biological Sciences 369, 20130244.

Devaiah B. N., Madhuvanthi R., Karthikeyan A. S., Raghothama K. G. (2009). Phosphate Starvation Responses and Gibberellic Acid Biosynthesis Are Regulated by the MYB62 Transcription Factor in *Arabidopsis*. Molecular Plant 2, 43-58.

Fernández-Calvo P., Chini A., Fernández-Barbero G., Chico J. M., Gimenez-Ibanez S., Geerinck J., et al. (2011). The *Arabidopsis* bHLH Transcription Factors MYC3 and MYC4 Are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. Plant Cell 23, 701-715.

Gallois J. L., Moury B., German-Retana S. (2018). Role of the Genetic Background in Resistance to Plant Viruses. Int J Mol Sci. 19, 2856.

Gambino G., Cuozzo D., Fasoli M., Pagliarani C., Vitali M., Boccacci P., et al. (2012). Co-evolution between *Grapevine rupestris stem pitting-associated virus* and *Vitis vinifera L.* leads to decreased defence responses and increased transcription of genes related to photosynthesis. Journal of Experimental Botany 63, 5919-5933.

Govrin E. M., Levine A. (2000). The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. Current Biology 10, 751-757.

Goyal R. K., Mattoo A. K. (2014). Multitasking antimicrobial peptides in plant development and host defense against biotic/abiotic stress. Plant Science 228, 135-149.

Griffith M., Walker J. R., Spies N. C., Ainscough B. J., Griffith O. L. (2015). Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLoS Comput Biol. 11, e1004393.

Grimplet J., Agudelo-Romero P., Teixeira R. T., Martinez-Zapater J. M., Fortes A. M. (2016). Structural and Functional Analysis of the GRAS Gene Family in Grapevine Indicates a Role of GRAS Proteins in the Control of Development and Stress Responses. Frontiers in Plant Science 7, 353.

Gusberti M., Gessler C., Broggini G. A. L. (2013). RNA-Seq Analysis Reveals Candidate Genes for Ontogenic Resistance in *Malus-Venturia* Pathosystem. Plos One 8, e78457.

Han Q., Chen R., Yang Y., Cui X., Ge F., Chen C., et al. (2016). A glutathione S-transferase gene from Lilium regale Wilson confers transgenic tobacco resistance to *Fusarium oxysporum*. Scientia Horticulturae 198, 370-378.

Haque E., Taniguchi H., Hassan M. M., Bhowmik P., Karim M. R., Śmiech M., Zhao K., Rahman M., Islam T. (2018). Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges. Front Plant Sci. 9, 617.

Huang D. W., Sherman B. T. & Lempicki R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 4, 44-57

Jones J. D., Dangl J. L. (2006). The plant immune system. Nature. 444, 323–329.

Kamber T., Buchmann J. P., Pothier J. F., Smits T. H., Wicker T., Duffy B. (2016). Fire blight disease reactome: RNA-seq transcriptional profile of apple host plant defense responses to *Erwinia amylovora* pathogen infection. Scientific Reports 6, 21600.

Kanfra X., Liu B., Beerhues L., Sørensen S. J., Heuer H. (2018) .Free-Living Nematodes Together With Associated Microbes Play an Essential Role in Apple Replant Disease. Front Plant Sci. 9, 1666.

Kukurba K. R., Montgomery S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harb Protoc. 11, 951-969.

Li C. W., Su R. C., Cheng C. P., Sanjaya, You S. J., Hsieh T. H., et al. (2011). Tomato RAV Transcription Factor Is a Pivotal Modulator Involved in the AP2/EREBP-Mediated Defense Pathway. Plant Physiology 156, 213-227.

Li H., Dong Q., Zhao Q., Shi S., Ran K. (2020). Isolation, sequencing, and expression analysis of 30 AP2/ERF transcription factors in apple. PeerJ. 8, e8391.

Li J., Liu C., Sun C., Yu-Ting. (2018). Plant stress RNA-seq Nexus: a stress-specific transcriptome database in plant cells. BMC Genomics. 19, 966.

Limera C., Sabbadini S., Sweet J. B., Mezzetti B. (2017). New Biotechnological Tools for the Genetic Improvement of Major Woody Fruit Species. Front Plant Sci. 8, 1418.

Li P., Zang W., Li Y., Xu F., Wang J., Shi T.(2011). AtPID: the overall hierarchical functional protein interaction network interface and analytic platform for *Arabidopsis*. Nucl. Acids Res 39, D1130-D1133.

Liu C., Zhang T. (2017). Expansion and stress responses of the AP2/EREBP superfamily incotton. BMC Genomics 18, 118.

Liu J. H., Wang W., Wu H., Gong X. Q., Moriguchi T. (2015). Polyamines function in stress tolerance: from synthesis to regulation. Frontiers in Plant Science 6, 827.

Liu J. Y., Osbourn A., Ma P. D. (2015). MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Molecular Plant 8, 689-708.

Lo Bianco R. (2011). Carbohydrate metabolism and source-sink relationships in peach. Tree Growth: Influences, Layers and Types. Nova Science Publishers. 79-91.

Marone D., Russo M., Laidò G., De Leonardis A., Mastrangelo A. (2013). Plant Nucleotide Binding Site–Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. Int. J. Mol. Sci. 14, 7302–7326.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17, 10-12.

Martinelli F., Dolan D., Fileccia V., Reagan R. L., Phu M., Spann T. M., et al. (2016). Molecular Responses to Small Regulating Molecules against Huanglongbing Disease. Plos One 11, e0159610.

Martinelli F., Ibanez A., Reagan R., Davino S., Dandekar A. M. (2015). Stress responses in *Citrus* peel: Comparative analysis of host responses to Huanglongbing disease and puffing disorder. Scientia Horticulturae 192, 409-420.

Martinelli F., Reagan R. L., Uratsu S. L., Phu M. L., Albrecht U., Zhao W., et al. (2013). Gene Regulatory Networks Elucidating Huanglongbing Disease Mechanisms. Plos One 8, e74256.

Martinelli F., Uratsu S. L., Albrecht U., Reagan R. L., Phu M. L., Britton M., et al. (2012). Transcriptome Profiling of *Citrus* Fruit Response to Huanglongbing Disease. Plos One 7, e38039.

Mayo S., Cominelli E., Sparvoli F., González-López O., Rodríguez-González A., Gutiérrez S., et al. (2016). Development of a qPCR Strategy to Select Bean Genes Involved in Plant Defense Response and Regulated by the *Trichoderma velutinum - Rhizoctonia solani* Interaction. Frontiers in Plant Science 7, 1109.

Moing A., Langlois N., Svanella L., Zanetto A., Gaudillere J. P. (1997). Variability in sorbitol: Sucrose ratio in mature leaves of different *Prunus* species. Journal of the American Society for Horticultural Science 122, 83-90.

Molina A., Garciaolmedo F. (1993). Developmental and pathogen-induced expression of three barley genes encoding lipid transfer proteins. Plant Journal 4, 983-991.

Moradpour M., Abdulah S. N. A. (2020). CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. Plant Biotechnol J. 18, 32-44.

Mou Z., Fan W. H., Dong X. N. (2003). Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. Cell 113, 935-944.

Muleo R., Morgante M., Cattonaro F., Scalabrin S., Cavallini A., Natali L., Perrotta G., Lopez L., Velasco R., Kalaitzis P. (2016). Genome sequencing, transcriptomics, and proteomics. In: Rugini E, Baldoni L, Muleo R, Sebastiani L, editors. The olive tree genome, Compendium of plant genomes, 141–61.

Nelson R., Wiesner-Hanks T., Wisser R., Balint-Kurti P. (2018). Navigating complexity to breed disease-resistant crops. Nat. Rev. Genet. 19, 21–33.

Petersen M., Brodersen P., Naested H., Andreasson E., Lindhart U., Johansen B., et al. (2000). *Arabidopsis* MAP kinase 4 negatively regulates systemic acquired resistance. Cell 103, 1111-1120.

Pieterse C. M. J., Leon-Reyes A., VanderEnt S., VanWees S. C. M. (2009). Networking by small-molecule hormones in plant immunity. Nature Chemical Biology 5, 308-316.

Pilet-Nayel M. L., Moury B., Caffier V., Montarry J., Kerlan M. C., Fournet S., Durel C. E., Delourme R. (2017). Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. Front. Plant Sci. 8, 1838.

Punelli F., Hassan M. A., Fileccia V., Uva P., Pasquinia G. Martinelli F. (2016). A microarray analysis highlights the role of tetrapyrrole pathways in grapevine responses to *"stolbur" phytoplasma, phloem virus* infections and recovered status. Physiological and Molecular Plant Pathology 93, 129-137.

Rakusová H., Abbas M., Han H., Song S., Robert H. S., Friml J., et al. (2016). Termination of Shoot Gravitropic Responses by Auxin Feedback on PIN3 Polarity. Current Biology 26, 3026-3032.

Rest J. S., Wilkins O., Yuan W., Purugganan M. D., Gurevitch J. (2016). Meta-analysis and meta-regression of transcriptomic responses to water stress in *Arabidopsis*. Plant J 85, 548-560.

Rojas-Vásquez R., Gatica-Arias A. (2020). Use of genome editing technologies for genetic improvement of crops of tropical origin. Plant Cell Tiss Organ Cult 140, 215–244.

Ryals J., Weymann K., Lawton K., Friedrich L., Ellis D., Steiner H. Y., et al. (1997). The *Arabidopsis* NIM1 protein shows homology to the mammalian transcription factor inhibitor I kappa B. Plant Cell 9, 425-439.

Sabbadini S., Capriotti L., Limera C., Navacchi O., Tempesta G., Mezzetti B. (2019). A plant regeneration platform to apply new breeding techniques for improving disease resistance in grapevine rootstocks and cultivars. BIO Web Conf. 12, 01019.

Schulze-Lefert P. (2004). Plant immunity: The origami of receptor activation. Current Biology 14, R22-R24.

Shen F., Huang Z., Zhang B., Wang Y., Zhang X., Wu T., Xu X., Zhang X., Han Z. (2019). Mapping Gene Markers for Apple Fruit Ring Rot Disease Resistance Using a Multi-omics Approach. G3 (Bethesda) 9, 1663-1678.

Shin S., Zheng P., Fazio G., Mazzola M., Main D., Zhu Y. (2016). Transcriptome changes specifically associated with apple (*Malus domestica*) root defense response during *Pythium ultimum* infection. Physiological and Molecular Plant Pathology 94, 16-26.

Shirasu K., Nakajima H., Rajasekhar V. K., Dixon R. A., Lamb C. (1997). Salicylic acid potentiates an agonist-dependent gain control that amplifies pathogen signals in the activation of defense mechanisms. Plant Cell 9, 261-270.

Silva K. J. P., Singh J., Bednarek R., Fei Z., Khan A. (2019). Differential gene regulatory pathways and co-expression networks associated with fire blight infection in apple (Malus × domestica). Hortic Res. 6, 35.

Singh J., Fabrizio J., Desnoues E., Silva J. P., Busch W., Khan A. (2019). Root system traits impact early fire blight susceptibility in apple (Malus × domestica). BMC Plant Biol. 19, 579.

Skadhauge B., Thomsen K. K., vonWettstein D. (1997). The role of the barley testa layer and its flavonoid content in resistance to *Fusarium* infections. Hereditas 126, 147-160.

Sohn K. H., Lee S. C., Jung H. W., Hong J. K., Hwang B. K. (2006). Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. Plant Molecular Biology 61, 897-915.

Spoel S. H., Mou Z., Tada Y., Spivey N. W., Genschik P., Dong X. (2009). Proteasome-Mediated Turnover of the Transcription Coactivator NPR1 Plays Dual Roles in Regulating Plant Immunity. Cell 137, 860-872.

Sweeney T. E., Haynes W. A., Vallania F., Ioannidis J. P., Khatri P. (2017). Methods to increase reproducibility in differential gene expression via meta-analysis. Nucleic Acids Res 45, e1.

Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447–D452.

Thimm O., Bläsing O., Gibon Y., Nagel A., Meyer S., Krüger P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37, 914-939.

Treutter,D. (2006). Significance of flavonoids in plant resistance: a review. Environmental Chemistry Letters 4, 147-157.

Uji Y., Taniguchi S., Tamaoki D., Shishido H., Akimitsu K., Gomi K. (2016). Overexpression of OsMYC2 Results in the Up-Regulation of Early JA-Rresponsive Genes and Bacterial Blight Resistance in Rice. Plant and Cell Physiology 57, 1814-1827.

Usadel B., Nagel A., Steinhauser D., Gibon Y., Bläsing O. E., Redestig H., et al. (2006). PageMan an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics 7, 535.

Van Eck L., Davidson R. M., Wu S., Zhao B. Y., Botha A. M., Leach J. E., et al. (2014). The transcriptional network of *WRKY53* in cereals links oxidative responses to biotic and abiotic stress inputs. Functional & Integrative Genomics 14, 351-362.

Wang K., Zheng X., Zhang X., Zhao L., Yang Q., Boateng N. A. S., Ahima J., Liu J., Zhang H. (2019). Comparative Transcriptomic Analysis of the Interaction between Penicillium expansum and Apple Fruit (Malus pumila Mill.) during Early Stages of Infection. Microorganisms 7, 495.

Wang W. M., Liu P. Q., Xu Y. J., Xiao S. Y. (2016). Protein trafficking during plant innate immunity. Journal of Integrative Plant Biology 58, 284-298.

Wang X., Yan Y., Li Y., Chu X., Wu C., Guo X. (2014). GhWRKY40, a multiple stress-responsive cotton WRKY gene, plays an important role in the wounding response and enhances susceptibility to *ralstonia solanacearum* infection in transgenic *Nicotiana benthamiana*. PLoS One 9, e93577.

Xu J., Li M., Jiao P., Tao H., Wei N., Ma F., et al. (2015). Dynamic transcription profiles of "Qinguan" apple (*Malus x domestica*) leaves in response to *Marssonina coronaria* inoculation. Frontiers in Plant Science 6, 842.

Xu X. P., Chen C. H., Fan B. F., Chen Z. X. (2006). Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell 18, 1310-1326.

Yin Z. Y., Ke X. W., Kang Z. S., Huang L. L. (2016). Apple resistance responses against *Valsa mali* revealed by transcriptomics analyses. Physiological and Molecular Plant Pathology 93, 85-92.

Zhang Q. Y., Gao M., Wu L. W., Wang Y. D., Chen Y. C. (2016). Divergent Expression Patterns in Two Vernicia Species Revealed the Potential Role of the Hub Gene VmAP2/ERF036 in Resistance to *Fusarium oxysporum* in *Vernicia montana*. Genes 7, 109.

Zhu L., Ni W., Liu S., Cai B., Xing H., Wang S. (2017). Transcriptomics Analysis of Apple Leaves in Response to *Alternaria alternata* Apple Pathotype Infection. Frontiers in Plant Science 8, 22.

Zhu Q. H., Shan W. X., Ayliffe M. A., Wang M. B. (2016). Epigenetic Mechanisms: An Emerging Player in Plant-Microbe Interactions. Mol. Plant Microbe Interact. 29, 187–196.

Experiment 2

Gaining Insight into Exclusive and Common Transcriptomic Features Linked with Biotic Stress Responses in *Malus x domestica*.

DOI : https://doi.org/10.3389/fpls.2017.01569

1. Introduction

Apple (*Malus x domestica Borkh*) is one of the most important fruit crops in the world, which is having highly nutritional value and is strongly recommended in diet. There are several constrains which affect the cultivation of the apple trees, which can be classified as biotic stress and abiotic stress. The infections due to bacteria, fungus, and virus are severely affecting apple production and threatening grower's profits. Transcriptomic studies have been conducted in *Malus x domestica* (Chen et al., 2014; Kamber et al., 2016) as well as in other crops (Martinelli et al., 2012, 2013; Giovino et al., 2016; Cohen et al., 2019) in order to shed lights on the complex molecular mechanisms of plant-microbe interactions. The identification of important proteins with a key role in gene-gene and protein-protein interaction networks is extremely useful to improve early diagnosis and therapeutic and genetic resistance strategies. RNA interference (RNAi), relatively new technique, which triggers gene silencing typically by double-stranded RNA (dsRNA), has become a significant tool to knockdown target genes in plants as well as in insects (Limera et al., 2017). The first key step towards developing an efficient RNAi-mediated pest control technique is to find suitable target genes. To develop RNAi-mediated pest control methods, it is critical to find suitable target genes and transcriptomes have been reported to be useful genetic resources for high-throughput screening of RNAi target genes (Khraiwesh et al., 2012; Li et al., 2013). Dual RNA-seq is another recent promising approach to study molecular mechanisms of interactions between plant pathogens and their hosts (Kovalchuk et al., 2019).

A comparison of the molecular mechanisms behind different stress conditions allows the discovery of potential candidate genes involved in specific and exclusive plant biotic stress responses. It also allows gaining insight into general and common features linked with disease status. This permits obtaining an early alert of the plant pathological status and addressing the most sustainable management strategies. Considering that transcriptomic studies are performed only in one season, often with no biological replications and in one

specific environment, the importance of performing meta-analysis is getting higher. The presence, of many transcriptomic studies using different techniques (RNA-seq, microarrays, cDNA libraries etc.) for each crop, allows gaining insight into common and specific genes, pathways, and functional gene categories associated with different pathogens and commonly modulated between environmental stresses. It is known how some genes are affected by multiple environmental factors, involved in different metabolic, physiological, developmental, and organ-specific processes. Indeed it is essential to compare transcriptomic data dealing with multiple research objects in order to determine the pattern of expression of each gene in different physiological processes. This will help filtering biotic stress responses from unspecific features related to multiple physiological conditions. In *Malus x domestica*, several transcriptomic studies have been conducted to elucidate important plant physiological and developmental processes such as tree and root architecture, development and morphology, flavonoid pathway, and fruit physiological disorders (Krost et al., 2013; Mellidou et al., 2014; Ferrero et al., 2015; Wang et al., 2015; Li et al., 2016).

The identification of commonalities between similar independent studies to study the same factor would allow identifying which genes are more strongly associated with the subject of the study and focus the functional analysis only on those common findings. A clear discordance between different omic levels is usually observed in integrated approaches due to the fine-tuned molecular mechanisms of gene regulation developed by cells. Transcriptomic findings often did not closely match with miRNAome, proteome and metabolome data. Despite these issues, the extreme progress obtained in the development of sophisticated machines for omic analysis has allowed researchers to generate “omic” data with low budget requirements. How is it possible to extract the most useful information from the huge amount of produced data? How complex has to be the experimental design of these studies to obtain meaningful and trustful information? “Omic” experiments should be considered reliable if replications in different seasons and environments are performed.

Thus, a meta-analysis of all the transcriptomic studies plays a vital role to select the most frequent and most significant differentially expressed genes (DEG) among the complete list of differentially regulated genes.

The aim of this meta-analysis study was to identify key genes and proteins involved in general plant pathological conditions and those involved in specific and unique pattern of

biotic stress responses in *Malus x domestica*. A customized pipeline of meta-analysis was developed, which could be applied to gain insight into similar studies in other crops.

2. Materials and methods

2.1 Search strategy of published study identification for meta-analysis

The published transcriptomic studies in *Malus x domestica* were identified from Scopus and PubMed using the combination of keywords "Transcriptomics" and "malus" or "Transcriptomics" and "apple" in computer-based searches, and were published on or before March 2017. The identified studies were first divided into two major groups (1) "Biotic Stress" and (2) "Others". The studies with the raw data are publically available were only considered for the "Biotic Stress" group. There were total twelve studies related to the purpose of meta-analysis where, six articles related with "Biotic Stress" and the rest six articles in the "Others" group. The raw data of all the "Biotic Stress" group studies were downloaded and performed RNA-Seq analysis using a single analysis pipeline to obtain the differentially expressed genes. The common genes present in the two groups were eliminated during the analysis in order to get more accurate results pertaining to the objective of the study. According to the type of the pathogen, Biotic Stress studies were further divided into three groups (a) ASGV, (b) *Erwinia amylovora* (*E. amylovora*) (c) Fungal pathogens. Also, rest of the studies were divided into three groups: (d) "Tree Architecture" (e) "Fruit" (f) "Root." The complete work flow of this meta-analysis was given in Fig. 2.1.1.

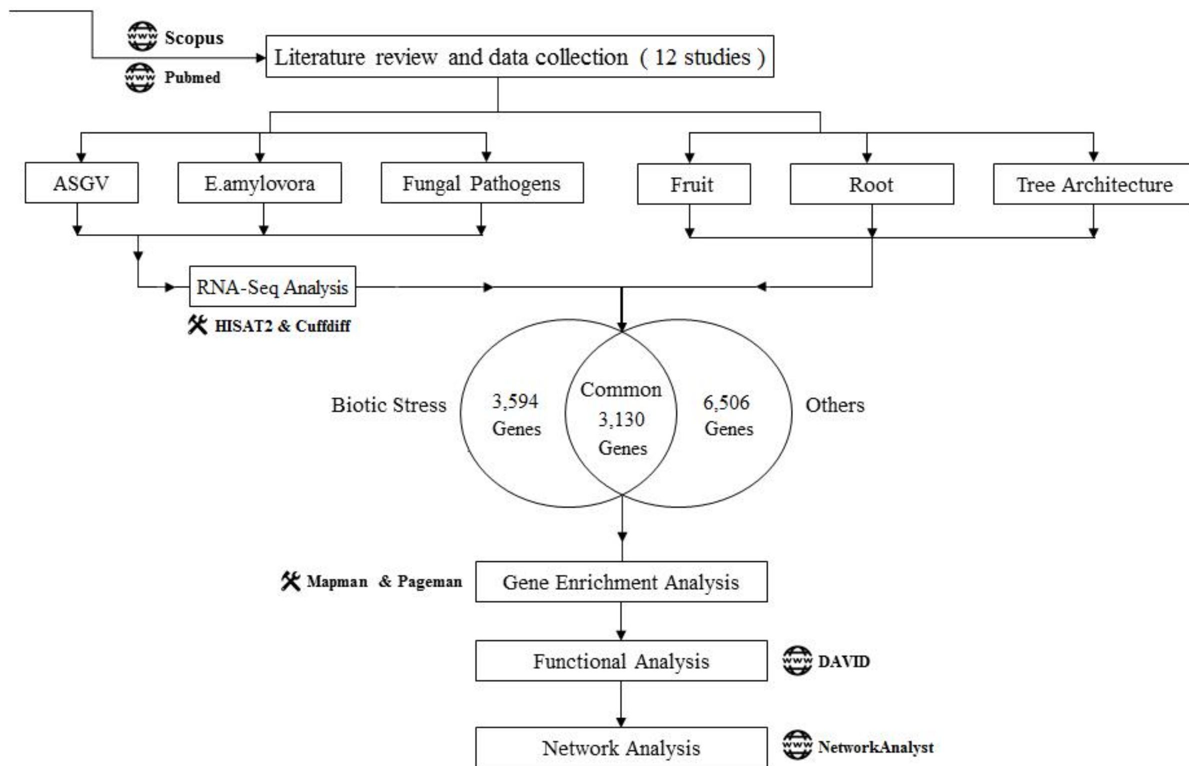


Figure 2.2.1–Work flow of the meta-analysis of the 12 *Malus x domestica* transcriptomic studies. Number of genes (up-regulated and down-regulated) uniquely modulated in biotic stress-related articles and in the rest of the studies was shown. Functional data mining tools were indicated.

2.2 Differentially expressed gene selection and annotation

The *Malus x domestica* genome v1.0 and annotation file were downloaded from Phytozome (https://phytozome.jgi.doe.gov). The Raw files (SRA format) of the six articles dealing with biotic stress responses in *Malus x domestica* were downloaded from NCBI SRA and then converted to FASTQ format using SRA toolkit version 2.3.5. The article Gusberti et al. (2013) contains the differential gene expression information related to “Biotic Stress” and “Leaf development” in *Malus x domestica* and downloaded the RAW files only for the samples dealing with “Biotic Stress.” The Raw reads were filtered to obtain high-quality clean reads by trimming low-quality bases followed by adaptor sequence removal using cutadapt version 1.8.1. The pre-processed reads were mapped to the *Malus x domestica* genome v1.0 with HISAT2 version 2.0.5 (Kim et al., 2015) using default parameters. The identification of differentially expressed genes was performed using Cuffdiff algorithm in Cufflinks version

2.2.1 pipeline with default parameters. The up and down regulated gens obtained with fold change cutoff (log2 FC >1 or log2 FC <−1) and p-value < 0.05 were only considered for the meta-analysis. The details of the selected articles are given in Table 2.1. The sample information, SRA IDs and alignment information are given in Table2.2.The sample comparison plan and the counts of differentially expressed genes obtained are given in Table2.3.

No	**Article**	**Objective**	**Tissue**	**DEGs**		**Pathogen**	**Group**
				Up	**Down**		
1	Chen et al., 2014	*Apple stem grooving virus*	Shoot	263	404	Virus	ASGV
2	Kamber et al., 2016	Responses to *Erwinia amylovora*	Flower	147	108	Bacteria	*F. amylovora*
3	Yin et al., 2016	Resistance to *Valsa mali*	Twig	247	22	Fungi	Fungal Pathogens
4	Zhu et al., 2017	Response to *Alternaria alternate*	Leaf	358	621	Fungi	
5	Shin et al., 2016	Response to *Pythium ultimum*	Root	355	923	Fungi	
6	Gusberti et al., 2013	Resistance to *Venturia*	Young Leaf	751	567	Fungi	
			Mature Leaf	244	208	Fungi	
		Leaf Development	Young & Mature Leaf	1,655	4,089	-	Tree Architecture
7	Krost et al., 2013	Tree architecture	Shoot	315	697	-	
8	Ferrero et al., 2015	Fruitlet Abscission	Fruit	470	37	-	Fruit
9	Mellidou et al., 2014	Flesh browning disorder	Fruit	44	26	-	
10	Wang et al., 2015	Flavonoid content	Fruit	88	25	-	
11	Petersen et al., 2015	Root architecture	seed, shoot, leaf	215	168	-	Root
12	Li et al., 2016	Root growth	Root	63	120	-	

Table 2.1–Analyzed articles, objective of the studies, tissue number of up- or down-regulated genes, and assigned group.

No.	Study	Sample Name	SRA ID	Read orientation	Total number of reads	Alignment %
1	Chen et al., 2014	Control	SRR1089478	Single	7,591,042	76.05%
		Infected	SRR1089477	Single	7,430,428	85.47%
2	Kamber et al., 2016	Control	ERR1189573	Single	3,553,043	65.45%
		Inoculated	ERR1189574	Single	12,839,290	59.43%
3	Yin et al., 2016	Control	SRR1917391	Paired	60,300,888	89.07%
		Infected	SRR1063452	Paired	96,640,082	78.73%
4	Zhu et al., 2017	0HPI	SRR4431586	Single	11,960,715	89.47%
		72HPI	SRR4431634	Single	12,360,687	84.78%
5	Shin et al., 2016	0T	SRR1603673 SRR1603674 SRR1603676 SRR1603677 SRR1603678 SRR1603675	Single	21,879,041	86.68%
		96M	SRR1603721 SRR1603722 SRR1603723 SRR1603724 SRR1603725	Single	20,000,000	87.00%
6-1	Gusberti et al., 2013	1.I.96.1_E1	ERR313218	Paired	65,987,624	83.78%
		1.I.96.2_E1	ERR313239	Paired	74,738,806	82.20%
		1.I.96.3_E1	ERR313221	Paired	83,789,050	80.64%
		1.N.96.1_E1	ERR313236	Paired	80,149,630	77.72%
		1.N.96.2_E1	ERR313225	Paired	93,291,312	92.30%
		1.N.96.3_E1	ERR313238	Paired	94,326,608	88.12%
6-2	Gusberti et al., 2013	7.I.96.1_E1	ERR313235	Paired	69,912,968	92.10%
		7.I.96.2_E1	ERR313219	Paired	90,906,390	88.71%
		7.I.96.3_E1	ERR313230	Paired	64,317,912	78.92%
		7.N.96.1_E1	ERR313229	Paired	76,635,396	78.88%
		7.N.96.2_E1	ERR313233	Paired	75,873,582	85.86%
		7.N.96.3_E1	ERR313220	Paired	76,611,594	90.96%

Table 2.2– SRA and alignment information.

Comparison	Total DEGs	Total Up-regulated genes	Total Down-regulated genes
Control vs. Infected	667	263	404
Control vs. Inoculated	255	147	108
Control vs. Infected	269	247	22
0HPI vs 72HPI	979	358	621
0T vs 96M	1,278	355	923
(1.N.96.1_E1,1.N.96.2_E1,1.N.96.3_E1) vs. (1.I.96.1_E1,1.I.96.2_E1,1.I.96.3_E1)	1,318	751	567
(7.N.96.1_E1,7.N.96.2_E1,7.N.96.3_E1) vs. (7.I.96.1_E1,7.I.96.2_E1,7.I.96.3_E1)	452	244	208

Table 2.3– Sample comparison plan and total number of DEGs obtained.

The up and down regulated gens with fold change cutoff (log2 FC >1 or log2 FC <−1) and p-value < 0.05 were collected from the rest six articles dealing with transcriptomic studies in *Malus x domestica* other than "Biotic Stress" (Table 2.1).

All the differentially expressed gene ids were annotated using the *Malus x domestica* genome v1.0 mapping file downloaded from the Phytozome database (https://phytozome.jgi.doe.gov). The common and unique genes among different groups were identified. The common genes present in "Biotic Stress" and "Other" groups were eliminated from "Biotic Stress" gene list and were considered for the rest of the analysis. I wrote custom made perl scripts for the selection of gens and mapping.

2.3 Gene enrichment analysis

I used MapMan (Thimm et al., 2004) with the *Malus x domestica* mapping file (Mdomestica_196.txt) (http://mapman.gabipd.org/) to map the gene ids and visualize the metabolic overview, hormone regulation, transcription factors, and protein targeting of the Biotic stress gene sets (a) ASGV, (b) *E. amylovora*, (c) Fungal Pathogens.

The PageMan (Usadel et al., 2006) analysis plugin of MapMan was used to visualize differences among metabolic pathways using Wilcoxon tests, no correction, and an over-representation analysis (ORA) cutoff value of 3. I considered all the differentially expressed genes present in all 6 gene sets for the PageMan analysis: (a) ASGV (b) *E. amylovora* (c) Fungal Pathogens (d) Tree Architecture (e) Fruit (f) Root.

2.4 Functional analysis

All the homologous TAIR IDs of the Biotic Stress genes were searched against the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (Huang et al., 2009) Web server (https://david.ncifcrf.gov/). The unique list of TAIR IDs for each group were collected and used for the DAVID pathway analysis. The gene ontology information related to Biological process were extracted (FDR cutoff = 0.05) from the DAVID result (Table 2.4).

Group	Up/Down	GO ID	GO Term	Count	pval	Fold Enrichment	FDR
Fungal Pathogens	Down	GO:0006355	regulation of transcription, DNA-templated	90	1.95E-05	1.555852518	1.69E-03
	Down	GO:0015979	Photosynthesis	15	2.49E-05	3.981704692	3.79E-02
	Down	GO:0007018	microtubule-based movement	12	1.11E-06	6.868440594	3.79E-02
	Up	GO:0055114	oxidation-reduction process	64	1.46E-05	1.752317802	2.97E-02
Tree Architecture	Down	GO:0006511	ubiquitin-dependent protein catabolic process	50	3.99E-08	2.312224705	1.65E-03
	Down	GO:0006886	intracellular protein transport	40	9.92E-07	2.319412969	6.99E-03
	Down	GO:0042254	ribosome biogenesis	31	6.76E-06	2.443142217	1.56E-02
	Down	GO:0000027	ribosomal large subunit assembly	14	4.95E-06	4.47646703	0.0073212
	Down	GO:0000059	protein import into nucleus, docking	11	4.41E-06	5.862040158	6.62E-05
	Down	GO:0018279	protein N-linked glycosylation via asparagine	8	4.21E-06	8.952934059	1.12E-02
	Up	GO:0015996	chlorophyll catabolic process	6	1.02E-05	18.20767717	2.22E-02
Root	Down	GO:0009734	auxin-activated signaling pathway	11	1.31E-08	12.74993734	1.09E-02
	Up	GO:0009734	auxin-activated signaling pathway	9	8.24E-06	8.762684211	8.22E-03
	Down	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	8	1.92E-06	13.55238095	1.62E-05

Table 2.4– Significantly regulated biological processes in the analyzed transcriptomic studies (FDR < 0.05).

2.5 Protein-protein interaction network

Individual data annotation and analysis were performed using NetworkAnalyst (Xia et al., 2014), a web-based tool for protein–protein interaction network analysis and visual exploration. The unique list of homologous TAIR IDs of each "Biotic Stress" groups were uploaded and mapped against the STRING interactome database with default parameters (confident score cutoff = 900 and with experimental evidence) provided in NetworkAnalyst. I selected "Minimum Network" to simplify the network and to study the key connectivities. The common genes present in the three biotic stress groups (a) ASGV, (b) *E. amylovora*, (c) Fungal pathogens were highlighted in Figure 2.5.1. The genes present in each biotic stress groups and the common genes among them were highlighted separately in Figure2.5.2 andFigure 2.5.3.

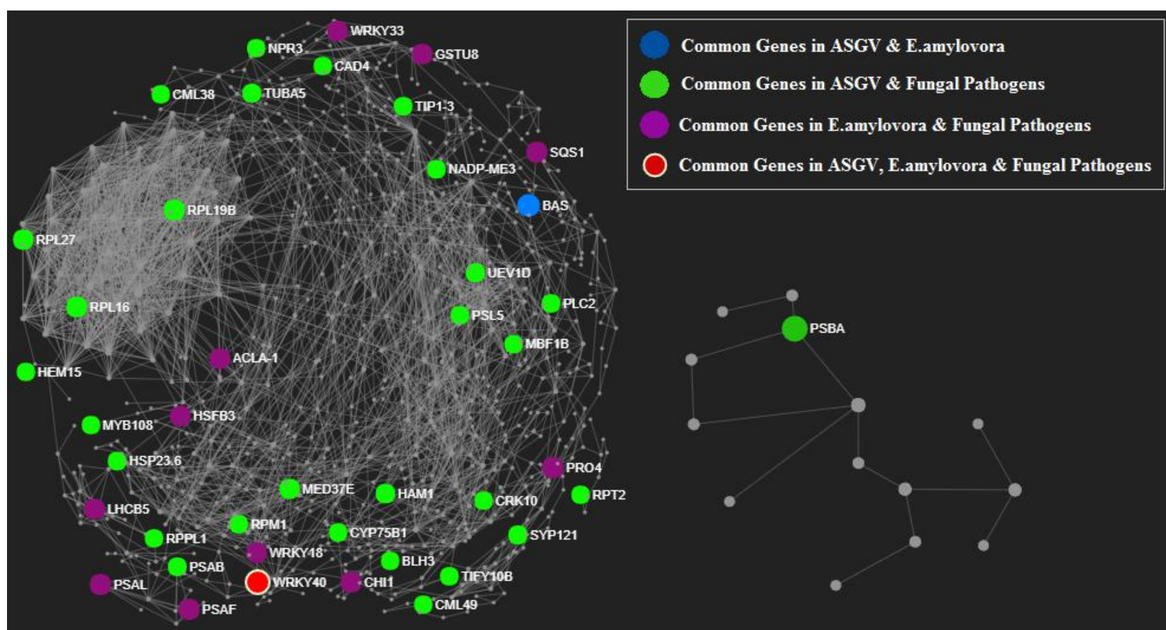


Figure 2.5.1–Protein-protein interaction network analysis predicted in *Malus x domestica* based on *Arabidopsis* knowledgebase. Proteins encoded by transcriptionally modulated genes were shown in different color basing on the type of pathogens. Proteins encoded by genes commonly modulated by 2 of 3 types of pathogens were shown in red.

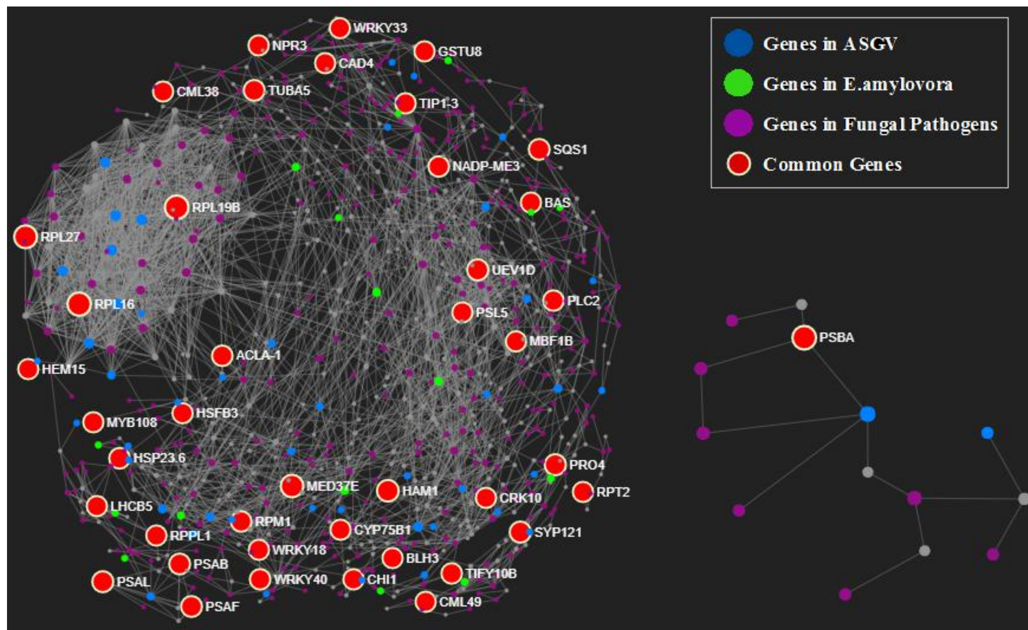


Figure 2.5.2–Proteins encoded by transcriptionally modulated genes were shown in different color basing on the type of pathogens. The commonly modulated (2 of 3 type of pathogens) genes were shown in red.

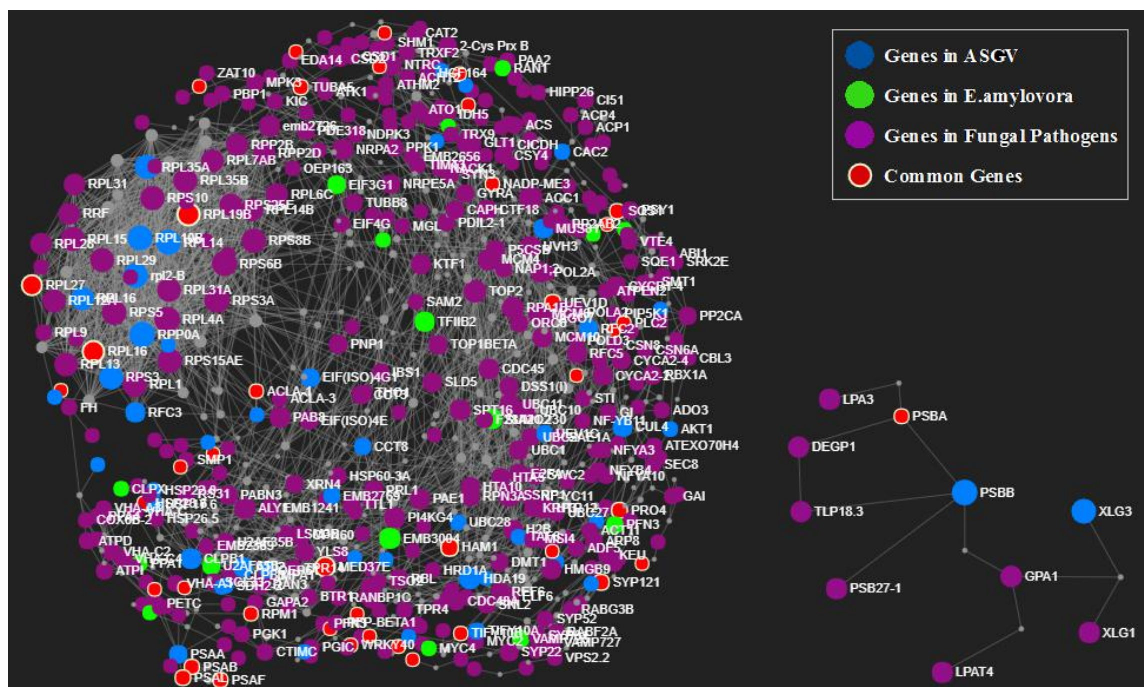


Figure 2.5.3–Protein-protein interaction network analysis predicted in *Malus x domestica* based on *Arabidopsis* knowledgebase. Proteins encoded by transcriptionally modulated genes were shown in different color basing on the type of pathogens.

3. Results

In this study, I collected a total 12 transcriptomic studies published in *Malus x domestica* in Pubmed and Scopus databases and compared the significantly regulated genes in each of these research subjects (p-value < 0.05, log2 FC >1 or log2 FC <−1).

Article details, titles, analyzed tissues, and numbers of up- and down-regulated genes were listed in Table 2.1. The first six articles deal with biotic stress responses (*Apple stem grooving virus*, *E. amylovora*, and fungal pathogens) (Gusberti et al., 2013; Chen et al., 2014; Kamber et al., 2016; Shin et al., 2016; Yin et al., 2016; Zhu et al., 2017). The other six studies were dealing with the understanding of molecular mechanisms of fruitlet abscission, flesh browning disorder (physiopathological fruit disorder), flavonoid biosynthesis in fruit, tree and root architecture, growth and morphology (Krost et al., 2013; Mellidou et al., 2014; Ferrero et al., 2015; Petersen et al., 2015; Wang et al., 2015; Li et al., 2016). One study was divided in three datasets: responses to *Venturia inaequalis* in young and mature leafs and gene expression involved in leaf development (Gusberti et al., 2013). Although great data variability was observed between the different studies regarding the number of significantly regulated genes, a strict p-value cut-off was kept in order to increase data reliability. The developed meta-analysis workflow was shown in Figure 2.2.1. Total 13,230 genes were analyzed: 5,215 were up-regulated, 8,015 were down-regulated. Biotic stress-related works significantly regulated 5,218 while the rest of articles related to fruit processes, tree and root architecture and leaf development affected the expression of 8,012 genes. A part of these genes were commonly modulated (3,130). Among the two main categories (biotic stress and "others"), I independently analyzed subgroups of studies. Biotic stress was divided in responses to *Apple Stem Grooving Virus*, *E. amylovora* and fungal pathogens. The "others" group was divided in fruit processes, root and tree architecture. All these transcriptomic analysis were functionally mined with an integrated approach composed by gene set enrichment analysis (Pageman; Usadel et al., 2006), pathway and gene ontology analysis (DAVID; Huang et al., 2009), gene visualization (MAPMAN; Thimm et al., 2004), network analysis (NetworkAnalyst; Xia et al., 2014).

3.1 Gene set enrichment analysis

Gene set enrichment analysis showed that photosynthesis was repressed at transcriptional level in two biotic stress-related studies while it was enhanced during leaf growth and development. The gene categories related to primary metabolism such as photosynthetic light reactions, Calvin cycle, major carbohydrate metabolism were expressed in Tree architecture. Trehalose pathway induction was linked with modifications on *Malus x domestica* tree architecture. RNA processing was more generally expressed by *E. amylovora* while these genes were not affected in response to other biotic stresses. Indeed, *Alternaria alternata* drives more the upregulation of genes involved in hormone metabolism ethylene, biotic stress, and protein degradation signaling compared to the other pathogens whereas it repress protein synthesis. As expected, root architecture was highly linked with the induction of auxin signaling and responsive genes as previously reported (Overvoorde et al., 2010).

Different hormone categories were linked with specific research subjects. Gibberellin-related pathways were inhibited by the attack of *Pythium ultimum*. *Apple stem grooving virus* mainly repressed jasmonic acid-mediated responses. Gene encoding key players in biotic stress responses was linked with modifications of shoot architecture as well as with attacks of A. alternata, *E. amylovora*, and *Venturia*. Abiotic stress-related genes were up-regulated by tree architecture modifications while an increase of induced genes involved in redox detoxifying pathways (ascorbate, glutathione-s-transferase, thioredoxin,) and biodegradation of xenobiotics were related to leaf development.

The different categories of transcription factors were exclusively linked to different studies. Basic helix-loop-helix transcription factors were expressed by *P. ultimum*, Zn C2-CO-like was enhanced by leaf development and Zn C2-DOF was inhibited in fruits. MADS-box, SNF7, and MYB-related were also induced by leaf development processes. Homeobox transcription factor family proteins were more down-regulated by *Valsa mali* infections. The bZIP transcription factor family proteins were more expressed in Root.

It is worthy to notice that different receptor kynases were involved in different physiological processes: receptor kinases were mostly repressed in fruit architecture changes. In protein synthesis, process expressed more in virus as well as *P. ultimum* infections. Protein degradation repression was linked to shoot development. Leucine rich repeat XI repression was linked with root development. DUF26 category upregulation was clearly linked to tree

architecture. Aminoacid transport induction was associated only with root architecture processes. As expected, major transport related (sugar, nitrate, sulfate, phosphate, and nucleotides transport) genes were induced by leaf development processes.

The aim of this work was to focus on the biotic stress responses in order to identify genes related to general mechanisms of plant responses to biotic attacks and genes specifically modulated by different types of pathogens (virus, bacteria, and fungal pathogens). Indeed, I visualize only those significantly regulated genes in each of the three pathogen groups eliminating those genes that were also affected by other physiological processes and related to unspecific plant responses. Although the list of fungal pathogen-regulated genes were higher than the other two types of pathogens and this may disturb the meta-analysis it is clear that specific pathogens and some gene categories were specific for each pathogen.

Anthranilate N-hydroxycinnamoyl flavonoid related genes were more induced by ASGV. Alcohol-dehydrogenases flavonoid-related genes, nucleotide-related genes were more induced fungal pathogens. AGSV repressed few specific genes involved in phenylpropanoids, aminoacids primary metabolism (TCA, lipids, carbohydrates) (Figure 3.1.1). A low amount of genes were commonly modulated by the different stresses. They were involved in photosynthesis, minor CHO, and phenylpropanoid pathways.

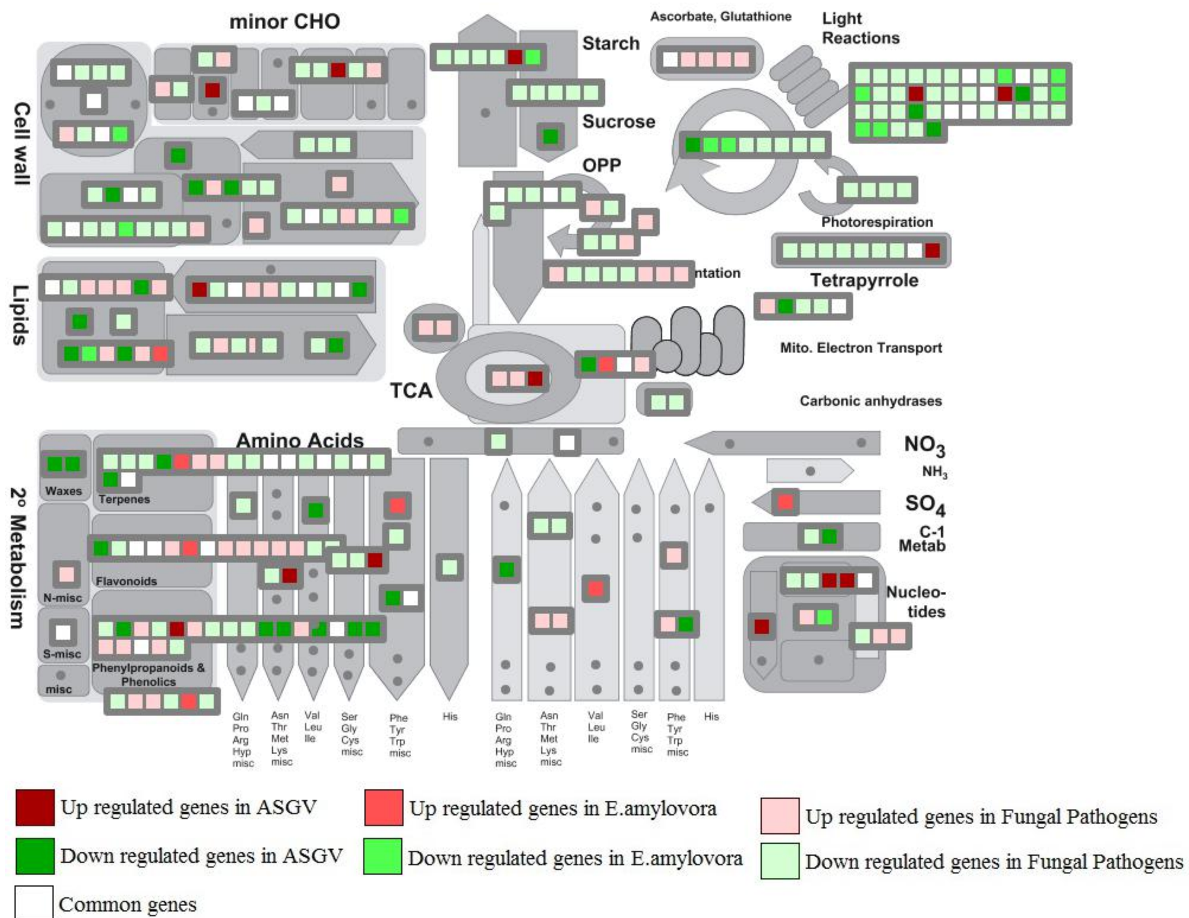


Figure 3.1.1– Mapman overview of biotic stress-related genes in response to *apple stem grooving virus* (ASGV), *Erwinia amylovora*, fungal pathogens and commonly regulated in at least 2 of 3 types of pathogens. Only those genes that were uniquely modulated by biotic stress were indicated. Those affected in at least one of the rest of the transcriptomic studies were eliminated.

3.2 Biological process enrichment analysis

DAVID software was used to identify which gene ontologies (biological process, cellular component, molecular function) were significantly affected by six groups of transcriptomic works (responses to *apple stem grooving virus*, *E. amylovora* and fungal pathogens, fruit responses, root morphology, and architecture and tree architecture). Related to biological processes, only responses to fungal pathogens, tree architecture and root responses showed significantly modulated biological processes (FDR < 0.05). Fungal pathogens repressed DNA-templated transcription regulation, photosynthesis, and microtubule-based movement

while tree architecture repressed ubiquitin-dependent protein catabolic process, intracellular protein transport, protein N-linked glycosylation via asparagine, protein import into nucleus, docking, ribosomal large subunit assembly and ribosome biogenesis (Table 2.4). The chlorophyll catabolic process and oxidation-reduction process showed more expression in Tree architecture studies. In root, transcriptomic studies showed a significant up and down regulation in auxin-activated signaling pathway and showed inhibition of transmembrane receptor protein tyrosine kinase signaling pathway.

3.3 Hormone-related pathways

Infection of fungal pathogens enhanced expression of isoforms of *ILR1*, *ATB2*, and has opposite effects on the expression of different aldo/keto reductase (Figure 3.3.1). Two key brassinosteroid genes were down-regulated by fungal pathogens and two were commonly modulated between different biotic stresses. Several genes involved in ethylene biosynthesis and signaling were enhanced by fungal pathogens such as oxidoreductase *2OG-Fe (II)* oxygenases and *ERF1*. Many ethylene-related genes were commonly regulated by different biotic stress studies. Jasmonic acid-related genes were mostly repressed by *apple stem grooving virus* (OPDA reductase3, allene oxide synthase). *E. amylovora* enhanced the expression of a gene involved in salicylic acid response. Interestingly gibberellin-related and ARA-related genes were mostly repressed by fungal pathogens although four key GA-related genes were commonly modulated between different biotic stresses.

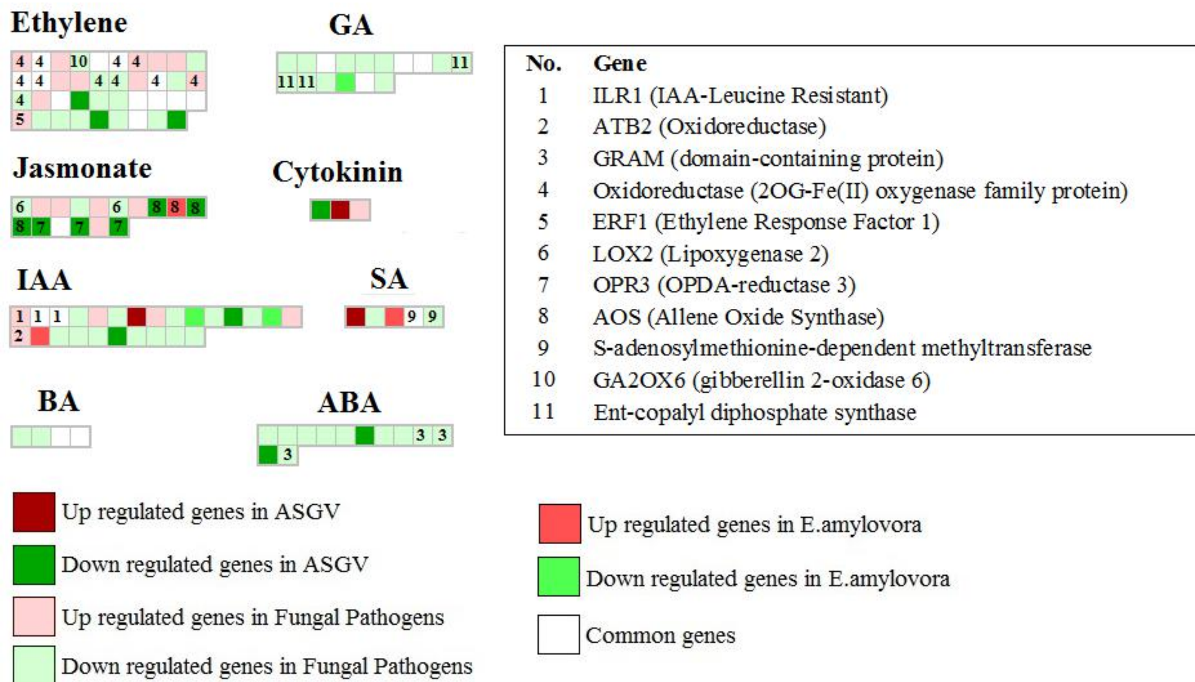


Figure 3.3.1–Gene expression changes involved in hormone-related pathways in response to *Apple stem grooving virus*, *Erwinia amylovora*, fungal pathogens, and commonly modulated in at least 2 of the 3 types of pathogens. Some key genes were indicated.

3.4 Secondary metabolism

The expression of genes involved in secondary metabolism was peculiarly modulated by the different analyzed transcriptomic studies (Figure 3.4.1). Leaf development up-regulated some genes of the non-MVA pathway (*CLA1*, *ISPF*, *CSB3*), it repressed other genes such as a zinc ion binding, a *GGPS1*, some 2-dehydro-3-deoxyphosphoheptonate aldolases and a shikimate synthase. On the other hand, transferases, a hydroxycinnamoyl-coa shikimate transferase, and a cinnamyl-alcohol dehydrogenase were enhanced by fungal pathogens. *E. amylovora* induced a chalcone synthase. Several genes involved in dehydroflavonol and carotenoid pathways were up-regulated during leaf development.

Non MVA pathway
Shikimate pathway
MVA pathway
Terpenoids
Tocopherol
Phenlypropanoids
Simple phenols
Lignin and lignans
Alkaloids-like
Carotenoids
Flavonoids
Chalcones
Isoflavonoids
Flavonols
Flavonones
Dihydroflavonols
Anthocyanins
Glucosinolates
Cyanogenic glycosides
Sulfur containing
Betains
Wax

Up regulated genes in ASGV
Up regulated genes in E.amylovora
Down regulated genes in ASGV
Down regulated genes in E.amylovora
Up regulated genes in Fungal Pathogens
Common genes
Down regulated genes in Fungal Pathogens

Figure 3.4.1–Genes significantly regulated in biotic stress responses and involved in secondary metabolism and grouped based on exclusively or commonly expression in response to the three types of pathogen attacks.

3.5 Protein targeting and transcription factors

Few key genes involved in secretary pathways differentially affected by the different biotic stresses (Figure 3.5.1). Nuclear transport factor 2 and VPS28-1 (vacuolar protein sorting-associated protein) were repressed by fungal pathogens while a signal peptidase subunit family protein was induced by *E. amylovora*. Specific transcriptomic changes were observed

in relation to different types of pathogens. Fungal pathogens mostly inhibited MYB-related genes as well *WRKY*s and TCP transcription factors. *Apple stem grooving virus* inhibited seven genes encoding bHLH and two WRKY members while it induced two trihelix members (Figure 3.5.2). *E. amylovora* induced one gene encoding bHLH and repressed one C2C2-CO-like gene.

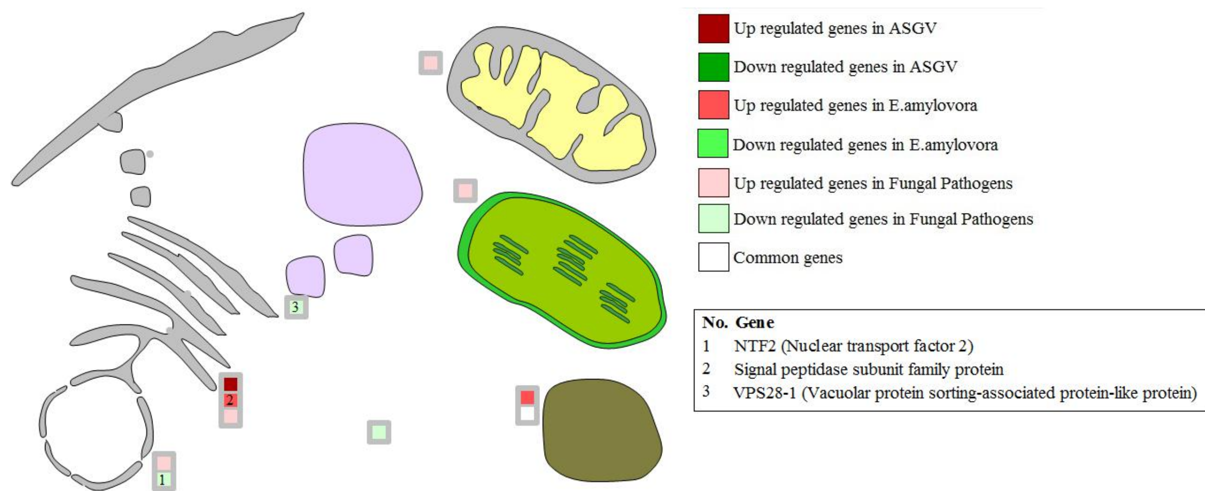


Figure 3.5.1–Gene expression changes involved in targeting-related genes in response to *Apple stem grooving virus*, *Erwinia amylovora*, fungal pathogens, and commonly modulated in at least 2 of the 3 types of pathogens.

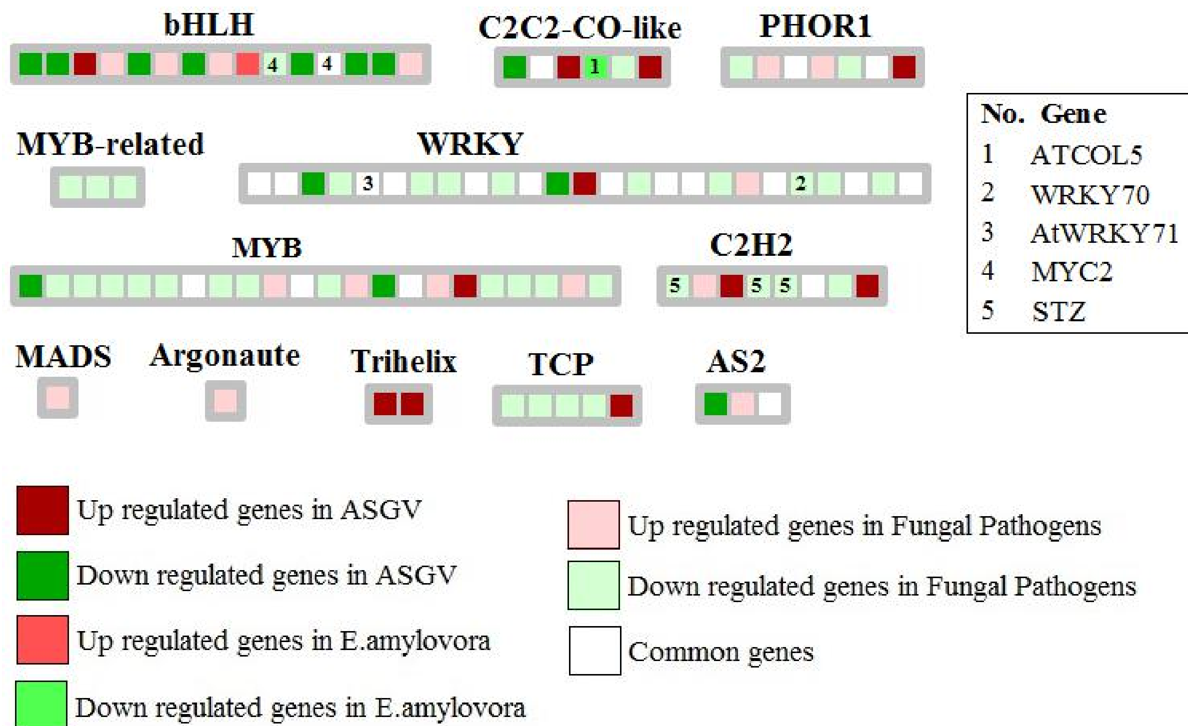


Figure 3.5.2–Significantly regulated transcription factor genes in response to *Apple stem grooving virus*, *Erwinia amylovora*, fungal pathogens, and commonly modulated in at least 2 of the 3 types of pathogens. Only few key TF categories were shown. Some key genes were also indicated.

3.6 Protein-protein interaction network analysis

A protein-protein interaction (PPI) network analysis was predicted in *Malus x domestica* based on *Arabidopsis* knowledgebase (Figure 2.5.1, Figure 2.5.2 and Figure 2.5.3). The list of biotic stress-related genes was determined. Those genes which were significantly affected in response to the three types of pathogens but not in the rest of transcriptomic studies were considered. Those biotic stress-related genes that showed to be modulated by other physiological processes were eliminated and considered to be unspecific. At the end, only the PPI network of the biotic stress-related genes and their partners were shown. The aim of this analysis was to identify some small interactive networks specific to each pathogens. GPA1, LPAT4 and XLG1 were closely connected and involved fungal pathogen responses (Figure 2.5.3). The protein network contain 28 proteins present in both ASGV and fungal pathogens, 11 proteins common between *E. amylovora* and fungal pathogens and one protein (BAS) in

common between ASGV and *E. amylovora*. WRKY40 protein is the only one protein present in the network, which is present in all three biotic stress groups (Figure 2.5.1).

4. Discussion

The large number of transcriptomic works published in plants really requires more meta-analysis studies that would identify common and specific features (genes, gene categories, pathways) linked with the different object of studies (plant developmental, agronomic, and environmental responses). Plant transcriptomic data are highly variable depending of different environmental conditions and gene expression is finely modulated by a high number of variables such as timing, genotypic differences, environmental factors and experimental conditions, tissues and their developmental stages. Here, I compared 12 transcriptomic studies in *Malus x domestica* in order to deliver functional genomic information linked with common or exclusive molecular responses to specific types of biotic stresses. In order to identify only those features related to biotic stresses, I also used RNA-seq data related to other apple physiological processes. The aim was to perform a comparison analysis among transcriptomic datasets clarifying the role of key genes previously identified and shade lights on the different crosstalk played by important biological molecules such as hormones. Different transcriptomic studies are generally performed using different transcriptomic platforms and using different experimental design. In order to compare them, it is necessary to use at least the same bioinformatic pipeline. Meta-analysis in *Malus x domestica* is important to create a database of curated transcriptomic data that could be used also by the scientific community working on other crop species.

The repression of photosynthetic pathways at transcriptomic level in response to biotic stresses is a feature widely seen in previous transcriptomic studies (Martinelli et al., 2012, 2013). My meta-analysis study showed that, among biotic stresses, fungal pathogens strongly inhibited primary metabolism genes. This evidence agrees with data obtained with imaging methodologies that analyzed chlorophyll and multicolor fluorescence. These published data demonstrated their possible application in improving early detection of infections of virus, bacteria, and fungi (Barón et al., 2016). The transcriptomic data of the meta-analysis related to secondary metabolism confirmed findings obtained with imaging methodologies. The integration of different techniques is essential to drive pre-symptomatic stress detection. The

reduction of photosynthesis has been observed in virus-infected leaves at symptomatic level (Pérez-Bueno et al., 2006; Pineda et al., 2010) and even before symptoms appears (Chaerle et al., 2007). Interestingly, the integrated use of imaging and statistical analysis was used to determine infections of *P. syringae* in *Arabidopsis* before symptoms were visible (Berger et al., 2007). Fluorescence signals were also increased in sugar beet infected by powdery mildew and authors concluded that fluorescence indices could be considered as good indices of stress conditions (Leufen et al., 2014). Significant changes in photosynthetic activities linked with fungi infections are also spatially and temporally determined (Barón et al., 2016). Infections of bean leaves by rust fungi have been linked with changes in fluorescence induction kinetics (Peterson and Aylor, 1995).

Experimental algorithms have been developed to determine differences between affected and unaffected plants treated with phytotoxins of *Alternaria brassicae* (Soukupova et al., 2003). The use of molecular and phenotypic stress indicators would allowing manage pathogenesis and guiding effective management procedures of biotic stresses in plants. These techniques highlighted the role of pathogen in repressing photosynthetic performance and affect secondary metabolism. Indeed, this let me to speculate that the integration of meta-analysis of transcriptomic works with the data obtained by techniques of chlorophyll measurements may improve both field and greenhouse management of plant diseases. A complementary use of molecular, remote sensing, and volatile sensor devices have shown to efficiently contribute in the early diagnosis of plant diseases and disorders (Dandekar et al., 2010). The use of these innovative integrated approaches represents the new frontier of plant pathology (Martinelli et al., 2015).

Auxins are considered the most regulator hormones of plant development (Taylor-Teeples et al., 2016). Lateral root development is one of the most well-known organogenesis process mediated by auxins. Auxin Response Factor (ARF) transcription factors are known to be key players in the auxin-mediated regulation of root development. Indeed, ARFs have been found be repressed by the interactions with (Aux/IAAs) repressor proteins and the corepressor Topless. Proteins of the TIR1/AFBs bind auxins in a complex with the Aux/IAAs controlling phyllotaxy. *AFB5*, *TIR1*, *F-box* were up-regulated on grafted apple and linked to root growth (Li et al., 2016) and they were not involved in the other studied physiological processes in *Malus x domestica*. This evidence highlighted their exclusive role in root development and growth. My meta-analysis found out that, among the 12 analyzed *Malus x domestica* studies,

root development processes uniquely induced GRAM-domain proteins. The GRAM domain has a length of 70 amino acids that is usually present in membrane-associated proteins and in glucosyltransferases (Doerks et al., 2000). Although some functions of these proteins remain unclear, the function of this domain seems to be linked with membrane-associated processes such as intracellular binding signaling pathways (Doerks et al., 2000).

Transcriptomic responses to *E. amylovora* (Kamber et al., 2016) showed that this pathogen is more linked with gibberellin response than the other studies as shown by the upregulation of four *2OG-Fe(II)* oxygenase, *GA2OX6* and the repression of others (*GASA4* and unknown genes). The role of gibberellins in response to fire blight has been previously reported (Maxson and Jones, 2002). Indeed apple trees were treated with prohexadione calcium (Apogee) and trinexapac-ethyl (Palisade) well-known inhibitors of gibberellin biosynthesis. This work was showed to be effective in enhance resistance to *E. amylovora* (Maxson and Jones, 2002). This effect was mediated by a reduction of tree growth. However, in the meta-analysis, I observed more GA-related genes modulated by fungal pathogens instead of *E. amylovora*. More studies are needed to define the role of gibberellins in plant responses to *E. amylovora*.

Interestingly jasmonic acid-mediated responses were generally repressed by leaf development process (Gusberti et al., 2013; Noir et al., 2013) and *apple stem grooving virus* infections (Chen et al., 2014). This latter evidence was expected since viruses are considered hemibiotrophic pathogen. Jasmonic acid (JA) and ethylene (ET) are critical for inducing immediate and effective responses against necrotrophs (Glazebrook, 2005) and they are usually repressed by Salicylic acid-mediated responses (Pieterse et al., 2009). My meta-analysis highlighted that *A. alternata* infections showed to downregulate *LOX2* in a peculiar way (Zhu et al., 2017). *LOX2*, requires the expression of the F-box protein COI1 (CORONATINE INSENSITIVE1) that forms a ternary complex with JAZ repressor proteins (Zander et al., 2010). This let me to speculate that this might be detrimental for the infected *Malus x domestica* tree.

Two genes were commonly modulated by different biotic stresses agreeing with published literature confirming the important role of brassinosteroid in hormonal crosstalk in plants in responses to biotic stresses. Brassinosteroids have been known to be important player of biotic and abiotic stresses, although their mechanisms are still not well-elucidated. A

homeodomain transcription factor OsBIHD1 is known to be involved in biotic and abiotic stress responses. The overexpression of this gene or its deficiency modulated the expression of several brassinosteroid-related genes causing brassinosteroid insensitivity (Liu et al., 2017). Indeed, the function of this gene seems to modulate the trade-off between resistance and growth by regulating brassinosteroid-ethylene pathway (Liu et al., 2017). In addition, it is worthy to notice that a squalene monoxygenase and squalene epoxidase3 were induced exclusively by *Pythium* infections among the 13 analyzed studies. In contrast, a key positive regulator was repressed by *apple stem grooving virus*. These data agreed with previous data that showed how the silencing of a N. benthamiana squalene synthase, an important player of phytosterol biosynthesis, counteracted non-host resistance of *Pseudomonas syringae* and *Xanthomonas campestris*, increasing the growth of the host pathogen *P. syringae pv tabaci* by enhancing nutrient efflux into the apoplast. In addition, squalene epoxidase was induced in *Calendula tropicalis* by *Aspergillus niger* and this was linked with enhanced ginsenosides biosynthesis.

Extracting the data published by Gusberti et al. (2013) and dividing them in two datasets (one related to *Venturia* infection and one related to leaf development), I observed an increased expression of detoxifying pathways when leaves are developing and this implies that chemical defense pathways are induced during ontogenetic development against xenobiotic agents. As far as it concerns, different ontogenetic, development, and physiological process activated specific classes of transcription factors. This evidence could be helpful in elucidating the diverse gene regulatory networks modulating plant responses to different pathogen attacks. This will allow the development of specific strategy of genetic resistance to different pathogens.

The meta-analysis showed that 12 genes encoding WRKYs were commonly modulated between different biotic stresses. Their modulation would be important to create genotypes resistant to the presence of multiple pathogens. Fungal pathogens mostly repressed *WRKY* genes implying that there might be a mechanism of repression of beneficial plant biotic-related genes. *WRKY*s represent a large family of transcription factors mostly found in plants with a key role in stress signaling among the several role where they are involved (Jiang et al., 2017). More than 100 and almost 200 *WRKY* superfamily members were discovered in Glycine max and Oryza sativa, (Rushton et al., 2010; Fan et al., 2015). Their expression is typically up-regulated in plants when they are subjected to a great variety of stresses and they

are activated by stress signals such as salicylic acid (SA) or other molecules. Their expression is rapid, transient and it is tissue-specific (Jiang et al., 2017). The identification of specific *WRKY*s modulated by different pathogens and abiotic factors would allow addressing the genetic improvement to develop genotypes resistant to agronomical limiting factors. The complex network of protein-protein interactions may be visualized using software such as bioconductor package of R, Graphviz, Cytoscape. The main aim of this analysis was to identify which highly interactive proteins are specifically or commonly modulated by each kind of the considered biotic stresses. Among them, I pointed my attention of *WRKY*s such as *WRKY18*, *WRKY33*, and *WRKY40*. Interestingly, the PPI network showed that *WRKY40* was affected by all three kind of biotic stresses. Physical and functional interactions have been reported between *WRKY18*, *WRKY40*, and *WRKY60* in response to pathogen infection in *Arabidopsis thaliana* (Xu et al., 2006). The PPI network analysis confirmed the important role played by the interaction between *WRKY18* and *WRKY40* since these two genes were shown to be affected by both *E. amylovora* and fungal pathogens. It is well-known that these two *WRKY*s play an important role in PAMP-triggered basal defense (Pandey et al., 2010; Bai et al., 2018; Chen et al., 2019). These two *WRKY*s negatively affect *EDS1* and *PAD4*, but positively up-regulated some key JA-signaling genes.

5. Conclusions

My meta-analysis was effective in confirming the effects of fungal pathogen attacks on reduction photosynthesis at transcriptomic level highlighting the importance of integrating different molecular, imaging and high-throughput platforms in early diagnose of plant stress status. In addition, it showed how specific hormones and transcription factor classes play specific roles in plant signaling responses to different pathogens. The PPI network highlighted the role of terpenoids in the response to pathogen attacks in *Malus x domestica*. The integrated meta-analysis approach and pipeline could be employed in comparing transcriptomic studies and deciphering common and exclusive features in the gene regulatory networks of other crop species.

6. References

Bai Y., Sunarti S., Kissoudis C., Visser R. G. F., van der Linden C. G. (2018). The Role of Tomato WRKY Genes in Plant Responses to Combined Abiotic and Biotic Stresses. Front Plant Sci. 9, 801.

Barón M., Pineda M., Pérez-Bueno M. L. (2016). Picturing pathogen infection in plants. Z. Naturforsch. C. J. Biosci. 71, 355–368.

Berger S., Benediktyová Z., Matous K., Bonfig K., Mueller M. J., Nedbal L., et al. (2007). Visualization of dynamics of plant-pathogen interaction by novel combination of chlorophyll fluorescence imaging and statistical analysis: differential effects of virulent and avirulent strains of *P. syringae* and of oxylipins on *A. thaliana*. J. Exp. Bot. 58, 797–806.

Chaerle L., Lenk S., Hagenbeek D., Buschmann C., Van Der Straeten D. (2007). Multicolor fluorescence imaging for early detection of the hypersensitive reaction to *tobacco mosaic virus*. J. Plant Physiol. 164, 253–262.

Chen S., Ye T., Hao L., Chen H., Wang S., Fan Z., et al. (2014). Infection of apple by *apple stem grooving virus* leads to extensive alterations in gene expression patterns but no disease symptoms. PLoS ONE 9, e95239.

Chen X., Li C., Wang H., Guo Z. (2019). WRKY transcription factors: evolution, binding, and action. Phytopathol Res 1, 13.

Cohen S. P., Leach J. E. (2019). Abiotic and biotic stresses induce a core transcriptome response in rice. Sci Rep. 9, 6273.

Dandekar A. M., Martinelli F., Davis C. E., Bhushan A., Zhao W., Fiehn O., et al. (2010). Analysis of early host responses for asymptomatic disease detection and management of specialty crops. Crit. Rev. Immunol. 30, 277–289.

Doerks T., Strauss M., Brendel M., Bork P. (2000). GRAM, a novel domain in glucosyltransferases, myotubularins and other putative membrane-associated proteins. Trends Biochem. Sci. 25, 483–485.

Fan X., Guo Q., Xu P., Gong Y., Shu H., Yang Y., et al. (2015). Transcriptome-wide identification of salt-responsive members of the *WRKY* gene family in *Gossypium aridum*. PLoS ONE 10, e0126148.

Ferrero S., Carretero-Paulet L., Mendes M. A., Botton A., Eccher G., Masiero S., et al. (2015). Transcriptomic signatures in seeds of apple (*Malus domestica L. Borkh*) during fruitlet abscission. PLoS ONE 10, e0120503.

Giovino A., Martinelli F., Saia S. (2016). *Rhynchophorus ferrugineus* attack affects a group of compounds rather than rearranging phoenix canariensis metabolic pathways. J. Integr. Plant Biol. 58, 388–396.

Glazebrook J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu. Rev. Phytopathol. 43, 205–227.

Gusberti M., Gessler C., Broggini G. A. (2013). RNA-Seq analysis reveals candidate genes for ontogenic resistance in *Malus-venturia* pathosystem. PLoS ONE 8, e78457.

Huang D. A. W., Sherman B. T., Lempicki R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Jiang J., Ma S., Ye N., Jiang M., Cao J., Zhang J. (2017). WRKY transcription factors in plant responses to stresses. J. Integr. Plant Biol. 59, 86–101.

Kamber T., Buchmann J. P., Pothier J. F., Smits T. H., Wicker T., Duffy B. (2016). Fire blight disease reactome: RNA-seq transcriptional profile of apple host plant defense responses to *Erwinia amylovora* pathogen infection. Sci. Rep. 6, 21600.

Khraiwesh B., Zhu J. K., Zhu J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. (2012). Biochim Biophys Acta. 1819, 137-48.

Kim D., Langmead B., Salzberg S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.

Kovalchuk A., Zeng Z., Ghimire R. P., Kivimäenpää M., Raffaello T., Liu M., Mukrimin M., Kasanen R., Sun H., Julkunen-Tiitto R., Holopainen J. K., Asiegbu F. O. (2019). Dual RNA-seq analysis provides new insights into interactions between Norway spruce and necrotrophic pathogen Heterobasidion annosum s.l. BMC Plant Biol. 19, 2.

Krost C., Petersen R., Lokan S., Brauksiepe B., Braun P., Schmidt E. R. (2013). Evaluation of the hormonal state of columnar apple trees (*Malus x domestica*) based on high throughput gene expression studies. Plant Mol. Biol. 81, 211–220.

Leufen G., Noga G., Hunsche M. (2014). Proximal sensing of plant-pathogen interactions in spring barley with three fluorescence techniques. Sensors 14, 11135–11152.

Li G., Ma J., Tan M., Mao J., An N., Sha G., et al. (2016). Transcriptome analysis reveals the effects of sugar metabolism and auxin and cytokinin signaling pathways on root growth and development of grafted apple. BMC Genomics 17, 150.

Li H., Jiang W., Zhang Z., Xing Y., Li F. (2013). Transcriptome Analysis and Screening for Potential Target Genes for RNAi-Mediated Pest Control of the Beet Armyworm, Spodoptera exigua. PLoS One. 8, e65931.

Limera C., Sabbadini S., Sweet J. B., Mezzetti B. (2017). New Biotechnological Tools for the Genetic Improvement of Major Woody Fruit Species. Front Plant Sci. 8, 1418.

Liu H., Dong S., Gu F., Liu W., Yang G., Huang M., et al. (2017). NBS-LRR Protein Pik-H4 Interacts with OsBIHD1 to balance rice blast resistance and growth by coordinating ethylene-brassinosteroid pathway. Front. Plant Sci. 8, 127.

Martinelli F., Ibanez A., Reagan R., Davino S., Dandekar A. (2015). Stress responses in citrus peel: comparative analysis of host responses to huanglongbing disease and puffing disorder. Sci. Hortic. 192, 409–420.

Martinelli F., Reagan R. L., Uratsu S. L., Phu M. L., Albrecht U., Zhao W., et al. (2013). Gene regulatory networks elucidating huanglongbing disease mechanisms. PLoS ONE 8, e74256.

Martinelli F., Uratsu S. L., Albrecht U., Reagan R. L., Phu M. L., Britton M., et al. (2012). Transcriptome profiling of citrus fruit response to huanglongbing disease. PLoS ONE 7, e38039.

Maxson K. L., Jones A. L. (2002). Management of fire blight with gibberellin inhibitors and SAR inducers. Acta. Hortic. 590, 217–223.

Mellidou I., Buts K., Hatoum D., Ho Q. T., Johnston J. W., Watkins C. B., et al. (2014). Transcriptomic events associated with internal browning of apple during postharvest storage. BMC Plant Biol. 14, 328.

Noir S., Bömer M., Takahashi N., Ishida T., Tsui T. L., Balbi V., et al. (2013). Jasmonate controls leaf growth by repressing cell proliferation and the onset of endoreduplication while maintaining a potential stand-by mode. Plant Physiol. 161, 1930–1951.

Overvoorde P., Fukaki H., Beeckman T. (2010). Auxin control of root development. Cold Spring Harb. Perspect. Biol. 2, a001537.

Pandey S. P., Roccaro M., Schon M., Logemann E., Somssich I. E. (2010). Transcriptional reprogramming regulated by *WRK18* and *WRKY40* facilitates powdery mildew infection of *Arabidopsis*. Plant J. 64, 912–923.

Pérez-Bueno M. L., Ciscato M., VandeVen M., García-Luque I., Valcke R., Barón M. (2006). Imaging viral infection: studies on *Nicotiana benthamiana* plants infected with the *pepper mild mottle tobamovirus*. Photosyn. Res. 90, 111–123.

Petersen R., Djozgic H., Rieger B., Rapp S., Schmidt E. R. (2015). Columnar apple primary roots share some features of the columnar-specific gene expression profile of aerial plant parts as evidenced by RNA-seq analysis. BMC Plant Biol. 15, 34.

Peterson R. B., Aylor D. E. (1995). Chlorophyll fluorescence induction in leaves of *phaseolus vulgaris* infected with bean rust (*Uromyces appendiculatus*). Plant Physiol. 108, 163–171.

Pieterse C. M., Leon-Reyes A., Van der Ent S., Van Wees S. C. (2009). Networking by small-molecule hormones in plant immunity. Nat. Chem. Biol. 5, 308–316.

Pineda M., Sajnani C., Barón M. (2010). Changes induced by the *pepper mild mottle tobamovirus* on the chloroplast proteome of *Nicotiana benthamiana*. Photosyn. Res. 103, 31–45.

Rushton P. J., Somssich I. E., Ringler P., Shen Q. J. (2010). WRKY transcription factors. Trends Plant Sci. 15, 247–258.

Shin S., Zheng P., Fazio G., Mazzola M., Main D., Zhu Y. (2016). Transcriptome changes specifically associated with apple (*Malus domestica*) root defense response during *Pythium ultimum* infection. Physiol. Mol. Plant Pathol. 94, 16–26.

Soukupova J., Smatanova S., Nedbala L., Jegorovd A. (2003). Plant response to destruxins visualized by imaging of chlorophyll fluorescence. Physiol. Plant. 118, 399–405.

Taylor-Teeples M., Lanctot A., Nemhauser J. L. (2016). As above, so below: auxin's role in lateral organ development. Dev. Biol. 419, 156–164.

Thimm O., Bläsing O., Gibon Y., Nagel A., Meyer S., Krüger P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939.

Usadel B., Nagel A., Steinhauser D., Gibon Y., Bläsing O. E., Redestig H., et al. (2006). PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics 7, 535.

Wang N., Zheng Y., Duan N., Zhang Z., Ji X., Jiang S., et al. (2015). Comparative Transcriptomes analysis of red- and white-fleshed apples in an F1 population of *Malus sieversii f. Niedzwetzkyana* crossed with *M. domestica* ‘Fuji’. PLoS ONE 10, e0133468.

Xia J., Benner M. J., Hancock R. E. (2014). NetworkAnalyst–integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. 42, W167–W174.

Xu X., Chen C., Fan B., Chen Z. (2006). Physical and functional interactions between pathogen-induced *Arabidopsis WRKY18*, *WRKY40*, and *WRKY60* transcription factors. Plant Cell 18, 1310–1326.

Yin Z., Ke X., Kang Z., Huang L. (2016). Apple resistance responses against *valsa mali* revealed by transcriptomics analyses. Physiol. Mol. Plant Pathol. 93, 85–92.

Zander M., La Camera S., Lamotte O., Métraux J. P., Gatz C. (2010). *Arabidopsis thaliana* class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. Plant J. 61, 200–210.

Zhu L., Ni W., Liu S., Cai B., Xing H., Wang S. (2017). Transcriptomics analysis of apple leaves in response to *Alternaria alternata* apple pathotype infection. Front. Plant Sci. 8, 22.

Experiment 3

Identifying Host Molecular Features Strongly Linked With Responses to Huanglongbing Disease in *Citrus sinensis* Leaves.

DOI : https://doi.org/10.3389/fpls.2018.00277

1. Introduction

Citrus huanglongbing (HLB), called citrus greening disease, a destructive disease of citrus that represents a major threat to the world citrus industry and is slowly invading new citrus growing areas (Dala-Paula et al., 2018). HLB, whose name in Chinese means "yellowd ragon disease", was first reported from southern China in 1919 and is now known to occur in next to 40 different Asian, African, Oceanian, South and North American countries (Bove, 2006). Even though Citrus huanglongbing (HLB), caused by *Candidatus Liberibacter spp*, is not yet present in the Mediterranean basin, this devastating disease is threatening the very survival of citrus in most parts of the world. As management of HLB would be difficult in the Mediterranean area, in particular because of the small size of most citrus orchards, all efforts should be concentrated towards preventing HLB entrance and spread in the area (Duran-Vila et al., 2014).

The disease causes yellow color of young shoots and the leaves show blotchy mottling, yellow veins and mineral deficiency symptoms, like that induced by zinc in particular. Fruits are lopsided, with color inversion and aborted seeds (Fig. 1.1). So far, seed transmission of HLB has not been demonstrated. Eventually, affected trees decline and become uneconomical. Since the initial discovery of HLB in Florida in 2005, the disease has affected most citrus-producing areas, causing the loss of about 100,000 acres of citrus, at a cost of approximately $3.6 billion in lost revenues and 6600 jobs (Alvarez et al., 2016; Ferrarezi et al., 2019). Currently, it is estimated that Las has infected over 95% of the mature trees in commercial citrus groves in Florida.

Control of HLB requires quarantine, clean stock, and certification programs in order to produce healthy plants and prevent movement of infected nursery stock. The psyllid vectors must be controlled. In areas where HLB is not already established, a three pronged approach to control is effective: regular surveys to identify early symptoms on trees which are then removed; control of the psyllid vector by survey and pesticide application; and use of clean

plant material for replanting (Bové, 2006). Detection of the bacterium associated with HLB is by PCR or real-time PCR. Testing psyllids for the presence of the bacterium associated with HLB by real-time PCR has proven to provide an earlier warning of the presence of the disease in an area where HLB is not already established (Manjunath et al., 2008).

Figure 1.1–Oranges with HLB are afflicted with a green color.

The pathogenetic mechanisms of Huanglongbing (HLB) disease remain unclear. The disease is caused by a phloem-limited bacterium, *Candidatus liberibacter asiaticus* (CaLas), transmitted by psyllids (Rao et al., 2019). The pathogen has three subspecies: *americanus*, *africanus*, and *asiaticus*. The first two subspecies infect *Citrus sinensis* in South America and Africa, respectively, while the asiaticus subspecies is widespread in North America and Asia. The pathogen is closely related to Rhyzobiaceae and has biotrophic behavior. The caused disease is the most threatening in *Citrus sinensis* worldwide and leads to tree death in few years, reduced tree growth, yellowing of leaves and malformed, unmarketable fruits characterized by small seeds, high acidity, small size, and altered ripening dynamics.

Although genomic sequences have been determined (Duan et al., 2009; Tyler et al., 2009) and putative toxins have been isolated, the toxic molecules are not the only cause of disease symptoms. Previous -omic approaches have provided insight into the molecular mechanisms provoking symptoms. The first studies conducted through microarrays highlighted

upregulation of genes involved in carbohydrate biosynthesis and metabolism, particularly those involved in starch pathways, such as AGPase and starch synthase (Albrecht and Bowman, 2008; Kim et al., 2009). Repression of photosynthesis and other primary metabolic pathways was also observed. RNA-Seq studies on different sink and source organs showed that sink-source tissue relationships were severely modified by the disease (Martinelli et al., 2012, 2013). Upregulation of glucose-phosphate-transporter2 was considered a key factor driving starch accumulation in infected leaves, a common symptom of the disease. CaLas-infected fruits remained green and photosynthesizing, while the mature leaves had decreased photosynthesis and yellowing due to starch accumulation, causing inversion of the usual relationship between developing fruit and mature leaves (Martinelli et al., 2015). Another important factor causing symptoms was reduced expression of genes encoding heat shock proteins (i.e., HSP82). The products of these genes protect protein folding and function during stress conditions. These genes help maintain normal function of key proteins, especially in phloem and leaves, maintaining correct function of proteins involved in primary metabolism. A third devastating effect of the pathogen is modified hormonal cross-talk. The pathogen induced upregulation of key genes involved in jasmonic acid-mediated responses in leaves (lox1, lox2, and lox3), probably in response to insect attacks. In addition, the salicylic acid-mediated response was more highly induced in fruits than in leaves, although fruits are not commonly the place of infection. Finally, induction of abscissic acid and auxin genes (*HVA22C*, *SAUR-like*, and *UGT71B6*) should counteract the action of salicylic acid responses, helping the pathogen grow and develop. Proteomic approaches performed to study pathogenetic mechanisms of the disease confirmed these findings, highlighting downregulation of photosynthesis-related proteins and modification of transport, carbohydrate metabolism, hormone biosynthesis, metabolism and xenobiotic responses. At the proteomic level, proteins involved in detoxification of oxidative stresses (glutathione S-transferases and nitrilases), in cell wall modification and pathogenesis-related processes were most effective in promoting *Citrus sinensis* tolerance (Martinelli et al., 2016).

Studies of the transcriptome, proteome, and metabolome led to design of new translational genomic tools to speed diagnosis and develop short- and long-term therapeutic and genetic resistance strategies (Dandekar et al., 2010; Ibáñez et al., 2014). Newly developed sensor devices that capture early-induced molecules produced by infected leaves and fruits could distinguish recently infected, still-asymptomatic trees from severely symptomatic ones

(Aksenov et al., 2014). This approach might be extended and applied to other specialty crops (Martinelli et al., 2016). Diagnostic methods focusing on host responses are highly desirable as a complement to traditional approaches targeting the pathogen, as the latter can seldom detect the pathogen before visual symptoms occur.

The -omic approaches do have weaknesses: low reliability and scarce or (often) absent experimental repetition. They are performed in different environments and seasons on trees grown under different agronomic conditions. There is an urgent need to perform duplicate field studies due to the many environmental variables that affect gene expression. The timing of disease progression from infection to tree death has also led to contradictory conclusions. Some of the issues can be addressed using meta-analysis to compare differentially regulated genes and affected pathways among different studies using the same bioinformatic methods (Rawat et al., 2015).

The aim of this study was to perform a bioinformatic analysis of previously published RNA-Seq studies on leaves of CaLas-infected *Citrus sinensis* using the same pipeline. In this study, I analyzed the raw datasets using the most-updated bioinformatic pipeline and an integrated functional data mining approach to identify common molecular patterns that were consistently linked with pathogen infection. Even while limited by the available stress-responsive transcriptome data, real trends were identifed, indicating it is possible to design experiments in less well-studied plant systems to use with my approach. It is also possible to expand my analysis approach to investigate the stress responses of other plants.

2. Materials and methods

2.1 Search Strategy to Identify Published Studies for Bioinformatic Analysis

The published RNA-Seq studies in *Citrus sinensis* related to HLB response and tolerance in leaf tissues were searched using Scopus and PubMed. I found three studies published on or before May 2017 (Martinelli et al., 2013; Fu et al., 2016; Wang et al., 2016). The first RNA-Seq study was divided into two datasets for young and mature leaves (Martinelli et al., 2013). The second study was performed under controlled conditions using artificial infection of young leaves (Fu et al., 2016). The third work (Wang et al., 2016) compared *Citrus sinensis* tolerance mechanisms with the list of HLB-regulated genes in common between the previous

two studies. I first analyzed the three transcriptomic datasets related to HLB-response. The next step was to identify the genes related to tolerance that were present in the three datasets dealing with HLB response. The raw data from the four datasets were downloaded and performed using the meta-analysis bioinformatics pipeline as described (Figure 2.1.1).

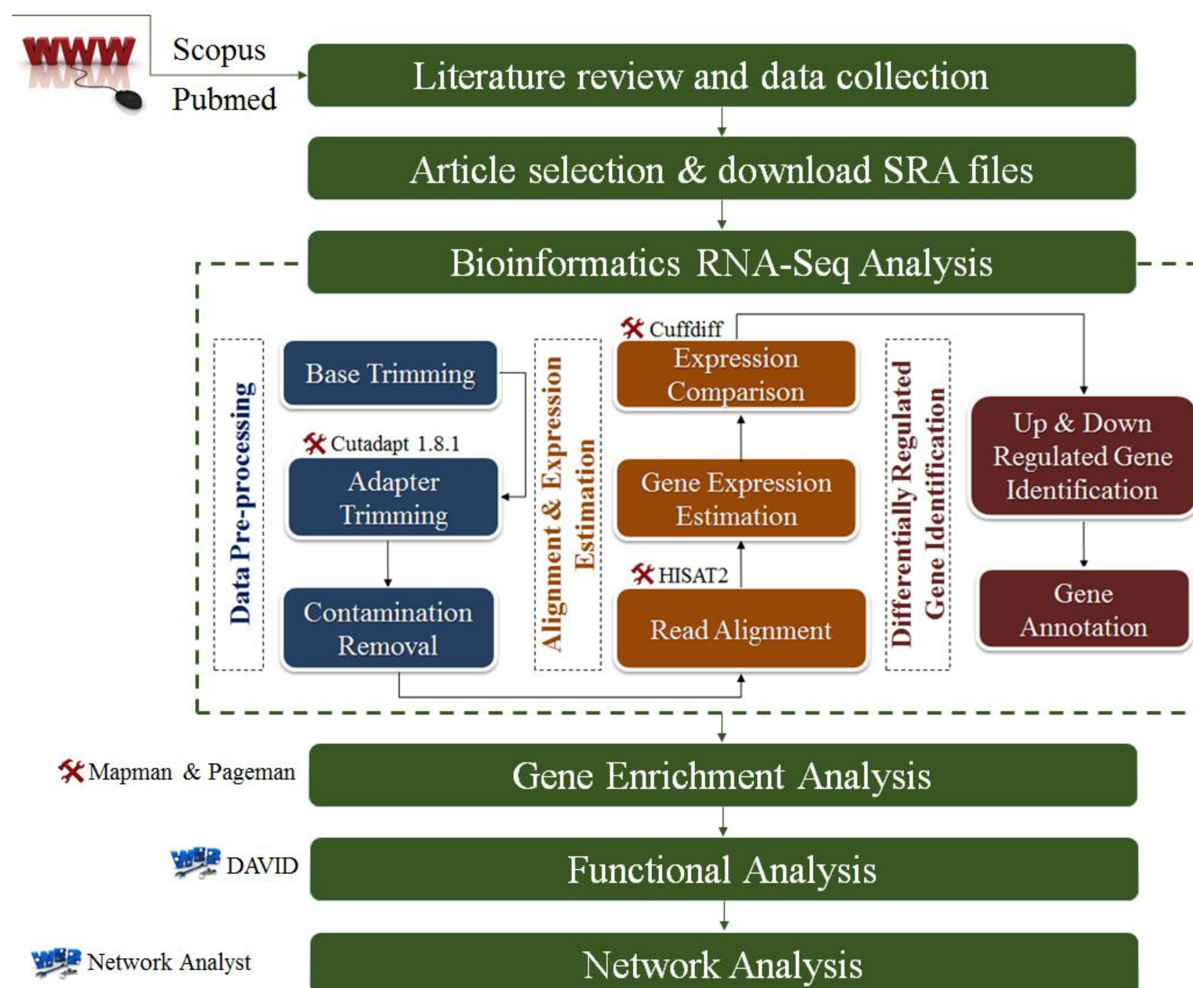


Figure 2.1.1–Meta-analysis workflow of the four RNA-Seq data dealing with Huanglongbing (HLB) response and tolerance in leaf tissues. Functional data mining tools were provided.

2.2 Bioinformatic Analysis of Raw Data

The *Citrus sinensis* v1.1 and annotation file were downloaded from Phytozome (https://phytozome.jgi.doe.gov). The SRA files of the three articles were downloaded from NCBI SRA (https://www.ncbi.nlm.nih.gov/sra) and then converted to FASTQ format using

SRAToolkit version 2.3.5. The raw reads were filtered to obtain high-quality clean reads by trimming low-quality bases followed by adaptor sequence removal using cutadapt version 1.8.1. The pre-processed reads were mapped to the *Citrus sinensis* genome v1.1 with HISAT2 version 2.0.5 (Kim et al., 2015) using default parameters. The identification of differentially expressed genes was performed using Cuffdiff algorithm in Cufflinks version 2.2.1 pipeline with default parameters (default false discovery rate, FDR is 0.05).

2.3 Differentially Expressed Gene Selection

The up- and down-regulated genes with fold change cutoff (log2 FC > 1 or log2 FC < -1) and p-value <0.05 were identified from the selected articles. *Citrus sinensis* gene ids were mapped to corresponding TAIR IDs using the mapping file downloaded from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html). The common and unique differentially HLB-regulated genes among different studies were identified. A custom made in-house Perl script was employed for gene selection and mapping.

2.4 Splice Analysis

The alternative splicing events of each samples were predicted by the ASTALAVISTA program (Foissac and Sammeth, 2007) (http://astalavista.sammeth.net/) on the web server3 using the GTF files generated by Cufflinks. Differential splicing analysis was done using MISO version 0.5.34 (Katz et al., 2010) (https://miso.readthedocs.io/en/fastmiso/) and rMATS version 4.0.15 (Shen et al., 2012) (http://rnaseq-mats.sourceforge.net/) using the default options. The Sashimi plots were generated in order to get the quantitative visualization of the aligned RNA-Seq reads which enables quantitative comparison of exon usage across the control and treated samples.

2.5 Gene Enrichment and Functional Analysis

MapMan (Thimm et al., 2004) was used with the *Citrus sinensis* mapping file (http://mapman.gabipd.org/mapmanstore) to map the gene ids and visualize the metabolic overview, large enzyme families, hormone-related genes, transcription factors and biotic-

stress related genes in all four transcriptomic datasets. PageMan (Usadel et al., 2006) analysis was used for gene set enrichment analysis and to visualize differences among metabolic pathways using Wilcoxon tests, no correction, and an over-representation analysis (ORA) cutoff value of 1.

Pathway enrichment analysis using gene ontologies was conducted using Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) (Huang et al., 2009). The complete DAVID pathway search results were provided. The gene ontology information related to Biological process was downloaded from the DAVID results (FDR cutoff = 0.05).

2.6 Protein–Protein Interaction Network

Individual data annotation and analysis were performed using NetworkAnalyst (Xia et al., 2014) (https://www.networkanalyst.ca/), a web-based tool to visualize protein–protein network analysis. The homologous TAIR IDs were uploaded and mapped against the STRING interactome database using default parameters provided in NetworkAnalyst. I selected 'Minimum Network' to simplify the network and highlight key connections. First, I performed network analysis between the two transcriptomic datasets using the same tissue (young leaves). Next, I compared the common HLB-regulated genes (Martinelli et al., 2013; Fu et al., 2016) with those linked to tolerance (Wang et al., 2016).

3. Results

3.1 Workflow, Bioinformatics Analysis, and Venn Diagrams

Twelve RNA-Seq raw datasets from three published articles were analyzed using a bioinformatic and functional data mining pipeline (Figure 2.1.1 and Table 3.1). The first study had two pairwise comparisons between apparently healthy and symptomatic young and mature leaf samples (Martinelli et al., 2013). Trees grew in the same orchard and were both infected by CaLas. Raw data submitted to NCBI were re-analyzed. Over 46 – 68 million reads were obtained from the four samples with an alignment percentage of 72.95 – 82.29%. The second study used a single pairwise comparison between healthy and infected leaf

samples (Fu et al., 2016). The number of reads from the control sample was much less than that from the infected one, although alignment percentages were similar. This may reduce the depth of the transcriptomic analysis. The third study compared susceptible and tolerant *Citrus sinensis* genotypes to identify which genes were related to molecular mechanisms of tolerance in leaves (Wang et al., 2016). Three biological replicates were used and 72–118 million reads were obtained, with 73.42–82.44% alignment.

Article	SRR ID	Orginal Sample Name	Sample Tissue	Given Sample Name	Total Reads	Alignment %
Martinelli et al., 2013	SRR867442 SRR867443	ML+AH	Mature Leaves	Control-A	68,263,920	81.69%
	SRR867431	ML+SY		Infected-A	62,503,214	72.95%
	SRR867426	YL+AH	Young Leaves	Control-B	64,186,912	82.29%
	SRR867431	YL+SY		Infected-B	46,296,354	81.20%
Fu et al., 2016	SRR3032893	Healthy (H)	Young Leaves	Control-C	8,523,448	86.76%
	SRR3032892	CaLas-B232		Infected-C	94,569,124	86.87%
Wang et al., 2016	SRR2224205	R19T23	Young Leaves	Susceptible	106,189,916	78.84%
	SRR2224296	R19T24			118,300,964	75.65%
	SRR2224411	R20T24			101,835,958	75.13%
	SRR2224421	R20T17		Tolerant	112,832,326	73.42%
	SRR2224429	R19T17			72,491,540	82.44%
	SRR2224406	R20T18			102,457,842	77.78%

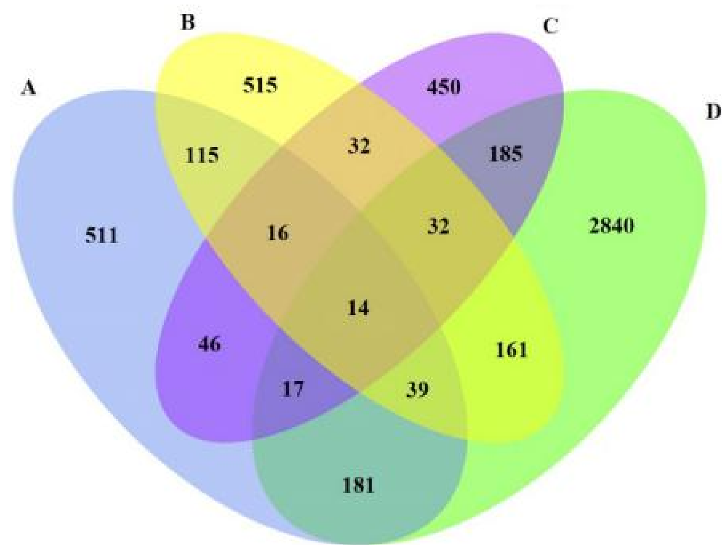
Table 3.1– Analyzed articles, SRA Ids, tissue, Read count, and alignment information.

Total 939 significantly CaLas-regulated genes were identified in mature leaves from the first study: 516 up-regulated and 423 down-regulated from apparently healthy samples (Table 3.2). In young leaves, 924 genes were significantly affected: 430 up-regulated and 494 down-regulated. In the third dataset, there were fewer repressed genes than CaLas-up-regulated ones. The molecular study addressing the understanding of tolerance highlighted 3,469 significantly regulated genes: 1,712 up-regulated and 1,757 down-regulated in tolerant genotypes compared to susceptible ones.

	Comparison			
	Control-A vs. Infected-A	**Control-B vs. Infected-B**	**Control-C vs. Infected-C**	**Susceptible vs. Tolerant**
Total Genes	939	924	792	3,469
Up-regulated Genes	516	430	707	1,712
Down-regulated Genes	423	494	85	1,757

Table 3.2– The count of up- and down- regulated genes in each study is given.

Only 16 genes were commonly modulated between the three transcriptomic leaf datasets and 14 were commonly modulated in all four transcriptomic works (Figure 3.1.1). Between young and mature leaves infected in the field, 115 genes were commonly modulated by HLB. Five hundred fifteen were significantly regulated in young leaves and 511 in mature leaves. Six hundred twenty-nine genes linked with HLB tolerance were also modulated in the three leaf datasets.



A : Up & Down regulated genes in Control-A vs. Infected-A
B : Up & Down regulated genes in Control-B vs. Infected-B
C : Up & Down regulated genes in Control-C vs. Infected-C
D : Up & Down regulated genes in Susceptible vs. Tolerant

Figure 3.1.1–Venn diagram indicating number of HLB-regulated genes commonly modulated between the four datasets and specifically induced by HLB in each of the four RNA-Seq datasets.

3.2 Gene Set- and Pathway-Enrichment Analysis

In naturally HLB-infected mature leaves, no biological processes related to defense mechanisms were induced (Sample Name A). In young leaves, some key gene-ontologies were over-repressed in infected young leaves: defense response to bacterium, lipid oxidation, oxylipin biosynthesis, and response to salicylic acid (Sample Name B) (Table 3.3). In artificially infected young leaves (Sample Name C), other gene set categories were affected such as jasmonic acid biosynthesis, plant-type hypersensitive response, and response to wounding.

Comparison	Up/Down	GO ID	GO Term	Count	pval	Fold Enrichment	FDR
Control-A vs. Infected-A	Up	GO:0080167	response to karrikin	12	1.22E-05	5.488231804	1.74E-02
	Down	GO:0015979	photosynthesis	20	1.56E-11	7.616518445	2.28E-08
	Down	GO:0009658	chloroplast organization	12	2.65E-05	5.045181818	3.88E-02
Control-B vs. Infected-B	Down	GO:0034599	cellular response to oxidative stress	9	2.72E-05	7.390403054	3.99E-02
	Down	GO:0009773	photosynthetic electron transport in photosystem I	6	8.79E-06	19.70774148	1.29E-02
Control-C vs. Infected-C	Up	GO:0006952	defense response	40	3.00E-07	2.493479536	4.46E-04
	Up	GO:0009611	response to wounding	30	8.86E-15	6.183578647	1.32E-11
	Up	GO:0042742	defense response to bacterium	21	1.51E-05	3.120914875	2.24E-02
	Up	GO:0009751	response to salicylic acid	15	2.65E-05	3.969458551	3.93E-02
	Up	GO:0009626	plant-type hypersensitive response	11	1.17E-05	6.097231378	0.017399625
	Up	GO:0031408	oxylipin biosynthetic process	8	1.16E-06	13.67257945	1.73E-03
	Up	GO:0009695	jasmonic acid biosynthetic process	7	2.40E-05	11.48496674	3.55E-02
	Up	GO:0034440	lipid oxidation	5	4.99E-06	34.18144863	7.40E-03
	Down	GO:0009408	response to heat	7	2.46E-05	11.90193015	0.030597268

Table 3.3– Significantly regulated biological processes in the analyzed transcriptomic studies (FDR < 0.05).

3.3 Molecular Responses to Huanglongbing Disease

3.3.1 Metabolism Overview

MapMan metabolism overview highlighted that HLB disease highly repressed photosynthesis in mature leaves and somewhat in young leaves (Figure3.3.1.1). Upregulation of starch and sucrose metabolism was shown by the induction of key genes encoding beta-amylase6, glucan phosphorylase in mature leaves, and phosphoglucan water dikinase in young leaves at 32 weeks after infection (w.a.i.). A contrasting expression pattern was observed for different genes from the same pathway. Some genes were up-regulated and others were repressed in primary metabolic pathways such as glycolysis, oxidative phosphate phosphorylation, fermentation, photorespiration, and tetrapyrrole. Trehalose-6-phosphate phosphatase and myo-inositol oxygenase were up-regulated in young leaves after artificial infection while aldo/keto reductase and 1,3–beta glucan synthase were induced in mature leaves. Cell wall-related genes were highly affected in all three gene expression datasets. UDP-D-glucuronate-4-epimerase6, pectate lyase, RD22 nutrient reservoir, and pectin methylesterase were enhanced in young leaves at 32 w.a.i. Genes up-regulated in young CaLas-infected leaves included UDP-glucose/UDP-galactose, 4-epimerase5, cellulose synthase, and the glycoside hydrolase 28 family. Lipid-related genes were particularly affected in immature leaves: choline kinase, glycerol-3-phosphate acyl transferase, and SUR4 membrane protein. Amino acid metabolism genes were mostly up-regulated in young leaves at 32 w.a.i.: shikimate dehydrogenase, chorismate mutase, and homoserine dehydrogenase. Genes involved in amino acid biosynthesis were more induced in young leaves (L-asparaginase, alanine-glyoxylate aminotransferase, and enoyl-CoA hydratase).

Terpene-related genes encoding myrcene synthase, beta-amyrin synthase, and terpene synthase 21 were enhanced in young leaves 32 w.a.i., while homogentisate phytyltransferase 1 was repressed. Genes encoding chalcone and stilbene synthase and isoflavone reductase were up-regulated in young leaves, while two methyltransferase family 2 proteins were induced in mature leaves.

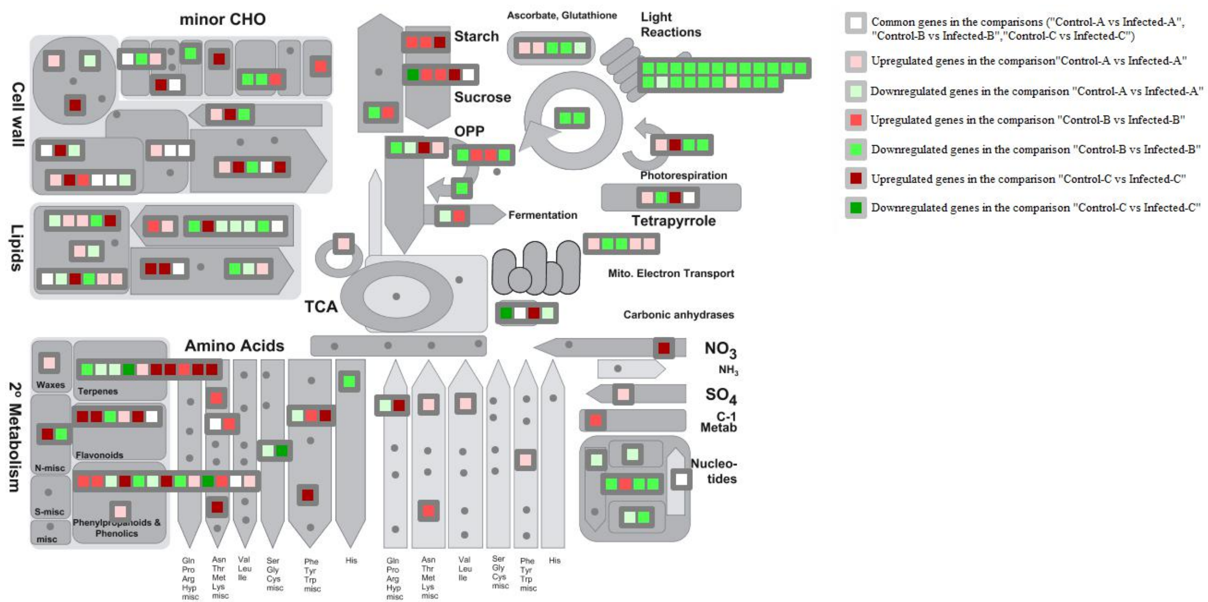


Figure 3.3.1.1–MapMan overview of three transcriptomic datasets related to HLB response. Up- and down-regulated genes in the three pairwise comparisons were shown. Common HLB-regulated genes between at least 2 of 3 pairwise comparisons were also shown.

Key genes involved in sucrose and starch metabolism were induced by CaLas in multiple studies. However, the same gene isoforms were not always affected. In mature leaves of naturally infected field trees, invertase1 and invertase2 were up-regulated while beta-fructosidase4 was commonly modulated by different studies. Starch branching enzyme2, glucan phosphorylase, and beta-amylase6 were induced in mature leaves (Martinelli et al., 2013), while a phosphoglucan water dikinase was enhanced in young leaves after artificial infection (Fu et al., 2016).

3.3.2 Hormone Overview

Comparison between datasets showed that hormone crosstalk was severely modified by CaLas infection. Jasmonic acid-mediated response was highly induced in young leaves at 32 w.a.i.: *lox1*, *lox2*, and *lox3* genes were up-regulated (Figure 3.3.2.1). S-adenosylmethyltransferase was affected by HLB in multiple datasets. However, the induction of genes involved in auxin and abscissic acid synthesis might counteract the beneficial effects of this gene. Abscissic acid-response up-regulated genes included GRAM-domain containing

protein and benzodiazepine receptor-related, while auxin-related genes induced in young leaves included IAA-alanine resistant3, IAA-amino acid conjugate hydrolase, aldo/keto reductase, and AILP1. In contrast, Fe(II) oxygenase, senescence-related gene1, and ethylene response factor1, all involved in ethylene response, were induced in concert with the *GID1* gene involved in gibberellin signaling. In young leaves, some key up-regulated genes were involved with auxins (*NGA1*), ethylene (flavonol synthase, ethylene-response-element-binding protein), and gibberellin (GASA proteins).

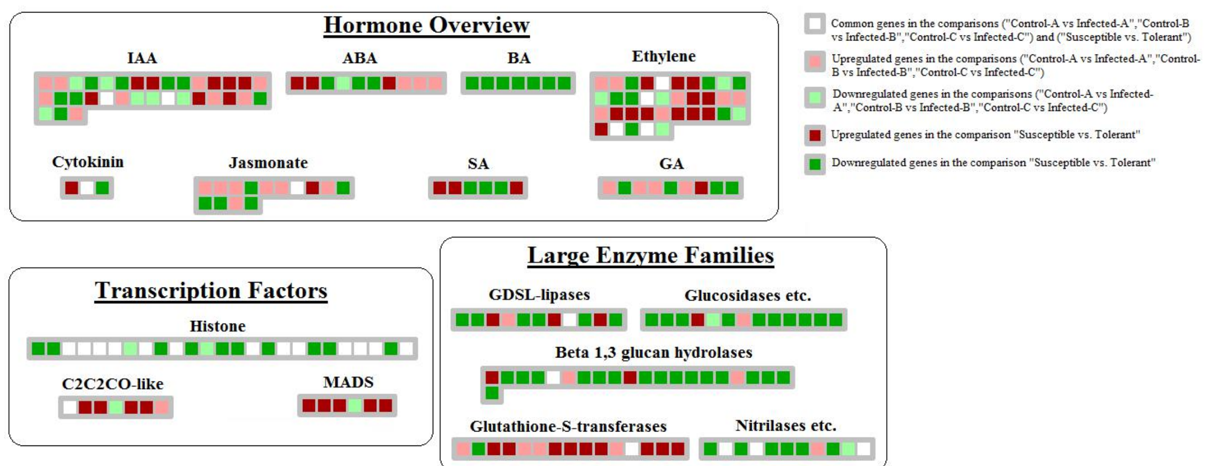


Figure 3.3.2.1–Genes linked with HLB tolerance and encoding transcription factors, enzymes, proteins involved in hormone-related pathways. Those genes commonly modulated between this dataset and the 3 related to HLB-response were shown

3.3.3 Transcription Factors

Huanglongbing induced key genes encoding AP2-EREBPs such as *SHN1*, *CRF1*, and two *AP2* domain-containing transcription factors in young leaves after artificial infection (Fu et al., 2016), another AP2 domain protein and a WRI1 in mature leaves and an ERE-BP in young leaves of symptomatic field trees (Figure 3.3.3.1). Other transcription factor categories induced by HLB were *HB* (Homeobox), *MYB*s, *C2H2*, pseudo *ARR*, and *GRAS*. HB TFs were induced in young leaves 32 w.a.i., including *HB40*, *HAT9*, *HB17*, and *KNAT7*. Two *MYB*s were enhanced in mature leaves (*MYB82* and *MYB116*). Genes encoding C2H2 transcription

factors up-regulated in young leaves 31 w.a.i. included STZ, zinc finger protein 7 and ZAT10 (Fu et al., 2016). Genes encoding pseudo ARR were also induced: *PRR3*, *PRR5*, and *PRR7*.

Figure 3.3.3.1–HLB-regulated genes encoding transcription factors. Up- and down-regulated genes in the three pairwise comparisons were shown. Common HLB-regulated genes between at least 2 of 3 pairwise comparisons were also shown.

WRKY transcription factors are involved in biotic stress responses. Several members of this family were induced by the three RNA-Seq datasets. *WRKY60*, *WRKY70*, *WRKY40*, and *WRKY33* were up-regulated in young leaves 32 w.a.i. *WRKY31* was induced in mature CaLas-infected leaves, while *WRKY2*, *WRKY47*, *WRKY42*, and *WRKY7* were enhanced in young leaves of field trees. Only *WRKY48* was commonly modulated by CaLas in two of the three datasets.

3.3.4 Biotic Stress Responses

Genes encoding glutathione S-transferases such as *GST7* and *GST25* were generally up-regulated by HLB disease in young leaves after artificial infection and in mature CaLas-infected leaves (Figure 3.3.4.1). Heme-binding was induced in young leaves, peroxidases were up-regulated in mature leaves and two peroxidases were enhanced in artificially infected immature leaves. Genes encoding cell wall modification and restructuring such as *GAE6*,

PRP4, *XTR6*, *RD22*, and pectate lyase were up-regulated in young leaves after artificial infection. UGE5, CSLG2, RGP2, glycoside hydrolase family 28 proteins, and expansin4 were induced in young leaves.

Seven pathogenesis-related proteins were up-regulated in immature leaves (Fu et al., 2016). A transmembrane signaling receptor involved in the SAR response, *EDS1*, was up-regulated in immature leaves after artificial infection (Fu et al., 2016), but not after natural infection (Martinelli et al., 2013).

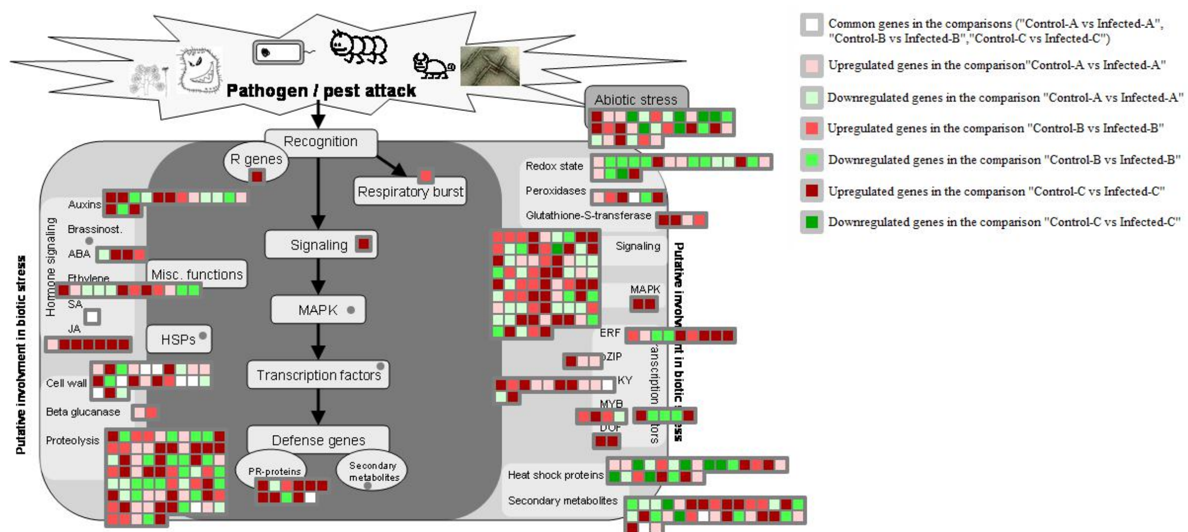


Figure 3.3.4.1–HLB-regulated genes involved in biotic stress responses. HLB-regulated genes in the three pairwise comparisons and commonly modulated between two of the three pairwise comparisons were shown.

3.3.5 Genes Commonly Involved in HLB Response Between Datasets

Genes with the same altered pattern of expression in more than one experiment were identified. More genes involved in primary than secondary metabolism were observed. The category of genes that was most commonly modulated in published transcriptomic datasets was cell wall modification and restructuring (Figure3.3.1.1). These genes were: *CESA8*, pectinesterase, expansin8, expansin beta 3.1, and pectate lyase. Some other genes linked with more than one transcriptomic study were related to lipids such as brassinosteroid sulfotransferase, myzus persicae-induced lipase 1, and mitochondrial acyl carrier protein 3.

Among genes affecting hormonal crosstalk, an S-adenosylmethionine-dependent methyltransferase was up-regulated in multiple transcriptomic datasets. Other key genes involved in biotic stress responses were affected in both young and mature leaves: *WRKY48*, peroxidase, and F-box family protein. In starch and sucrose metabolism, an invertase was modulated by HLB in multiple studies. Among transcription factors, genes encoding phytochrome interacting 3-like 1, MYB transcription factor, and IAA14 were clearly and specifically linked to HLB response.

3.4 Molecular Mechanisms of HLB Tolerance

The comparison between susceptible and tolerant species highlighted that many more genes involved in photosynthesis and the Calvin cycle were repressed by HLB in susceptible than in tolerant plants (Figure3.4.1). However, there were very few genes in common between the HLB response and tolerance datasets. Tolerant genotypes showed downregulation of genes encoding chlorophyll binding, oxygen-evolving complex-related, thylakoid luminal 20 kDa protein, and two ferredoxin-related proteins. Photosystem II reaction PSB28 protein was commonly repressed. In the tolerant genotype, there was a repression of genes involved in tetrapyrrole (NADH-ubiquinone oxidoreductase 20 kDa subunit, alternative, NADH dehydrogenase, ubiquinonl-cytochrome C reductase complex 14 kDa, cytochrome c oxidase, and ATP synthase) and the TCA cycle (*LTA2*, succinyl-CoA ligase, and malate dehydrogenase). In comparison to HLB response, there were more repressed genes involved in call wall modification such as 4 pectinesterases, *PME1*, and *PME3*. There was also decreased transcript abundance of genes involved in cellulose synthesis, cellulases and beta-1,4'glucanases, poligalacturonases, cell wall precursor synthesis, fatty acid synthesis and fatty acid elongation. Genes involved in starch degradation were up-regulated, including alpha-amylase2, beta-amylase8, glycoside hydrolase, and starch excess4. Sucrose biosynthesis was enhanced (sucrose-phosphate-synthase, sucrose-phosphate1, and transferase). Secondary metabolism genes involved in terpenes and phenylpropanoids were induced: homogentisate phytyltransferase1, amino_oxidase, carotenoid isomerase, cycloartenol synthase, O-methyltransferase2, and isoflavone-7-O-methyltransferase9. More up-regulated genes involved in flavonoid synthesis than in phenylpropanoid synthesis were

observed: O-methyltransferase, oxidoreductase, isoflavone reductase, and pinoresinol reductase. More genes involved in amino acid metabolism than biosynthesis were repressed.

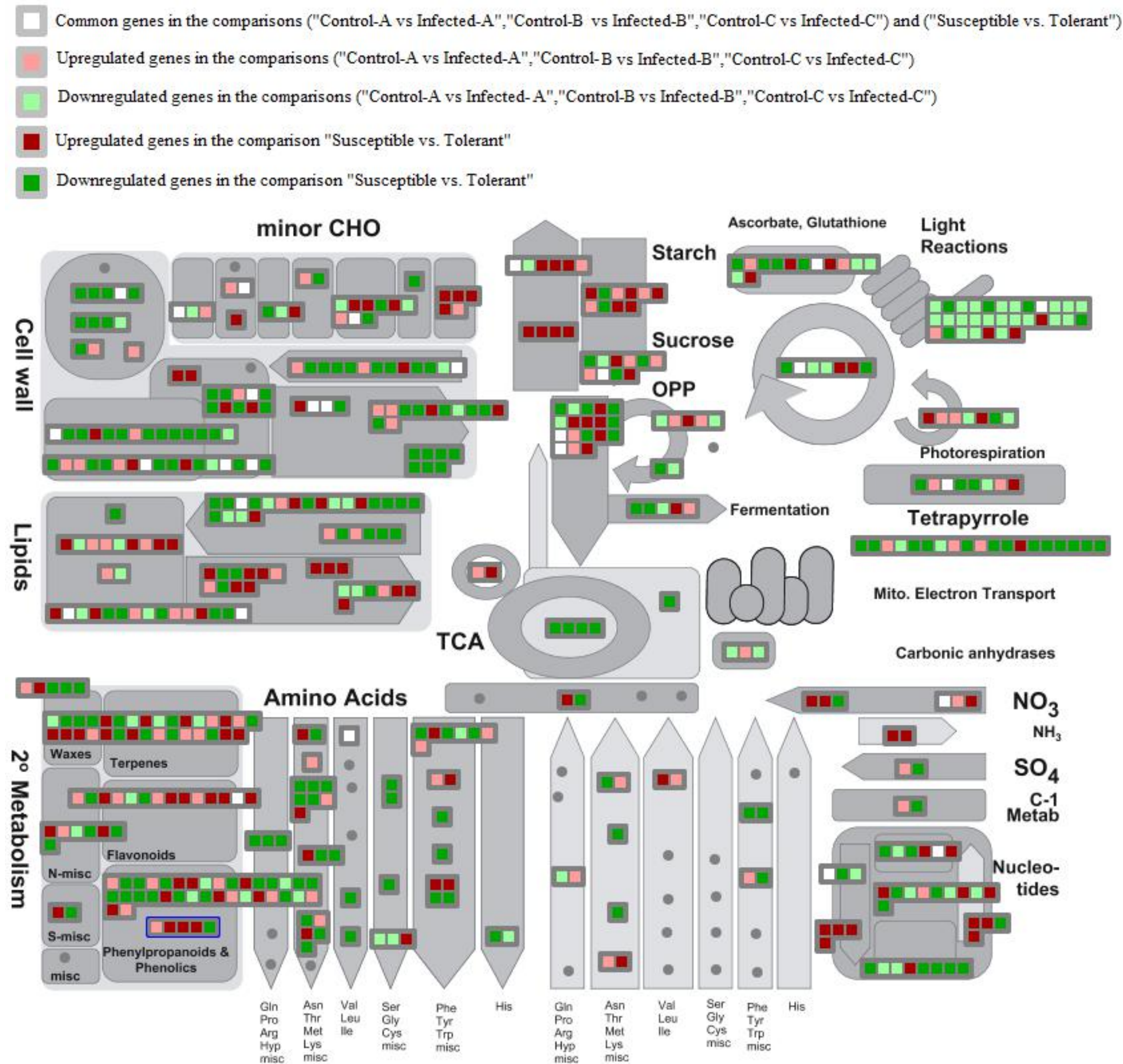


Figure 3.4.1–Mapman overview showing differentially regulated genes between susceptible and tolerant genotypes. Genes commonly modulated between this dataset and the three related to HLB response are shown.

3.4.1 Hormone Overview

Repression of brassinosteroids and salicylic and jasmonic acid-mediated responses were more linked with tolerance than HLB response. The following genes were down-regulated: steroid

5-alpha-reductase, sterol methyl tranferase2, C-8 sterol isomerase, C-5 sterol desaturase, BR1-EMS-suppressor, lipoxygenase, electron carrier, and allene oxide synthase (Figure3.3.2.1). Several genes involved in ethylene-related pathways were up-regulated: oxidoreductase, 2-oxoglutarate-dependent dioxygenase, *ACS6* and *ACS12*, *HLS1*, and *ERF104*. Several ethylene-related genes were commonly modulated by HLB in the two types of datasets (oxidoreductase, gibberellin-2-beta-dioxygenase, ethylene-regulated nuclear protein, and universal stress protein). A similar number of genes involved in auxin-related pathways were up- or down-regulated. While some GA-related genes were up-regulated in response, others involved in the same hormone pathways were repressed: GASA4, gibberellin-responsive protein, and GASA protein.

3.4.2 Transcription Factors

More *MYB*s were up-regulated in the tolerant genotype: *MYB59*, *MYB55*, *MYB15*, *MYB30*, *MYB73*, and *MYB52* (Figure3.3.4.1). Other transcription factor categories were induced, including MADS (*AGL7*, *AGL22*, and *AGL42*), B3 DNA binding protein (VRN1), histone ATse (*ADA2B*, *HAF01*, and *HAC12*), C2C2-CO-like (*COL9* and zinc finger B-box type), and homeobox (*HB-1*, *HB-7*, and *HAT9*). There were many genes involved in chromatin structure remodeling that were commonly modulated by both response and tolerance: histone4, *HTA7*, histone H3.2, *HMGA*, and *HTA5*.

3.4.3 Biotic Stress Responses

In addition to the differentially regulated genes previously mentioned, tolerance was linked with downregulation of genes involved in cellulose and cell wall precursor synthesis such as UDP-glucuronate decarboxylase, UDO-6-glucose-6-dehydrogenase, GDP-mannose 4,6,-dehydratase, rhamnose biosynthesis 1, and cellulose synthase like C4 and D3. Several beta-glucanases were also repressed (Figure 3.4.3.1). Twenty-one pathogenesis proteins were induced in tolerant genotypes encoding TIR-NBS-LRR proteins. Detoxifying pathways were up-regulated as shown by the induction of several glutathione S-transferases (*GSTU19*, *GST8*, *GSTU19*, *GST-TAU20*, and *GST14*). Several genes were commonly modulated in HLB

response datasets and the tolerance one: phloem protein 2 A5 (R genes), *GSTU7* (detoxifying pathways), *TGA1* (bZIP), *AIL5*, and *TINY2* (AP2-EBEPB transcription factors).

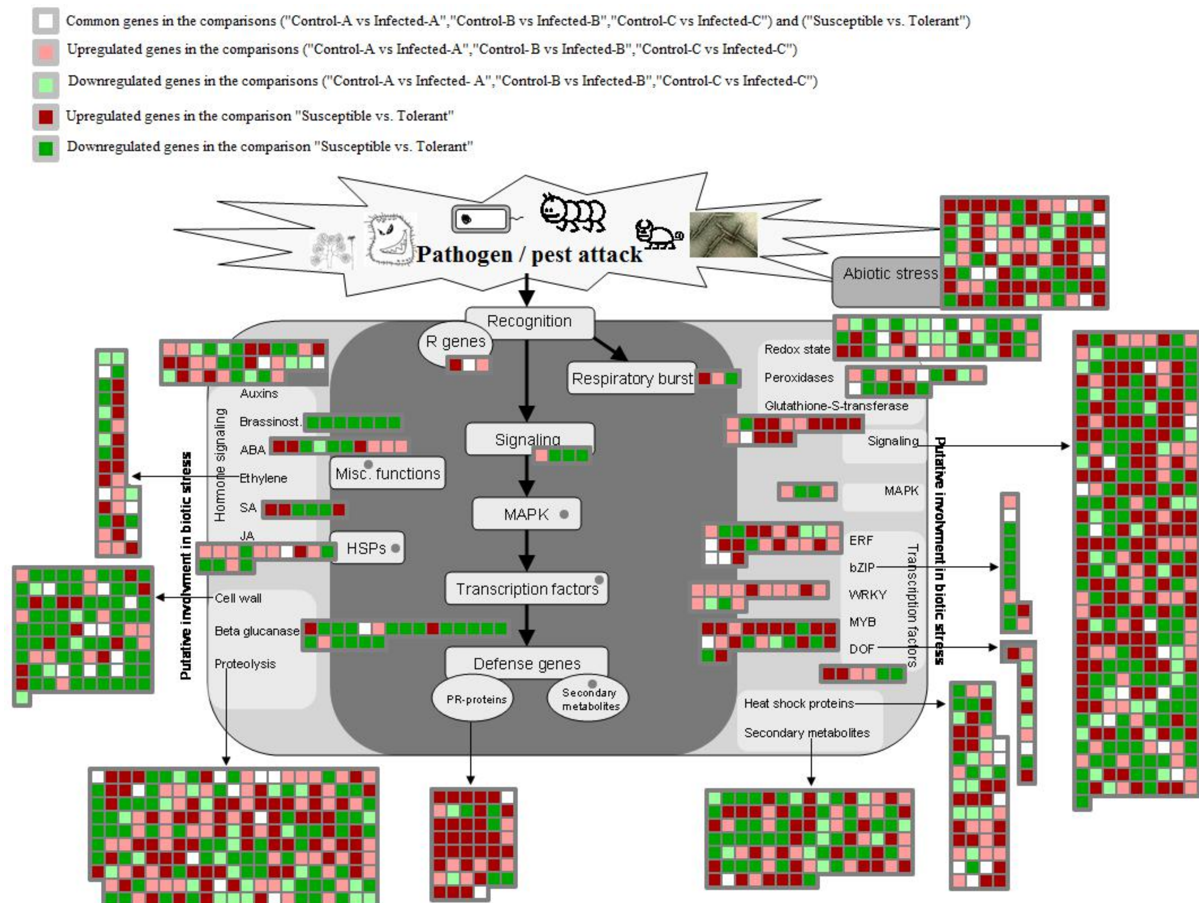


Figure 3.4.3.1–Biotic stress response genes showing differentially regulated genes between susceptible and tolerant genotypes. Genes commonly modulated between this dataset and the three related to HLB response are shown.

3.4.4 Large Enzyme Families

Among the large enzyme families, upregulation of glutathione S-transferases was linked with tolerance, as were oxidases (copper amine oxidase, NADP-dependent oxidoreductase, flavin-containing monooxygenase, and CTF2A). Genes involved in cytochrome P450-related reactions were more up-regulated than down-regulated in tolerant genotypes than in

susceptible ones. There was general repression of glucosidases, beta-1,3-glucan hydrolases, GDSL-lipases, and nitrilases. Several key genes belonging to large enzyme families were commonly regulated between tolerance and response: FAD-binding domain containing protein, glucose-methanol-choline, MES17, and two peroxidases.

3.5 Protein–Protein Network Analysis

Protein–protein interaction (PPI) network analysis based on an *Arabidopsis* knowledgebase compaired two pairwise comparisons performed on the same type of leaf tissue (young leaves): one related to HLB response (dataset B) and one linked with HLB tolerance (dataset D) (Figure 3.5.1). Four highly interactive proteins encoded by genes commonly regulated between the two datasets were identified: UBQ4, CYCD1-1, RPS19A, and STP1. A second PPI network analysis was performed to identify proteins commonly modulated by HLB in the three HLB-response datasets A, B, and C and HLB tolerance dataset D (Figure 3.5.2). Only a CSD2 protein was commonly present in the four pairwise comparisons. The comparison between the three leaf RNA-Seq datasets involved in HLB response showed that three heat shock proteins (HSP70-5, HSFB1, and HSP25.3) were encoded by genes that were significantly regulated in all three datasets.

Genes common in both "Control-B vs Infected-B" and "Susceptible vs Tolerant"
Genes present only in the comparison "Control-B vs Infected-B"
Genes present only in the comparison "Succptable vs Tolerant"

Figure 3.5.1–Protein–protein network analysis based on *Arabidopsis* knowledge base. Genes commonly modulated between datasets from Martinelli et al. (2013) (HLB response) and Wang et al. (2016) (HLB tolerance) are indicated. Genes present only in the comparison between healthy and infected in young leaves are also indicated (dataset B; Martinelli et al., 2013). Genes present in the comparison between susceptible vs. tolerant genotype are shown (Wang et al., 2016).

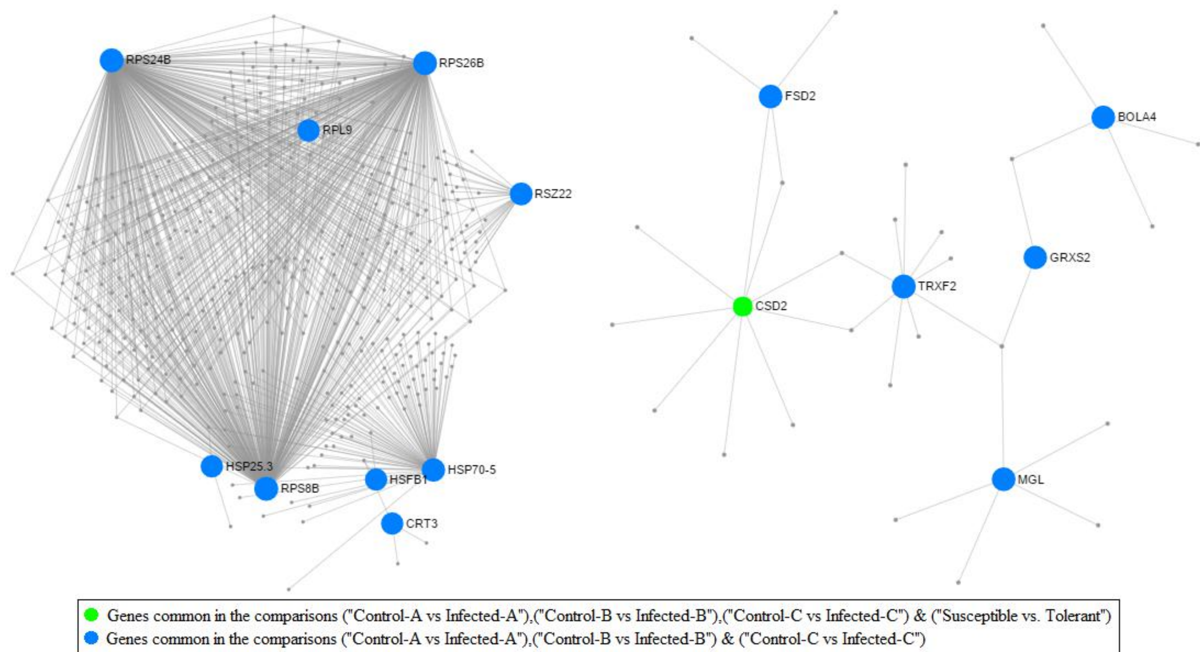


Figure 3.5.2–Genes commonly modulated between the three RNA-Seq datasets related to HLB response and all four datasets.

3.6 Splice Analysis

The online tool ASTALAVISTA3 predicted the splice events intron retention (IR), alternative splice donor (AD), alternative splice acceptor (AA), exon skipping (ES), and the combination of the above mentioned splice mechanisms from the eight samples. I observed that the splice event IR is the most abundant type (28.5–43.4%), followed by AA (16.1–29.1%), AD (8.3–13.6%) (Figure 3.6.1 and Supplementary Table S5). A considerable amount of the combination of splice events (classified as”Other Events,” 18.7–31.6%) were also observed from all the samples.

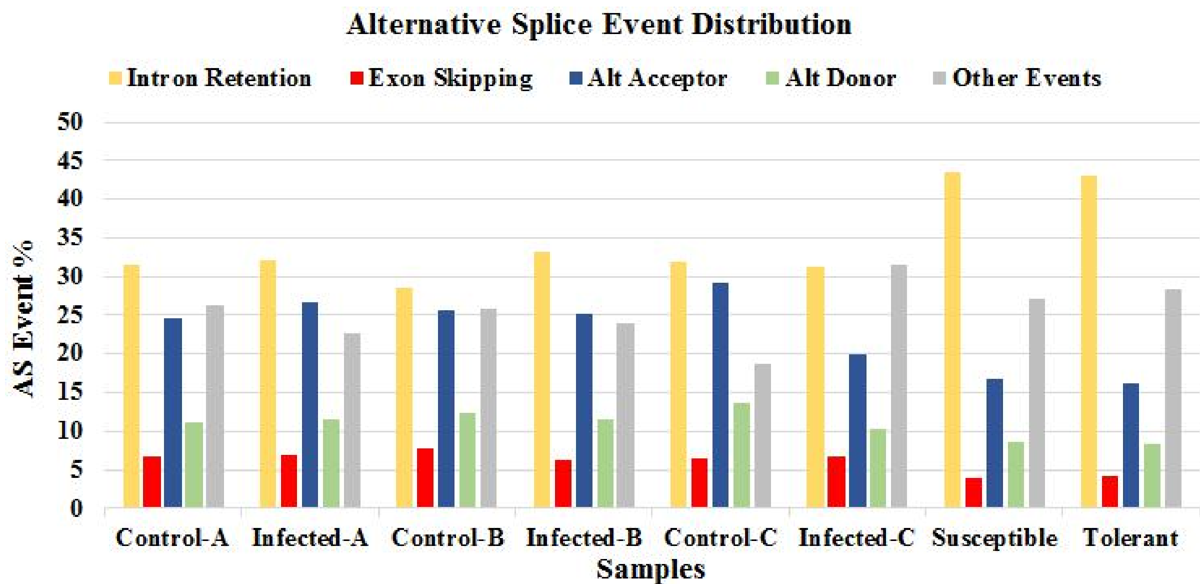


Figure 3.6.1–Distribution of the predicted splice events by ASTALAVISTA online tool for each sample. The splice categories are intron retention (IR), alternative splice donor (AD), alternative splice acceptor (AA), exon skipping (ES) and the combination of the above mentioned splice mechanisms (Other Events).

The complete AS events (AS landscapes) identified in this study can be downloaded from the link https://drive.google.com/drive/folders/1oorwtZmEcwSAs1n6x4g_fEGS-HOrxIbK?usp=sharing.

The data contain the information of exon-intron structure of the AS events, chromosomal coordinates, the IDs of the transcripts, involved in the given AS event. For the exon-intron structure of the AS event, ES is indicated by 1–2∧,0, alternative donor (AD) by 1∧,2∧, alternative acceptor (AA) by 1-,2- and IR by 1∧2-,08.

The differential regulations of alternative spliced forms of the commonly modulated genes between the four datasets were observed in response to HLB. The quantitative visualization of splice junction of the genes showing significant psi score difference was done using "sashimi_plot" program in MISO (Mixture of Isoforms) tool (https://miso.readthedocs.io/en/fastmiso/) along with uninfected sample as sashimi plot. The sashimi plot shows the number of reads corresponding to specific exon–exon junctions was labeled for each junction. I observed higher bayes factor for the splice event 'orange1.1g023621m.g.v1.1' in all datasets except "Control-C vs. Infected-C", which showed

that the isoform is more likely to be differentially expressed (Supplementary Table S5). Exon-skipping events found for the splice event ‘orange1.1g010747m.g.v1.1’ in samples ‘Control-A’ and ‘Control-C’. The splicing event ‘orange1.1g021628m.g.v1.1’ was not detected in the control sample ‘Control-C.’

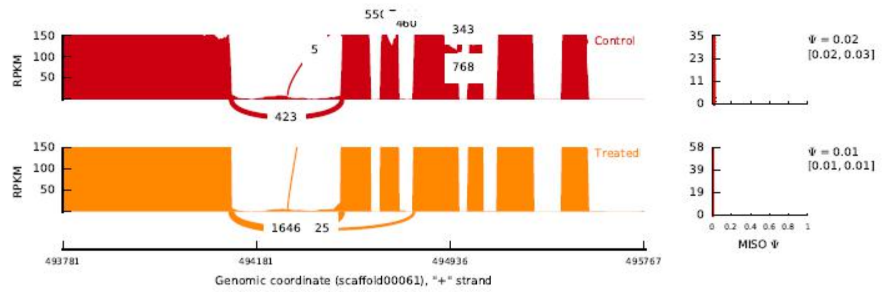
The MISO differential expression result files for each comparison were given below.

1) Control-A vs. Infected-A: https://drive.google.com/open?id=1caFTINtMPt2YAcS8IP7hNTRPpxsIBbk4
2) Control-B vs. Infected-B: https://drive.google.com/open?id=1g6PLt8sG3_RRs0LlZ8BwDNibczYvCSzs
3) Control-C vs. Infected-C: https://drive.google.com/open?id=1rW1o7JIifvE2Z00KS8z7cljb0k_tV6Uq
4) Susceptible vs. Tolerant: https://drive.google.com/open?id=17EhM6q96uDBrcO4wpij5xA96cKardt-x

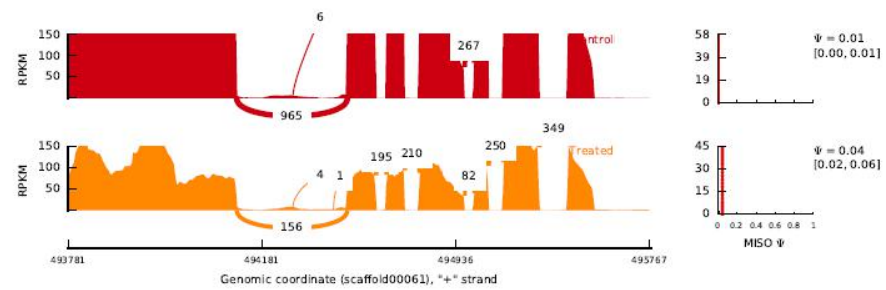
Multivariate analysis of transcript splicing (MATS; http://rnaseq-mats.sourceforge.net/) provides a statistical framework that determines the junction counts supporting the inclusion or the exclusion of specific splice events in Treated sample against Control. I ran MATS for all four comparisons and extracted the AS events only for the common genes. Only one gene ‘*orange1.1g023621m.g.v1.1*’ reported AS event alternative 5′ splice site (A5SS) in the rMATS results of all four comparisons (Figure 3.6.2) and reported a skipped exon (SE) AS event only in the comparison “Susceptible vs. Tolerant.”

Splice Event : orange1.1g023621m.g.v1.1

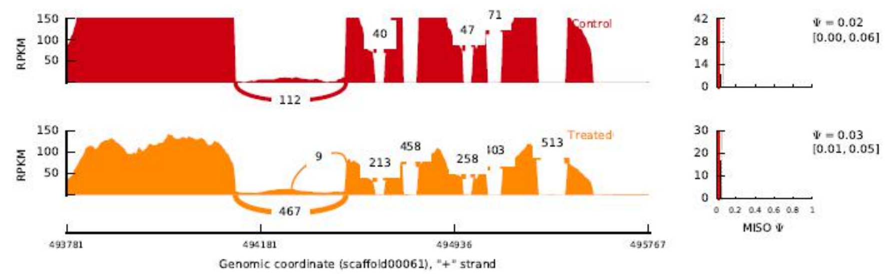
Control-A vs. Infected-A



Control-B vs. Infected-B



Control-C vs. Infected-C



Susceptible vs. Tolerant

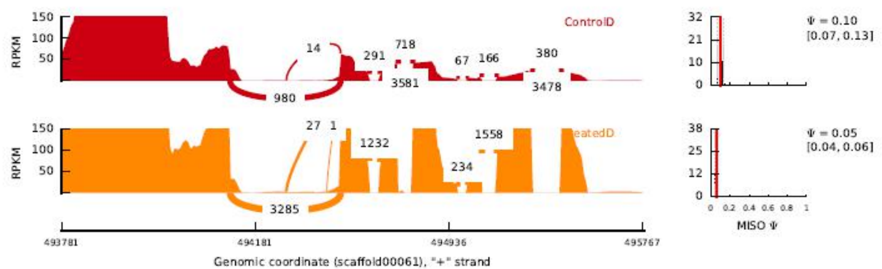


Figure 3.6.2–Splice event 'Orange1.1g023621m.g.v1.1'. The Sashimi plots of the splice event Orange1.1g023621m.g.v1.1 in all four datasets were plotted.

The rMATS result files for each comparison were given below.

1) Control-A vs. Infected-A: https://drive.google.com/open?id=1j-1JaOolzZAiqCbBYoj8Fq8UYGF23V4q
2) Control-B vs. Infected-B: https://drive.google.com/open?id=1Yy-zF_g3gq09Hd15cC6vYUtB2A9ik523
3) Control-C vs. Infected-C: https://drive.google.com/open?id=1QEQvoZZ4qGRCLpBKhvXh-yVnUoBHJXxd
4) Susceptible vs. Tolerant: https://drive.google.com/open?id=1Rvqbx7KF5-LDVN5APZw9qvU-bk4MXkuy

4. Discussion

The aim of this work was to identify genes and pathways commonly modulated by HLB disease in different published RNA-Seq datasets examining leaf tissues and tolerance mechanisms. The high variability in transcriptomic data requires more bioinformatic analysis. Most transcriptomic studies on HLB response were performed in only one season and using different agronomic, developmental and physiological conditions, weakening data reliability. My work compared all available RNA-Seq datasets related to HLB responses in *Citrus sinensis* leaf tissues. First, I compared three transcriptomic datasets performed on leaves infected by CaLas and then I sought common findings between these three studies and one examining HLB tolerance.

Although repression of photosynthesis and upregulation of starch and sucrose-related genes were observed in all three leaf datasets, few genes were commonly regulated. These data agreed with published findings that genes involved in photosynthetic reactions are generally down-regulated by HLB disease (Albrecht and Bowman, 2008; Kim et al., 2009; Martinelli et al., 2013). The comparison between the study on artificial infections (Fu et al., 2016) and that performed under field conditions (Martinelli et al., 2013) showed that different genes may have contrasting expression trends in the same tissue, young leaves. This implies that experimental conditions may drastically affect data, leading to contrasting conclusions.

The present study showed how different variables (developmental, agronomic and physiological conditions and infection method) affect expression of key genes in primary

metabolism. Some common features between the three transcriptomic datasets involved cell wall modifications. Six genes involved in cell wall-related pathways were commonly modulated by HLB in all three leaf datasets. I speculate that these genes may affect plant signaling responses to CaLas infection because of the role played by cell wall restructuring in sensing pathogen infections (Corwin and Kliebenstein, 2017).

Sugar and starch metabolism has been linked to a possible pathogenetic mechanism of CaLas (Martinelli and Dandekar, 2017; Rao et al., 2019). The induction of genes involved in sucrose degradation (invertase), starch biosynthesis (starch branching enzyme and starch synthase), and starch degradation (amyl-amylase, beta-amylase, and phosphoglucan water dikinase) were clearly induced by HLB in leaves (Albrecht and Bowman, 2008; Martinelli et al., 2013). Starch accumulation is a clear symptom of HLB progression in leaves (Bove, 2006). Unfortunately, my work found no HLB-induced gene involved in starch metabolism that was commonly modulated in different datasets. Only a beta-fructosidase involved in sucrose degradation was commonly modulated by HLB in the three datasets, perhaps due to the many differences in physiological, developmental, environmental, and agronomic conditions between the two studies (Martinelli et al., 2013; Fu et al., 2016). These findings confirmed the difficulty in finding common, specific host biomarkers to complement traditional diagnostic approaches relying on pathogen detection. Further works on RNA-Seq studies will demonstrate whether key markers can identify natural CaLas infections under field conditions.

Among all the hormone categories, only S-adenosylmethionine-dependent methyltransferase was commonly modulated by HLB in at least two of the three leaf datasets. Although more genes involved in hormonal crosstalk were expected be commonly regulated among studies, this evidence highlighted that SAR responses were activated in leaf tissues: an expected result, since CaLas is a biotrophic pathogen. The upregulation of several genes involved in jasmonic-mediated responses (*lox1*, *lox2*, and *lox3*) confirmed that typical defense responses against necrotrophic pathogens are induced by CaLas infection. A possible pathogenetic mechanism of CaLas is its modulation of hormonal-mediated defense responses for its own benefit (Martinelli et al., 2012, 2013; Martinelli and Dandekar, 2017; Rao et al., 2019). Although *MYC2*, a gene involved in jasmonic acid inhibition of salicylic acid responses, was not altered in any dataset, this present work confirmed a possible role for jasmonic acid in counteracting SAR responses in CaLas-infected leaves (Kazan and Manners, 2013). Abscissic acid and auxin genes can negatively affect SAR responses (Pieterse et al., 2009).

Although the upregulation of key genes involved in auxin biosynthesis, metabolism and response in artificially infected young leaves may inhibit salicylic acid responses, these genes were not affected in leaves under field conditions. Other auxin genes such as *GH3.1*, *GH3.9*, and *GH3.17* were induced (Martinelli et al., 2013). That no commonly regulated genes were found among the three transcriptomic datasets is another illustration of the high variability of transcriptomic data taken under different experimental conditions. Data obtained by Fu et al. (2016) highlighted the induction of two key genes involved in abscissic acid responses (GRAM domain containing protein and benzodiazepine receptor-related). Another gene, *HVA22*, was induced in field-grown young leaves. Taken together, these findings suggest that abscissic acid-related genes may aid pathogen colonization of the *Citrus sinensis* plant. The positive effect of gibberellins on SAR response is well known (Pieterse et al., 2009). Two genes involved in gibberellin pathways were up-regulated in field-grown mature leaves and one was induced in young leaves in the same study (Martinelli et al., 2013). Another hydrolase potentially involved in auxin pathways was induced in immature leaves (Fu et al., 2016). Taken together, these findings suggest that upregulation of gibberellin-related genes may compensate for negative effects of ABA and auxin on the SAR response. No gene involved in brassinosteroid was modulated by HLB in infected leaves. This contradicts findings that highlight the involvement of brassinosteroid-related genes on biotic stress responses.

Among transcription factors, *HB*, *AP2EREBP*, and Pseudo *ARR* were mostly induced in the dataset of Fu et al. (2016) while *GRAS* and *bHLH* were up-regulated in leaves analyzed by Martinelli et al. (2013). This evidence highlighted again how different experimental conditions affect expression of different key genes involved in CaLas responses. *WRKY*s are a family of transcription factors mostly involved in environmental plant stress responses (Jiang et al., 2016; Balan et al., 2017, 2018). Although only *WRKY48* was commonly regulated between the three transcriptomic leaf datasets, several *WRKY*s were highly up-regulated by HLB disease: five genes were induced in artificially infected young leaves, four in immature field-grown leaves and one in mature field-grown leaves. Because hundreds of different *WRKY*s are documented in crops (Rushton et al., 2010; Fan et al., 2015), only one *WRKY* may be involved in any specific biotic or abiotic stress. Specific *WRKY*s are induced by almost all environmental stresses and their expression is often tissue-specific (Jiang et al., 2016). However, analysis of this group of transcription factors clearly linked to

environmental stress may help early diagnosis of CaLas infections, when HLB disease is at an early stage. Although this must be confirmed by further experiments focusing on early disease stages, plant diagnostic approaches relying on host responses have been proposed (Ibáñez et al., 2014; Martinelli et al., 2016). This approach may complement traditional diagnostic methods based on PCR that target the pathogen, but will not replace them. This approach may be particularly helpful for plant diseases characterized by long incubation times.

Some genes involved in glutathione S-transferases were HLB-modulated in all three analyzed leaf datasets. More genes encoding pathogenesis-related proteins were up-regulated in the datasets of Fu et al. (2016) than in the one obtained by Martinelli et al. (2013). Artificial infection may induce a stronger response in infected *Citrus sinensis* than natural infection. This is also confirmed by the upregulation of *EDS1* after artificial, but not natural, infection. This gene is the receptor for salicylic acid-mediated responses (Parker et al., 1996). From my analysis, I speculate that the infection method (artificial or natural) deeply affects host responses to pathogen attack, driving diverse hormone-mediated defense responses.

Protein–protein interaction network analysis was conducted to identify which HLB-modulated genes play a key role at the PPI level in both HLB response and tolerance. The identification of three heat shock proteins commonly modulated between the three leaf HLB-response datasets confirmed their key role in disease progression and symptomatology (Martinelli et al., 2012, 2013; Martinelli and Dandekar, 2017).

In recent decades, the effective application of genetic engineering and genome editing technologies have substantially improved the ability to make precise changes in the genomes and to obtain disease-resistant (bacterial, fungal, and virus) crops (Sun et al., 2019). The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-mediated genome editing seems to be the most promising strategy to improve crop cultivars without introducing foreign genes which have the potential to be called non-GMO, can be cultivated and sold without regulatory monitoring (Jaganathan et al., 2018; Sun et al., 2019). In recent studies, the use of genome editing to target plant's susceptibility (S) genes for the development of transgene-free and durable disease-resistant crop varieties. Large-scale compatible as well as incompatible plant–pathogen interaction transcriptomes followed by comparative coexpression network analyses may discover such novel, nutritional immunity-related

sensitivity genes (or S genes) and tolerance genes (or T genes) in plants (Zaidi et al., 2018; Song et al., 2017). Thus a curated transcriptomic database of S genes and T genes in *Citrus sinensis* using the comprehensive transcriptome analysis dealing with plant pathogen interaction will be a fruitful resource for the plant researchers to choose the suitable target genes for the CRISPR/Cas9-mediated genome editing to create *Citrus* cultivars that are less susceptible to HLB.

This bioinformatic analysis highlights how different transcriptomic studies dealing with the same subject tend to show few commonly regulated genes. This may be due to the high environmental variability of field studies, leading to large differences in physiological and environmental conditions. However, identification of common features between studies helps clarify the role of CaLas in this devastating *Citrus sinensis* disease.

5. Conclusions

The present study deals with meta-analysis of transcriptomic studies related to Huanglongbing (HLB) response and tolerance in *Citrus* leaf tissues. It was found that several WRKY transcription factos were regulated and *WRKY48* was commonly modulated by CaLas in two of the three datasets. The genes encoding glutathione S-transferases such as *GST7* and *GST25* were upregulated and many genes involved in chromatin structure remodeling were commonly modulated. These responses helps the plant to recover the homeostatic state, which was disturbed due to CaLas infection. Gene co-expression network analysis confirmed a possible role for heat shock proteins by revealing the existence of highly inter-correlated stress-specific and consensus modules.

Altogether, the results from my study shows that different transcriptomic studies dealing with the same subject tend to show few commonly regulated genes, which may due to the high environmental variability of field studies, leading to large differences in physiological and environmental conditions. More works will be useful once more RNA-Seq datasets are available. A new bioinformatic analysis comparing microarray and RNA-Seq data is highly desirable such as previously performed in Citrus sinensis (Martinelli et al., 2015).

6. References

Aksenov A. A., Pasamontes A., Peirano D. J., Zhao W., Dandekar A. M., Fiehn O., et al. (2014). Detection of Huanglongbing disease using differential mobility spectrometry. Anal. Chem. 86, 2481–2488.

Albrecht U., Bowman K. D. (2008). Gene expression in *Citrus sinensis (L.)* Osbeck following infection with the bacterial pathogen *Candidatus* Liberibacter asiaticus causing Huanglongbing in Florida. Plant Sci. 175, 291–306.

Alvarez, S., Rohrig, E., Sol´ıs, D., and Thomas, M. H. (2016). Citrus greening disease (Huanglongbing) in Florida: Economic impact, management and the potential for biological control. Agric. Res. 5, 109-118.

Balan B., Caruso T., Martinelli F. (2017). Gaining insight into exclusive and common transcriptomic features linked with biotic stress responses in Malus. Front. Plant Sci. 8, 1569

Balan B., Marra F. P., Caruso T., Martinelli F. (2018). Transcriptomic responses to biotic stresses in *Malus x domestica*: a meta-analysis study. Sci. Rep. 8, 1970.

Bove J. M. (2006). Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. J. Plant Pathol. 88, 7–37.

Corwin J. A., Kliebenstein D. J. (2017). Quantitative resistance: more than just perception of a pathogen. Plant Cell 29, 655–665.

Dala-Paula B. M., Plotto A., Bai J., Manthey J. A., Baldwin E. A., Ferrarezi R. S., Gloria M. B. A. (2018). Effect of Huanglongbing or Greening Disease on Orange Juice Quality, a Review. Front Plant Sci. 9, 1976.

Dandekar A. M., Martinelli F., Davis C. E., Bhushan A., Zhao W., Fiehn O., et al. (2010). Analysis of early host responses for asymptomatic disease detection and management of specialty crops. Crit. Rev. Immunol. 30, 277–289.

Duan Y., Zhou L., Hall D. G., Li W., Doddapaneni H., Lin H., et al. (2009). Complete genome sequence of citrus Huanglongbing bacterium, ‘*Candidatus* Liberibacter

asiaticus' obtained through metagenomics. Mol. Plant Microbe Interact. 22, 1011–1020.

Duran-Vila N., Janse J.D., Foissac X., Melgarejo P. & Bové J.M (2014). Addressing the threat of huanglongbing in the mediterranean region: A challenge to save the citrus industry. Journal of Plant Pathology 96, S4.3-S4.8.

Fan X., Guo Q., Xu P., Gong Y., Shu H., Yang Y., et al. (2015). Transcriptome-wide identification of salt-responsive members of the *WRKY* gene family in *Gossypium aridum*. PLoS One 10, e0126148

Ferrarezi R. S., Qureshi J. A., Wright A. L., Ritenour M. A., Macan N. P. F. (2019). Citrus Production Under Screen as a Strategy to Protect Grapefruit Trees From Huanglongbing Disease. Front Plant Sci. 10, 1598.

Foissac S., Sammeth M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res. 35, W297–W299

Fu S., Shao J., Zhou C., Hartung J. S. (2016). Transcriptome analysis of sweet orange trees infected with '*Candidatus* Liberibacter asiaticus' and two strains of *Citrus Tristeza Virus*. BMC Genomics 17, 349.

Huang D. W., Sherman B. T., Lempicki R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Ibáñez A. M., Martinelli F., Reagan R. L., Uratsu S. L., Vo A., Tinoco M. A., et al. (2014). Transcriptome and metabolome analysis of citrus fruit to elucidate puffing disorder. Plant Sci. 21, 87–98.

Jaganathan D., Ramasamy K., Sellamuthu G., Jayabalan S., Venkataraman G. (2018). CRISPR for Crop Improvement: An Update Review. Front Plant Sci. 9, 985.

Jiang M., Liu Q., Liu Z., Li J., He C. (2016). Over-expression of a *WRKY* transcription factor gene *BoWRKY6* enhances resistance to downy mildew in transgenic broccoli plants. Australas. Plant Pathol. 45, 327–334.

Katz Y., Wang E. T., Airoldi E. M., Burge C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods 7, 1009–1015.

Kazan K., Manners J. M. (2013). MYC2: the master in action. Mol. Plant 6, 686–703.

Kim D., Langmead B., Salzberg S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360

Kim J. S., Sagaram U. S., Burns J. K., Li J. L., Wang N. (2009). Response of sweet orange (*Citrus sinensis*) to '*Candidatus* Liberibacter asiaticus' infection: microscopy and microarray analyses. Phytopathology 99, 50–57.

Manjunath K.L., Halbert S.E., Ramadugu C., Webb S., Lee R.F. (2008) Detection of 'Candidatus Liberibacter asiaticus' in Diaphorina citri and its importance in the management of citrus huanglongbing in Florida. Phytopathology. 98, 387-396.

Martinelli F., Dandekar A. M. (2017). Genetic mechanisms of the devious intruder *Candidatus* Liberibacter in citrus. Front. Plant Sci. 8, 904.

Martinelli F., Ibanez A. M., Reagan R. L., Davino S., Dandekar A. M. (2015). Stress responses in citrus peel: comparative analysis of host responses to Huanglongbing disease and puffing disorder. Sci. Hortic. 192, 409–420.

Martinelli F., Reagan R. L., Dolan D., Fileccia V., Dandekar A. M. (2016). Proteomic analysis highlights the role of detoxification pathways in increased tolerance to Huanglongbing disease. BMC Plant Biol. 16, 167.

Martinelli F., Reagan R. L., Uratsu S. L., Phu M. L., Albrecht U., Zhao W., et al. (2013). Gene regulatory networks elucidating Huanglongbing disease mechanisms. PLoS One 8, e74256.

Martinelli F., Uratsu S. L., Albrecht U., Reagan R. L., Phu M. L., Britton M., et al. (2012). Transcriptome profiling of citrus fruit response to Huanglongbing disease. PLoS One 7, e38039.

Parker J. E., Holub E. B., Frost L. N., Falk A., Gunn N. D., Daniels M. J. (1996). Characterization of eds1, a mutation in *Arabidopsis* suppressing resistance to

Peronospora parasitica specified by several different *RPP* genes. Plant Cell 8, 2033–2046.

Pieterse C. M., Leon-Reyes A., Van der Ent S., Van Wees S. C. (2009). Networking by small-molecule hormones in plant immunity. Nat. Chem. Biol. 5, 308–316.

Rao M. J., Ding F., Wang N., Deng X., Xu Q. (2019). Metabolic Mechanisms of Host Species Against Citrus Huanglongbing (Greening Disease). Critical Reviews in Plant Sciences, 1–16.

Rawat N., Kiran S. P., Du D., Gmitter F. G., Deng Z. (2015). Comprehensive meta-analysis, co-expression, and miRNA nested network analysis identifies gene candidates in citrus against Huanglongbing disease. BMC Plant Biol. 15, 184.

Rizzini F. M., Bonghi C., Chkaiban L., Martinelli F., Tonutti P. (2010). Effects of postharvest partial dehydration and prolonged treatments with ethylene on transcript profiling in skins of wine grape berries. Acta Hortic. 877, 1099–1104.

Rushton P. J., Somssich I. E., Ringler P., Shen Q. J. (2010). *WRKY* transcription factors. Trends Plant Sci. 15, 247–258.

Saia S., Ruisi P., Fileccia V., Di Miceli G., Amato G., Martinelli F. (2015). Metabolomics suggests that soil inoculation with *Arbuscular mycorrhizal fungi* decreased free amino acid content in roots of durum wheat grown under N-Limited, P-Rich field conditions. PLoS One 10, e0129591.

Shen S., Park J. W., Huang J., Dittmar K. A., Lu Z. X., Zhou Q., et al. (2012). MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. 40, e61

Song X., Bhattarai K., Lv D., Gao F., Ying X. (2017). Can CRISPR Win the Battle against Huanglongbing? J. Plant Pathol. Microbiol. 8, 2.

Sun L., Nasrullah, Ke F., Nie Z., Wang P., Xu J. (2019). Citrus Genetic Engineering for Disease Resistance: Past, Present and Future. Int J Mol Sci. 20, E5256.

Thimm O., Bläsing O., Gibon Y., Nagel A., Meyer S., Krüger P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939.

Tosetti R., Martinelli F., Tonutti P., Barupal D. K. (2012). Metabolomics approach to studying minimally processed peach (*Prunus persica*) fruit. Acta Hortic. 934, 1017–1022.

Tyler H. L., Roesch L. F., Gowda S., Dawson W. O., Triplett E. W. (2009). Confirmation of the sequence of '*Candidatus* Liberibacter asiaticus' and assessment of microbial diversity in Huanglongbing-infected citrus phloem using a metagenomic approach. Mol. Plant Microbe Interact. 22, 1624–1634.

Usadel B., Nagel A., Steinhauser D., Gibon Y., Bläsing O. E., Redestig H., et al. (2006). PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics 7, 535.

Wang, N., Pierson, E.A., Setubal, J.C., Xu, J., Levy, JG., Zhang, Y., Li, J., Rangel, LT., and Martins, J. (Jr). (2017). The Candidatus Liberibacter–host interface: insights into pathogenesis mechanisms and disease control. Annu. Rev. Phytopathol. 55, 20.1–20.32.

Wang Y., Zhou L., Yu X., Stover E., Luo F., Duan Y. (2016). Transcriptome profiling of Huanglongbing (HLB) tolerant and susceptible Citrus plants reveals the role of basal resistance in HLB tolerance. Front. Plant Sci. 7, 933.

Xia J., Benner M. J., Hancock R. E. (2014). NetworkAnalyst–integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. 42, W167–W174.

Zaidi S. S., Mukhtar M. S., Mansoor S. (2018). Genome Editing: Targeting Susceptibility Genes for Plant Disease Resistance. Trends Biotechnol. 36, 898-906.

General Conclusion

In a meta-analysis, the selection of candidate differentially expressed genes by the comparison of samples from individual studies was significant, before combining the data across studies. The overlapping of lists of differentially expressed genes may lead to potentially biased conclusions because of two reasons. First, the genes with the same pattern of expression in the majority of samples, but failed to surpass the minimum threshold in a few samples may not be detected with this approach. Second, the genes that may exhibit differential expression in more than one experiment may not reach differential expression when all the data across experiments is considered because the variation across experiments is higher than the variation within the experiment. More research into the application of meta-analysis to RNA-Seq studies remains necessary to help in identifying the most effective approach for the varying gene expression scenarios.

Meta-analyses allow an increase in reliability of transcriptomic data, reducing environmental variability due to a low number of biological replicates and repeated experiments. My first meta-analysis work conducted in *Malus x domestica*, highlights the role of *WRKY*s in the molecular response to biotic stresses at both transcript and protein-protein interaction levels. Although *WRKY40* was involved response to both fungal pathogens and *E. amylovora*, its interaction with other different *WRKY* may induce specific responses. In response to fungal pathogens, *WRKY* interacted with two other pathogen-regulated *WRKY6* and *WRKY18* while in response to *E. amylovora* it interacts with *WRKY33*. Specific hormones were differentially affected between the three types of stresses and drives to specific defense responses. Future studies in other crops investigating similar diseases will allow validate these findings and identify resistance mechanisms in gene regulatory networks of plant-microbe interactions.

In the second meta-analysis, I collected 12 transcriptomic works in *Malus x domestica* in order to identify which key genes, proteins, gene categories are involved in general plant pathological conditions and those features linked with exclusive biotic stress responses. Different transcriptomic studies are generally performed using different transcriptomic platforms and using different experimental design. In order to compare them, it is necessary to use at least the same bioinformatic pipeline. A pipeline composed by pathway and gene set enrichment analysis, protein-protein interaction networks and gene visualization tools were employed. Those genes that are only related to molecular responses to pathogen attacks and

those linked with other plant physiological processes were identified. Gene set enrichment analysis pointed out that photosynthesis was inhibited by *Erwinia amylovora* and fungal pathogens. Different hormonal crosstalk was linked with responses to different pathogens. Gibberellin-related pathways, ABA-related were mostly repressed by fungal pathogens. Relating to transcription factors, genes encoding MYBs and WRKY2 were down-regulated by fungal pathogens and 12 *WRKY*s were commonly regulated by different biotic stresses. The protein-protein interaction analysis discovered the presence of several proteins affected by more than one biotic stress including a WRKY40 and some highly interactive proteins such as heat shock proteins. This study represents a first preliminary curated meta-analysis of apple transcriptomic responses to biotic stresses.

In my final meta-analysis, I collected the raw data related to transcriptome studies dealing with Huanglongbing disease in *Citrus sinensis*. After the transcriptome analysis using the developed pipeline, I identified only 16 HLB-regulated genes which were commonly identified between the three leaf datasets. Among them were key genes encoding proteins involved in cell wall modification such as CESA8, pectinesterase, expansin8, expansin beta 3.1, and a pectate lyase. Fourteen HLB-regulated genes were in common between all four datasets. Gene set enrichment analysis showed some different gene categories affected by HLB disease. Although sucrose and starch metabolism was highly linked with disease symptoms, different genes were significantly regulated depending on leaf growth and infection stages and experimental conditions. Histone-related transcription factors were highly affected by HLB in the analyzed RNA-Seq datasets. HLB tolerance was linked with the induction of proteins involved in detoxification. Protein-protein interaction (PPI) network analysis confirmed a possible role for heat shock proteins in curbing disease progression.

Classification of samples into groups according to source, behaviors, treatments, and stages based on gene expression profiles is an important step in the meta-analysis on transcriptome studies because it will diminish the biases associated with individual studies. The re-analysis using an updated bioinformatics pipeline plays a crucial role in the data normalization of the list of differentially expressed genes. There is a high possibility to detect new genes using re-analysis due to a) updated bioinformatics tools minimize the errors b) usage of the updated reference genome, which results in the high alignment percentage of reads and in turn results more differentially expressed genes. In contrast, it helps us to identify robust classifier genes that overcome the limitations of previous approaches. A curated transcriptomic database will

be a fruitful resource for the plant researchers to understand the transcriptomic changes due to different stress condition in plants. This comprehensive meta-analysis study is a preliminary step for the creation of such a curated database in future.

I present my work as a pilot project : meta-analysis of diverse transcriptomic data sets is a well-grounded and robust approach to develop hypotheses for how plants respond to biotic stress in general. The analysis I describe enables researchers to investigate stress responses in other plants even with limited stress-responsive transcriptome data, with multiple tissue types and few replicated per treatment.

Appendix

1. TrimSeq.pl

Purpose: This perl script trim 'N" bases from 5' and 3' end of the input fastq files.

To execute: ***perl TrimSeq.pl Input.fastq 10 5***

(Trim 10 bases from 5' end and 5 bases from 3' end of the fastq file 'Input.fastq')

Perl Script:

```
#!/usr/bin/perl -w

use strict;



# Receiving input parameters #

my $infile = $ARGV[0];

my $start = $ARGV[1];

my $end = $ARGV[2];

chomp ( $infile,, $start,$end);



my $fileName = (split'\.',((split'\/',$infile,999)[-1]),999)[0];

my $ResultFile = "Trimmed_$fileName.fastq";



# If the fastq file is compressed#

if ($infile =~ /\.gz$/)

{

        open(IN,"gunzip -c $infile |") or die "can't open $infile for reading";

}
```

```
my $lineNo = 0;

open OUT,">$ResultFile" or die "Can't open $ResultFile for writing\n";

while(<IN>)

{

        chomp;

        next if(/^\s*$/);

        $lineNo++;

        if( ($lineNo%2 == 0 ) || ($lineNo%4 == 0 ) )

        {

                my $TrimmedSeq = reverse(unpack("x$end A*",reverse(unpack("x$start
A*",$_))));

                print OUT "$TrimmedSeq\n";

        }

        else

        {

                print OUT "$_\n";

        }

}

close IN;

close OUT;

print "Trimming of File : $fileName is completed !!\n";

exit;
```

2. CheckAfterAdapterTrimming.pl

Purpose: This perl script will remove the reads having length less than 30 bases after the adapter trimming is done. The bases with length less than 30 bases won't influence the influende in alignment to the reference genome. The script will work only for the paired end data

To execute: ***perl CheckAfterAdapterTrimming.pl Adaptertrimmed_R1.fastq Adaptertrimmed_R2.fastq Preprocessed_R1.fastq Preprocessed_R2.fastq***

(The input fastq files are 'Adaptertrimmed_R1.fastq' and 'Adaptertrimmed_R2.fastq' and the result files will be 'Preprocessed_R1.fastq' and 'Preprocessed_R2.fastq')

Perl Script:

```
#!/usr/bin/perl -w

use strict;


#Receiving input fastq files (paired end) #

open(INFP1,"<$ARGV[0]");

open(INFP2,"<$ARGV[1]");


# Openingoutput fastq files (pass output file names) #

open(OUTFP1,">$ARGV[2]");

open(OUTFP2,">$ARGV[3]");
```

```
my $Problems = 0;

while(my $r1_1=<INFP1>) {

my $r1_2=<INFP1>; my $r1_3=<INFP1>; my $r1_4=<INFP1>;

my $r2_1=<INFP2>; my $r2_2=<INFP2>; my $r2_3=<INFP2>; my $r2_4=<INFP2>;

        chomp($r1_1,$r1_2,$r1_3,$r1_4);

        chomp($r2_1,$r2_2,$r2_3,$r2_4);

my @a1 = split(" ",$r1_1);

my @a2 = split(" ",$r2_1);

        if($a1[0] eq $a2[0] && length($r1_2) >= 30 && length($r2_2) >= 30) {

                print OUTFP1 "$r1_1\n$r1_2\n$r1_3\n$r1_4\n";

                print OUTFP2 "$r2_1\n$r2_2\n$r2_3\n$r2_4\n";

        }

        elsif($a1[0] ne $a2[0]) {

                print "Problem\n";

                $Problems++;
```

```
            exit;
        }
}
close(INFP1);
close(INFP2);
close(OUTFP1);
close(OUTFP2);
if($Problems == 0)
{
        unlink $ARGV[0];
        unlink $ARGV[1];
}
```

2. ExtractDEGs.pl

Purpose: The script will extract the up- and down-regulated genes from cuffdiff comparison result files. In the script, the user should edit the input samples and can change the p-val cutoff. The result files will be the separate up- and down-regulated files and a fuke contain the statistics.

To execute: ***perl ExtractDEGs.pl PATH_TO_CUFFDIFF_RESULT_FOLDER***

(please provide the full path to cuffdiff result folder)

Perl Script:

```
my @InputFolders = qw($ARGV[0]); # cuffdiff result folder path #

my $cutoff = '0.05';

my %Samples = qw(sampleshortname1 sample1 sampleshortname2 sample2
sampleshortname3 sample3);

my %SampleMapping = ();

my %Stat = ();
```

```
foreach my $eachTypes(keys %Samples)

{

	foreach my $samples(keys %Samples)

	{

		if($eachTypes ne $samples)

		{

			$SampleMapping{$eachTypes}{$samples} =
"$Samples{$eachTypes}\_$Samples{$samples}";

		}

	}

}

my $ResultFolder = 'Results'.$cutoff;

my $StatResult = 'Sample_Regulation_Statistics_'.$cutoff.'.txt';


unless(-d $ResultFolder)

{

	mkdir $ResultFolder;

}
```

```
foreach my $eachFolder(@InputFolders)
{
	my $SeqType = 'All';
	my $ResultFolder1 = "$ResultFolder\/$SeqType";
	unless(-d $ResultFolder1)
	{
		mkdir $ResultFolder1;
	}
	foreach my $eachFiles(glob("$eachFolder/*_*.diff"))
	{
		my $type = (split'\_',((split'\/',$eachFiles,999)[-1]),999)[-2];
		if(($type eq 'isoform') || ($type eq 'gene') )
		{
			my $ResultFolder2 = "$ResultFolder1\/$type";
			unless(-d $ResultFolder2)
			{
				mkdir $ResultFolder2;
			}
```

```
&ParseFile($eachFiles,$type,$SeqType,$ResultFolder2,\%SampleMapping,\%Stat);
                }
        }
}


open OUT,">$StatResult" or die "Can't open $StatResult for writing\n";
print OUT "TYPE\tCONTROL\tSAMPLE\tUP REGULATED\tDOWN
REGULATED\tTOTAL\n";


foreach my $eachtypes(keys %Stat)
{
        foreach my $samplecompare(sort{$a cmp $b;} keys %{$Stat{$eachtypes}})
        {
                my ($S1,$S2) = split'\_',$samplecompare,999;
```

```
		my $upCount = 0;
		my $downCount = 0;
		my $totalCount = 0;
		$upCount = $Stat{$eachtypes}{$samplecompare}{'UP'}
if(defined($Stat{$eachtypes}{$samplecompare}{'UP'}));
		$downCount = $Stat{$eachtypes}{$samplecompare}{'DOWN'}
if(defined($Stat{$eachtypes}{$samplecompare}{'DOWN'}));
		$totalCount = $Stat{$eachtypes}{$samplecompare}{'TOTAL'}
if(defined($Stat{$eachtypes}{$samplecompare}{'TOTAL'}));
		print OUT "$eachtypes\t$S1\t$S2\t$upCount\t$downCount\t$totalCount\n";
	}
}
close OUT;


sub ParseFile
{
	my $InputFile = shift;
	my $type = shift;
	my $SeqType = shift;
	my $ResFolder = shift;
	my $RefSampleMapping = shift;
	my $RefStat = shift;
	my $FileHeader = '';
```

```
#print "$InputFile\t$type\t$SeqType\t$ResFolder\n";
my %ExpressionInfo = ();
open IN,"<$InputFile" or die "Can't open $InputFile for reading\n";
while(<IN>)
{
    chomp;
    next if(/^\s*$/);
    if(/^\s*test_id.*/)
    {
        $FileHeader = $_;
        next;
    }
    my @data = split"\t",$_,999;
    if(defined($$RefSampleMapping{$data[4]}{$data[5]}))
    {
        if( ( ($data[7]>=1) && ($data[8]>=1)) && ($data[11]<=0.05)) # Checking
P-Value #
        {

if(defined($ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'TOTAL'}{$data[1]
}))
```

```
                {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'TOTAL'}{$data[1]}.="\n$_";
                }
                else
                {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'TOTAL'}{$data[1]} = $_;
                }
                if($data[8]>$data[7]) # UP Regulated Gene/isoform #
                {
if(defined($ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'UP'}{$data[1]}))
                    {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'UP'}{$data[1]}.="\n$_";
                    }
                    else
                    {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'UP'}{$data[1]} = $_;
                    }
                }
```

```
                    if($data[8]<$data[7]) # DOWN Regulated Gene/isoform #
                    {
if(defined($ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'DOWN'}{$data[1]
}))
                         {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'DOWN'}{$data[1]}.="\n$_";
                         }
                         else
                         {
$ExpressionInfo{$$RefSampleMapping{$data[4]}{$data[5]}}{'DOWN'}{$data[1]} = $_;
                         }
                    }
               }
          }
     }
     close IN;
```

```
# Writing To Files #
foreach my $sampleComparison(keys %ExpressionInfo)
{
    my $ResFolder1 = "$ResFolder\/$sampleComparison";
    unless(-d $ResFolder1)
    {
        mkdir $ResFolder1;
    }
    foreach my $regulationType(keys %{$ExpressionInfo{$sampleComparison}})
    {
        my $ResultFile = "$ResFolder1\/$sampleComparison\_$regulationType.txt";
        my $dataType = (split'\/',$ResultFile,999)[2];
        my $ResultFile1 =
"$ResFolder1/$sampleComparison\_$regulationType\_$dataType\_exp_filtered.diff";
        print
"$sampleComparison\t$regulationType\t\t$ResultFile\t$sampleComparison\t<$ResultFile1>
\n";
        open OUT,">$ResultFile" or die "Can't open $ResultFile for writing\n";
        open OUT1,">$ResultFile1" or die "Can't open $ResultFile1 for writing\n";
        print OUT1 "$FileHeader\tRegulation\n";
        my $count = 0;
```

```
foreach my $Genes(sort{$a cmp $b;} keys
%{$ExpressionInfo{$sampleComparison}{$regulationType}})
{
	$count++;
	print OUT "$Genes\n";
	foreach my
$filedata(split'\n',$ExpressionInfo{$sampleComparison}{$regulationType}{$Genes})
	{
		my @data5 = split"\t",$filedata,999;
		$data5[4] = $Samples{$data5[4]} if(defined($data5[4]));
		$data5[5] = $Samples{$data5[5]} if(defined($data5[5]));
		my $infos = join"\t",@data5;
		my $regtype1 = 'NA';
		if(defined($ExpressionInfo{$sampleComparison}{'UP'}{$Genes}))
		{
			$regtype1 = 'Up';
		}

elsif(defined($ExpressionInfo{$sampleComparison}{'DOWN'}{$Genes}))
		{
			$regtype1 = 'Down';
		}
```

```
                    print OUT1 "$infos\t$regtype1\n";
                }
            }
            close OUT;
            close OUT1;
            $$RefStat{$type}{$sampleComparison}{$regulationType} = $count;
        }
    }
}
```

Acknowledgements

I praise God, the Almighty for providing me with this opportunity and granting me the capability to proceed successfully. I am very grateful to the many people who have been such great inspiration, support and help to me during this project, and my graduate career as a whole.

First and foremost, thank you to my doctoral advisors Professor Tiziano Caruso and Professor Federico Martinelli, for their scientific supervision, hard work, constant guidance, cooperation, patience, advice and motivation. Professor Tiziano Caruso has been a vital part to all of my doctoral successes, and a great role model and inspiration for all students in academia.

Thank you to Professor Francesco Paolo Marra and Dr Annalisa Marchese for continuing to support me beyond my PhD project and entrusting me with instructing their knowledge in alternative bearing mechanism in Pistachio. I am also very grateful to Dr. Annalisa Marchese for helping me enormously for the improvisation and final formatting of this dissertation.

Thank you to my PhD coordinator, Professor Vincenzo Bagarello and Carmelo Priolo for supported me throughout the doctoral process.

Thank you to Professor Abhaya M Dandekar for his support and words of encouragement and for giving me the possibility to work in his laboratory at the University of California, Davis.

Thank you to my lab colleagues Jubina, Silvia, Emilio, Rafael and Sebastiano for their support.

And last but certainly not least, my family, who has supported me throughout my academic trajectory despite the self-alienation that came as a result of pursuing my goals.