The Pennsylvania State University

The Graduate School

Genetics Program

GENOMICS OF THE *THEOBROMA CACAO L*. DEFENSE RESPONSE

A Dissertation in

Genetics

by

Andrew S. Fister

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2016

The dissertation of Andrew S. Fister was reviewed and approved* by the following:

Mark J. Guiltinan
Professor of Plant Molecular Biology
Dissertation Advisor

Majid Foolad
Professor of Plant Genetics
Chair of Committee

Siela Maximova
Professor of Horticulture

James Marden
Professor of Biology

Charles Anderson
Assistant Professor of Biology

Robert Paulson
Professor of Veterinary and Biomedical Sciences
Genetics Program Chair

*Signatures are on file in the Graduate School

ABSTRACT

Theobroma cacao, the source of cocoa and a cash crop of global economic importance, suffers significant annual losses due to several pathogens. While study of the molecular mechanisms of defense in cacao has been limited, the recent sequencing of two cacao genomes has greatly expedited the ability to study genes and gene families with roles in defense. Here, the pathogenesis-related (PR) gene families were bioinformatically identified, and family size and gene organization were compared to other plant species, revealing significant conservation throughout higher monocots and dicots. Expression of the PR families was also analyzed using a whole genome microarray to measure transcriptomic regulation in leaves after treatment of cacao seedlings with two pathogens, identifying the induced PR genes within each family. We found significant overlap between the PR genes induced by the pathogens, and subsequent qRT-PCR revealed up to 5000-fold induction of specific PR family members.

Next, the regulation of the defense response in cacao by salicylic acid, a major defense hormone, was analyzed. The study focused on two genotypes, the broadly resistant Scavina 6 and the widely susceptible ICS1. First, treatment of leaves of two cacao genotypes with salicylic acid was shown to enhance resistance of both. Moreover, overexpression of TcNPR1, a master regulator of systemic acquired resistance, is also shown to enhance the defense response, supporting the importance of salicylic acid and its downstream targets in cacao immunity. Microarray analysis of the transcriptomic response to salicylic acid revealed genotype-specific responses to hormone treatment. ICS1 appeared to show a more canonical response to salicylic acid, with more PR genes induced, while Scavina 6 exhibited increased expression of chloroplastic and mitochondrial genes. It was hypothesized that this induction was linked to increased ROS production, and subsequent ROS staining experiments confirmed higher concentration of superoxide in salicylic acid-treated Scavina 6 leaf tissue.

Third, a pilot study was performed to quantify genetic variability within defense genes. Using DNA samples representing three populations of cacao – Peruvian, Ecuadorian, and French Guianan – we amplified three genes involved in defense, two predicted to be more variable (cysteine-rich repeat secretory peptide 38 and a polygalacturonase inhibitor) and one predicted to harbor less polymorphism (pathogenesis-related 1). Population genetic analysis of variability suggested that the gene predicted to be more variable may be under diversifying selection, suggesting that they may directly interact with rapidly evolving pathogen proteins. The experiment validated previously described observations about the populations, in particular that the French Guianan population was less variable than the others. The study also supported the predictions regarding gene variability, indicating that our strategy for identifying genes with more variation appears to be applicable but will require further validation.

The Guiltinan-Maximova lab developed a protocol for transient transformation of cacao leaf tissue, which has been applied to characterizing gene function in several published analyses. Here the highly efficient protocol is presented in full, along with data collected in a series of

optimization experiments. We also use the protocol to demonstrate the effect of overexpression of a cacao chitinase after subsequent infection with *Phytophthora* mycelia.

A preliminary study describing a strategy for selection of high-priority candidate genes for functional characterization is described. Six genes were cloned and overexpressed using the transient transformation protocol; and while the study showed the ability of our protocol to significantly increase transcript abundance of the gene of interest, it did not validate the role of any of the genes in defense by showing decreased susceptibility.

This dissertation contributes to the study of genomics and molecular mechanisms of defense in four key ways: 1) 15 classes of defense genes are identified and their expression dynamics are characterized, 2) genotype-specific differences in defense response are identified, providing insight into different strategies for survival, 3) variability within defense genes is discovered, differentiating populations of cacao and providing evidence for diversifying selection, and 4) a rapid and efficient strategy for gene functional analysis, which will enhance future genetic analyses in cacao, is presented.

TABLE OF CONTENTS

List of Figures.....viii

List of Tables.....xiii

Preface.....xiv

Foreword and Acknowledgements.....xvi

Chapter 1 : Literature Review.....1

1.1. *Theobroma cacao*: Cultivation and biology.....1
Production and horticulture.....1
Physiology and taxonomy.....2
Genetics and genomics.....3
1.2. Pathogens of cacao.....4
Phytophthora spp. – Black pod rot.....4
Moniliophthora roreri – Frosty pod rot.....5
Moniliophthora perniciosa – Witches' broom.....5
Oncobasidium theobromae – Vascular streak dieback (VSD).....6
Cacao swollen shoot virus (CSSV).....7
Improving resistance to pathogens.....7
1.3. Plant-pathogen interactions and the defense response.....8
Pathogen recognition overview.....8
PAMP triggered immunity – Basal, broad spectrum defense.....9
Effector proteins suppress basal defenses.....10
Effector triggered immunity – Adapted defenses for co-evolved pathogens.....11
Hypersensitive response.....13
Phytohormones and the systemic response.....13
Induced defenses against microbes.....16
Durability of defense and immune memory.....17
Cacao and molecular studies of defense.....18
1.4. Evolution and plant defense mechanisms.....19
Plant genomes and their evolution.....19
Plant-pathogen coevolution.....21
1.5 Dissertation Overview.....22

Chapter 2: *Theobroma cacao L*. Pathogenesis-Related Gene Tandem Array Members Show Diverse Expression Dynamics in Response to Pathogen Colonization.....39

Abstract.....39
Background.....40
Materials and Methods.....44
Results.....48
Discussion.....58

Conclusions 65

Chapter 3: Two *Theobroma cacao* Genotypes with Contrasting Pathogen Tolerance Show Aberrant Transcriptional and ROS Responses after Salicylic Acid Treatment 73

Abstract 73
Introduction 74
Materials and methods 76
Results 81
Discussion 92
References 95

Chapter 4: Genetic Diversity Assessment of Three Cacao Populations using SNP Identification in Putative Defense Genes 101

Abstract 101
Background 101
Materials and Methods 103
Results 106
Discussion 115
References 119

Chapter 5: Protocol: Transient expression system for functional genomics in the tropical tree *Theobroma cacao* L. 123

Abstract 123
Background 123
Experimental Design 125
Reagents and Equipment 132
Protocol 134
Conclusions 141
References 142

Chapter 6: Retrospective 145

Outro 145
Defining cacao's induced defenses 146
Scavina 6 and ICS1: Unique strategies for (partial) tolerance 147
Assessing germplasm for variation in defense genes 149
Transient transformation: a strategy for rapid gene assessment 150
Gene prioritization strategy 151
In conclusion… 152
References 153

Appendix A: A strategy for functional characterization of defense genes 155

Introduction 155
Methods 158
Results 160

Discussion 164
Conclusions 165
References 166

Appendix B: Supplemental Data 169

Chapter 2 169
Chapter 3 171
Chapter 4 172
Chapter 5 177

LIST OF FIGURES

Figure 2.1 - Karyogram depicting position of PR genes along the length of chromosomes based on the Criollo genome sequence. Tandem arrays are labeled above the chromosomes with gene family and number of genes in the array in parentheses. Length of chromosomes is shown in Mb. Due to resolution of the image, lines representing nearby genes partially overlap……52

Figure 2.2 - Scatterplots comparing PR gene family size in the in the Criollo *T. cacao* genome to five plant species and the Matina *T. cacao* genome……53

Figure 2.3 - Maximum-likelihood phylogeny of Criollo and Arabidopsis PR-3 family members. Node labels represent bootstrap support from 100 replicates. Brackets denote members of tandem arrays. Arrows indicate cases where non-tandem array members group most-closely with a tandem array member. Branch lengths represent genetic distance in substitutions per site. AT5G05460, a cytosolic beta-endo-N-acetyglucosaminidase and member of the chitinase superfamily, was included as an outgroup……..54

Figure 2.4 - Microarray analysis of pathogen treatment on cacao PR gene expression. Scatterplots of normalized expression value for all probes for PR genes, comparing A) *P. palmivora* treatment and water-treated control and B) *C. theobromicola* with water-treated control. C) Heatmap showing fold change in transcript abundance after pathogen treatments compared to water-treated control for all 359 Criollo PR genes. Black bars correspond to genes with non-significant (Benjamini-Hochberg $p > 0.05$) fold change or genes removed from analysis in background filtration……56

Figure 3.1 - Inoculation of Salicylic Acid (SA) pre-treated stage C leaves from ICS1 and SCA6 with *Phytophthora tropicalis*. Stage C leaves were inoculated with agar plugs containing *P. tropicalis* mycelium 24 hours after water or 1 mM SA treatment. Representative images of (A) water-treated ICS1 leaves, (B) SA-treated ICS1 leaves, (C) water-treated SCA6 leaves, and (D) SA-treated SCA6 leaves three days after inoculation. Scale bars represent 1 cm. E. Average lesion areas in replicate leaves were evaluated by ImageJ. Data represent means ± SE of treated leaves from 24 replicates per genotype. Letters above the bar chart show the significant differences ($p<0.05$) determined by Fisher's PLSD analysis. F. Relative pathogen biomass was measured by qPCR with DNA isolated 48 hrs after inoculation and is expressed as the ratio of *P. tropicalis actin* to cacao *actin*. Bars represent means ± SE of four biological replicates, each with three technical replicates. Letters above the bar show the significant differences ($p<0.05$) determined by Fisher's PLSD analysis……..82

Figure 3.2 - Functional analysis of *TcNPR1*. **A.** Representative images of lesions from control and *TcNPR1* transiently transformed leaves two days after *Phytophthora tropicalis* inoculation. **B.** qRT-PCR analysis of *TcNPR1* transcript two days after vacuum infiltration (Control –Ctrl; *TcNPR1* expression – NPR1). Data represent means ± SE of three biological replicates. **C.** Average lesion areas from control and *TcNPR1* overexpressing leaves were measured three days after inoculation using ImageJ. Bar charts represent the means ± SE of measurements from 12 lesion spots from four leaf discs of each genotype. The asterisk denotes a significant difference determined by single factor ANOVA ($p<0.05$). **D.** Pathogen biomass was measured at the lesion

sites by qPCR to determine the ratio of pathogen DNA to cacao DNA two days after inoculation. Bar charts represent four biological replicates, each with three technical replicates. The asterisk denotes a significant difference determined by single factor ANOVA analysis ($p<0.05$)……84

Figure 3.3 - Gene induction differences in Sca6 and ICS1. X-axis represents log_2 expression change in SA-treated versus water-treated Sca6 leaves, after obtaining the general linear model mean of differences across leaf stages. Y-axis represents log_2 expression change in SA-treated versus water-treated ICS1 leaves, after obtaining the general linear model mean of differences across leaf stages. Points represent the 234 genes with statistically significant (BH $p<0.05$) expression changes in both genotypes……85

Figure 3.4 - Number of PR genes induced by SA treatment for all leaf stages in Sca6 alone, in ICS1 alone, and across both genotypes. Y-axis shows percentage of PR genes included on the microarray that were induced by SA. Light grey bars represent genes induced with BH $p<0.05$, dark grey bars represent genes induced with BH $p<0.10$……..86

Figure 3.5 - Graphical representation of GO Enrichment by Parametric Analysis of Gene Set Enrichment (PAGE). **A.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes in Sca6. **B.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes in ICS1. **C.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes averaging both genotypes across all stages. **D.** Z scores for select GO terms with statistically significant enrichment in both genotypes. See Supplemental Tables 5-9 for detailed data…….89

Figure 3.6 - Graphs comparing transcript levels of select genes from ICS1 and Sca6 genotypes and Salicylic Acid (SA) and water control treatments as detected by qRT-PCR. Each bar represents the mean of nine samples ± SE (three replicated from each developmental stage). Values are calculated relative to *TcActin*. **A.** Graph displaying effect of SA treatment on cacao PR genes. **B.** Graph displaying effect of SA treatment on cacao genes with Arabidopsis best hits encoded in chloroplasts and mitochondria. Primer sequences are listed in Supplemental Table 10……91

Figure 3.7 - Representative images showing NBT and DAB staining of cacao leaf discs 24 hours after Salicylic Acid (SA) or water (H_2O) treatment of leaves attached to trees: (**A**) NBT-stained ICS1 treated with water, (**B**) NBT-stained Sca6 treated with water (**C**) NBT-stained ICS1 treated with SA, (**D**) NBT-stained Sca6 treated with SA, (**E**) DAB-stained ICS1 treated with water, (**F**) DAB-stained Sca6 treated with water, (**G**) DAB-stained ICS1 treated with SA, (**H**) DAB-stained Sca6 treated with SA. Scale bars represent 1 cm. **I.** Graph displaying mean product area of leaf disc stained by NBT and mean grey value of stained area for each genotype and treatment. **J.** Graph displaying the mean product of area stained by DAB and mean grey value of stained area for each genotype and treatment. In both graphs, differences between bars marked with the same

letter are not statistically significant (Fisher's PLSD analysis p > 0.05). Standard errors in both graphs were calculated from five biological replicates. Each biological replicate is a plate containing three leaf discs……92

Figure 4.1 - Haplotype analysis of CRSP38 gene. A) Maximum-likelihood phylogenetic tree of detected CRSP38 haplotypes. Node labels represent bootstrap support from 100 replicates. Branch lengths represent genetic distance in substitutions per site. B) Bar graph displaying frequency of each haplotype within each population……112

Figure 4.2 - Haplotype analysis of PGIP gene. A) Maximum-likelihood phylogenetic tree of detected PGIP haplotypes. Node labels represent bootstrap support from 100 replicates. Branch lengths represent genetic distance in substitutions per site. B) Bar graph displaying frequency of each haplotype within each population……114

Figure 4.3 - Positioning of variable sites within predicted protein domains for A) PR-1, B) CRSP38, and C) PGIP. Sequence depicted is the amino acid sequence, with color coding representing polarity. Domain IDs for PR-1 and CRSP-38 are shown in annotation arrows. Domain database IDs for PGIP are as follows: LRRNT 2 (pfam 08263), LRR 4 (pfam 12799), LRR RI (cd00116), LRR 8 (pfam 13855)……116

Figure 5.1 - Workflow diagram for transient transformation of cacao leaf tissue……126

Figure 5.2 - Leaf stages and force to puncture measurements. **A)** Photograph displaying representative leaves of stages A (leftmost) to E (rightmost) collected from genotype Scavina 6. Scale bar represents 5 cm. **B-F)** Representative photographs of EGFP fluorescence taken 48 hours after infiltration of leaves (stages A-E) with *Agrobacterium.* Scale bars represent 1 mm. **G)** Measurement of force to puncture for each leaf stage. Bars represent mean of five measurements, each representing one leaf from that stage. Error bars represent standard deviation across five replicates. T-test p values are shown above bars for Stage D and Stage E, which are comparisons of measurements of Stage C leaves with those of the older stages. Differences between Stage A and C and B and C were not significant……128

Figure 5.3 - Time course of EGFP fluorescence intensity after infiltration of leaf tissue with *Agrobacterium.* Fluorescence is expressed as a percentage of the intensity measured at hour 45, the peak time point. Error bars represent standard deviation calculated from three biological replicates……129

Figure 5.4 - Transformation of eight cacao genotypes. **A)** Photograph showing stage C leaves selected from eight cacao genotypes. Some genotype identifiers are abbreviated: Sca6 = Scavina 6, Criollo = B97-61/B2, ICS1 = Imperial College Selection 1. Scale bar represents 5 mm. **B-H)** Representative images of EGFP coverage 48 hours after *Agrobacterium* infiltration using the eight genotypes shown in panel A. Scale bars represent 1 mm. **B)** Sca6; **C)** CCN51; **D)** CF2; **E)**

Criollo; **F)** ICS1; **G)** GU255**; H)** PA107**; I)** TSH1188. **J)** Bar graph depicting transformation efficiency expressed as a percentage of that calculated for Scavina 6 samples. Error bars represent standard deviation calculated from three biological replicates. Bars labeled with the same letter are not statistically significant ($p > 0.05$)……130

Figure 5.5 – Functional analysis of *TcChi1*. **A.** Representative images of lesions from control (Ctrl, transformed with pGH00.0126) and leaves transiently transformed to overexpress *TcChi1* two days after *Phytophthora tropicalis* inoculation. **B.** Average lesion areas from control and *TcChi1* overexpressing leaves were measured three days after inoculation using ImageJ. Bar charts represent the means ± SE of measurements from 12 lesion spots from four leaf discs of each genotype. **C.** Pathogen biomass was measured at the lesion sites by qPCR to determine the ratio of pathogen DNA to cacao DNA two days after inoculation. Bar charts represent four biological replicates, each with three technical replicates. **D.** qRT-PCR analysis of *TcChi1* transcript two days after vacuum infiltration. Data represent means ± SE of three biological replicates. The asterisk denotes a significant difference determined by single factor ANOVA ($p<0.05$)……132

Figure 5.6 - Images representing stages in leaf transformation process. **A)** Stage C cacao leaf with tip and base removed. Scale bar represents 1 cm. **B)** Two halves of a cacao leaf placed into petri dishes with wet paper towel and filter paper. Scale bar represents 1 cm. **C)** Ideal EGFP coverage seen 48 hours after vacuum infiltration of leaves with *Agrobacterium*. Scale bar represents 1 mm……..138

Figure 5.7 - Images representing leaf infection process. **A)** Mature (approximately 1 week since inoculation) plate of the cacao pathogen *Phytophthora palmivora*. **B)** Plate of *P. palmivora* with four agar plugs bored into V8 media. **C)** Inoculation of a new plate of 20% V8 media by transferring agar plugs, pathogen side down, onto the media. **D)** Typical size of pathogen growth 48 hours after inoculation of new plate. Agar plugs are bored around the edges of the cultures to be used for leaf inoculation. **E)** Inoculation of a Stage C cacao leaf with pathogen. Control (media only) plugs are placed on the left side, plugs containing pathogen are placed on the right. **F)** Lesion development 48 hours after inoculation. All scale bars represent 1 cm……139

Figure A.1 - Schematic representation of gene prioritization strategy. Genes that pass more filters have more evidence supporting the importance of their role in defense, and are considered higher priority candidates for functional analysis……156

Figure A.2 - Relative expression of target genes to two housekeeping genes 48 hours after Agrobacterium infiltration. Panels represent the target genes: A) PR-1, B) CRSP38, C) Myb251, D) WRKY50, E) PGI, F) GID1L3. Error bars represent standard deviation from A) 12 replicates, B) 19 replicates, C) 7 replicates, D) 6 replicates, E) 6 replicates, F) 12 replicates……162

Figure A.3 - Lesion size assessment of target gene overexpression 72 hours after inoculation with *Phytophthora palmivora.* Rows represent experiments using different genes: Row 1) PR-1, Row2) CRSP38, Row 3) Myb251, Row 4) WRKY50, Row 5) PGI, Row 6) GID1L3, Row 7) NPR1. Left column panels are representative photos from vector control transformed tissue, center panels are representative photos from target gene overexpression. Right panels are bar graphs representing average lesion size, with error bars representing standard deviation from: PR1: 13 replicates, CRSP: 21 replicates, MYB251: 7 replicates, WRKY50: 6 replicates, PGI: 7 replicates, GID1L3: 14, NPR1, 8 replicates. Asterisk indicates t-test $p < 0.05$…….163

Fig. A.4 – qRT-PCR analysis of leaf detachment effect. **A)** Measurement of PR-1 and **B)** measurement of CRSP38. Bars represent means and error bars are standard deviation from 5 replicates. Asterisk denotes t-test $p < 0.05$ comparing the labelled bar to the Detached 72 Hr bar for the same genotype……164

LIST OF TABLES

Table 1.1 - Summary of phytohormones, their regulation, and their role in resistance……..14

Table 2.1 - Summary of PR gene families and their functions……..42
Table 2.2 – Summary of PR gene families in the *Theobroma cacao* Criollo genome……..49

Table 2.3 - Regulation of Criollo PR genes as detected by microarray. Counts of up- and down-regulated genes represent the number of genes with Benjamini-Hochberg p < 0.05……..57

Table 2.4 - Validation of PR gene induction by qRT-PCR. N.S. indicates p-value was not significant. Genes shown as induced by microarray had BH p-values < 0.05. Inductions detected by qRT-PCR were calculated using REST software (Pfaffl et al., 2002) and represent the average of five pathogen-treated samples compared to five water-treated samples relative to TcTub1 (Tc06_g000360). Transcripts were considered undetected if the average Ct value across all treatments was greater than 35……..59

Table 3.1 - Number of genes up- and down-regulated (BH p value < 0.05) by SA treatment for ICS1 and Sca6 genotypes……..85
Table 4.1 - Genotype counts and frequencies for the variable site detected in PR-1 (Tc02_g002410)……107
Table 4.2 - Genotype counts and frequencies for variables sites detected in CRSP38 (Tc06_g009580)……..108
Table 4.3 - Genotype counts and frequencies for variable sites detected in PGIP (Tc05_g018290)……109
Table 4.4 - Nucleotide diversity and haplotype analysis for three genes and three populations……110
Table 4.5 - Haplotype coding variation forCRSP38……113
Table 4.6 - Haplotype coding variation for PGIP……..115
Table 5.1 – Media recipes for *Agrobacterium* growth and induction and pathogen growth……133
Table A.1 - Genes selected for functional characterization and traits used for their prioritization……161

PREFACE

Chapter 1, the literature review, was prepared by Andrew Fister, with revisions suggested by Dr. Mark Guiltinan and Dr. Siela Maximova.

For Chapter 2**,** Andrew Fister wrote Python scripts, curated gene families, analyzed microarray data, generated figures, performed qRT-PCR, and drafted the manuscript. Dr. Yufan Zhang performed initial BLASTp search on the Criollo genome. Dr. Luis Mejia and Dr. Edward Allen Herre designed and conducted the pathogen infection assays and related microarray experiments. Dr. Siela Maximova and Dr. Mark Guiltinan participated in designing the experiments and editing the manuscript. Dr. Mark Guiltinan was responsible for overall project management. All authors contributed to the writing and review of the manuscript, which was published in BMC Genomics

For Chapter 3, Andrew Fister composed the manuscript and performed the analysis of microarray gene expression data, qRT-PCR, PAGE analysis, and ROS staining. Dr. Shawn O'Neil performed statistical analysis of microarray data. Dr. Zi Shi performed pathogen assays, transient overexpression of NPR1, and contributed to manuscript writing. Dr. Yufan Zhang assisted in initial microarray data analysis and hypothesis formulation. Dr. Brett Tyler contributed to experimental design, statistical analysis of microarray data, and manuscript editing. Dr. Mark Guiltinan contributed to the experimental design, microarray design, and writing and editing the manuscript. Dr. Siela Maximova contributed to experimental design, microarray design, SA treatment, sample collection and RNA extractions, writing and editing of the manuscript and overall project management. The manuscript was published in The Journal of Experimental Botany.

For Chapter 4, Andrew Fister performed PCR amplification, PCR cleaning, sample preparation for sequencing, sequence analysis, and drafted the manuscript. Lin Zhou performed DNA extractions and assisted with PCR amplification. Dr. Dapeng Zhang conceived of the study and planned analyses. Dr. Siela Maximova and Dr. Mark Guiltinan planned experiments and oversaw the project's execution. All authors contributed to writing and editing the chapter, which is in preparation for publication

For Chapter 5, Andrew Fister optimized leaf selection parameters and infiltration conditions and drafted the manuscript. Dr. Zi Shi developed the experimental procedures and drafted a working protocol. Dr. Yufan Zhang and Dr. Emily Helliwell contributed to optimizing and editing the protocol. Dr. Siela Maximova and Dr. Mark Guiltinan oversaw experiments, contributed to optimization, and edited the manuscript.

For Chapter 6, Andrew Fister contributed to the development of gene prioritization strategy, performed molecular cloning of genes of interest, and performed the transient transformation assay. Dr. Desire Pokou and Dr. James Marden contributed to the development of the gene prioritization strategy. Dr. Siela Maximova and Dr. Mark Guiltinan oversaw the experiment, contributed to prioritization strategy development, and edited the chapter.

Chapter 7 was written by Andrew Fister, with revisions suggested by Dr. Siela Maximova and Dr. Mark Guiltinan.

FOREWARD AND ACKNOWLEDGEMENTS

As an undergraduate, I double majored in Biology, which satisfied my deep curiosity about the nature of life, and English Literature, which I think of as an approach to sociological philosophy. I love fiction, and even more, I love to think about why authors write what they do: what universal truths were they trying to capture, why did the ideas matter to them, why did the ideas matter to their contemporaries or within their time, etc. Even as early as high school I developed a special fondness for a particular poem, "Ozymandias," and over the years I've considered it in a variety of contexts. Here's the text:

I met a traveller from an antique land
Who said: "Two vast and trunkless legs of stone
Stand in the desert. Near them, on the sand,
Half sunk, a shattered visage lies, whose frown,
And wrinkled lip, and sneer of cold command,
Tell that its sculptor well those passions read
Which yet survive, stamped on these lifeless things,
The hand that mocked them and the heart that fed:
And on the pedestal these words appear:
'My name is Ozymandias, king of kings:
Look on my works, ye Mighty, and despair!'
Nothing beside remains. Round the decay
Of that colossal wreck, boundless and bare
The lone and level sands stretch far away."

- Percy Bysshe Shelley, 1817

As I near the completion of my dissertation, I've been considering the poem in the context of one's place in the world of science. The poem's central themes are the titular ruler's monomaniacal pride and the naiveté of his thinking that his achievements will transcend time. I'm proud of my work, but I think that's the extent of my similarity to Ozymandias. I'm proud of all I've learned, I'm proud to have published my findings, and I hope that my work has an impact. But in contrast, I think it is fun to remember that a scientist's body of work is a product of their time; he or she is limited by technology and resources and other ideas in that zeitgeist. So, 400 years ago a compound microscope was amazing, 40 years ago molecular cloning was the peak of ingenuity, now genome and transcriptome sequencing provide immense and unanticipated quantities of data, and who knows what will become possible over the next decades. But what is possible today doesn't invalidate the effort and innovation of the past. It's just a profound and wonderful thing to be a part of a legacy of people who are curious and who care enough to try to answer life's questions. So, I take a more forgiving view of Ozymandias

than did Shelley's traveller. To build, and to contribute, and to be a part of the community, is a beautiful thing.

So with that, I'd like to thank the community that who taught me to build and contribute.

To my advisors, Mark and Siela: Thank you for taking me into the lab. I'm embarrassed to think back to when we met because I had no training in plant biology or molecular biology. It's hard to believe I would have made a good impression. Nonetheless, you welcomed me into the group and fostered creativity and 'working smart.' I also have you to thank for the gift of clarity of thought. The hours in lab meeting rehashing the 'big questions,' reconsidering how to efficiently plan experiments, and laying out the perfect slides were time well spent.

To my committee: I'm grateful to my doctoral committee members, Jim Marden, Charlie Anderson, and Majid Foolad, and Mark and Siela again, as well as my comprehensive and candidacy exam committee members, Ross Hardison, Tim McNellis, Seogchan Kang, for your time, consideration, and insight over the years. I appreciate the feedback I received within the exams and on early drafts of manuscripts and proposals. I admire you all as scientists, and my work is stronger thanks to your feedback. Just as importantly, I've enjoyed getting to know you on a personal level. It was truly critical to my growth in academia to see that non-scientific interests and hobbies, from sailing, to photography, to triathlons, balance scientific rigor and provide a way to 'recharge the batteries.' Thank you.

To the greatest lab manager, Lena: Thanks for keeping the ship afloat. I'm sorry for complaining about being nagged. Graduate students are, at times, lazy and petulant, but we understand that your role is integral to our being successful. I don't remember ever needing to wait for reagents, and we could always rely on you to put our own needs before the stack of items on your to-do list. Your tireless help is most genuinely appreciated.

To my lab friends: First, thanks to Stone and Yufan for showing me the ropes. I remember watching you both carrying out all sorts of experiments when I joined the lab, and thinking to myself "How am I ever going to learn all of this?" I did, thanks to you. Thanks also to the others who've popped in along the way, Emily, Xiangling, Mariela, Monica, and Adriana, for taking the time to share your passion and ideas with me. New guys, Kevin and Ben, I've been enjoying getting to know you both, and I'm thrilled that I'll be sticking around to continue to work with you. Thanks also to a slew of undergrads I've had the pleasure of working with: Adam, Julia, Dan, Dominic, David, Cody, Brian, Grace, and Joe. Whether you were pruning trees, mixing fertilizer, making media, or measuring leaf lesions, you were a huge help.

To our collaborators: I can't believe how fortunate I've been to interact with so many great scientists with such diverse interests. Claude dePamphilis, Brett Tyler, Brian Staskawicz, Shawn O'Neil, Yinong Yang, Desire Pokou, Abu Dadzie, Dapeng Zhang, Luis Mejia, and Allen Herre, your time, patience, and input is tremendously appreciated. It has been a sincere pleasure to work with you, and I hope our paths continue to cross. Thanks also to Andy Clark. However little I knew starting graduate school, I knew far, far less as an undergraduate, and I

credit you and your lab members with beginning my acclimation to the world of science. Spending two years as a part of your group was a revelatory and inspiring experience.

To my Genetics pals, Elyse, Liron, Eric, Erin, Arslan, Bruce: The program lasted just long enough for us all to meet! We've been a great support system to each other. I'm thankful for your friendship and your willingness to sit through practice talks. To my other pals in the trenches, Thomas, Will, Nate, Matt, thanks for the support, best of luck to you all.

To my childhood friends, Mark, Nic, Austin, Ryan, Matt: Thanks for being there. It's comforting that I can count on some links to imgur nonsense a few times a week for a quick laugh. No, the transformed cacao trees aren't my army of super-soldiers. It's scary how poorly you understand genetics. Thanks for keeping me level.

To my brother, Matt: If I didn't have your model to follow, I certainly wouldn't have pushed myself as hard as I have. I remember you trying to teach me to code on weekends when I was in middle school because YOU THOUGHT IT WOULD BE FUN, FOR BOTH OF US. Your intelligence, passion, and weirdness are delightful. Thanks for help with Python, and thanks for being you.

To Lauren: You're perfect. Unconditional support is a trite thing to be thankful for, but it's exactly what you've given. Thanks for teaching me about mixed effect models, thanks for the motivation to keep working long after I've stopped wanting to, thanks for having dinner ready when I come home late. I can't thank you enough for letting me lean on you. You're brilliant and infinitely capable. I couldn't have finished this without you, and it definitely wouldn't have been as much fun. Thanks for the inspiration. I'll resume my wedding planning responsibilities soon. Love you.

To my parents, Joe and JoAnne, I dedicate this dissertation. You are infinitely giving. The work that I've done is thanks to dedication, which I learned from you. The persistence that you instilled in me is my favorite thing about myself. I can't thank you enough. Enjoy retirement, I'll see you on the bay.

Now:

Look on my works, ye Mighty, and despair!
(Just kidding, thanks for reading, hope you like it.)

Chapter 1 :
Literature Review

1.1. *Theobroma cacao*: Cultivation and biology

Production and horticulture

Theobroma cacao is the source of cocoa, the major raw ingredient used to make chocolate, and thus is a crop of great economic importance around the globe. Cacao's commercial value – and popularity – derive from the seeds, which are sources of cocoa butter and cocoa powder. In this dissertation, 'cacao' refers to the plant and cocoa refers to the dried and fermented seeds and all products derived from them. More than five million metric tons of cocoa are produced every year, creating a chocolate trade valued at around $80 billion (Ploetz, 2016). United States imports alone total more than $1.3 billion annually. This trade fuels a chocolate industry that leads to the consumption of more than 7 million tons of chocolate annually. Economic growth in Asia is leading to increased demand, particularly in China, requiring new innovations to improve yield and losses to pathogens.

The *T. cacao* tree itself is native to the Amazon Basin; it is known to have been propagated by indigenous peoples throughout Mexico and Central America before the settlement of the New World by Europeans (Bergmann, 1969; Holliday, 1971), and it is now grown in tropical regions around the world. Until the mid-1900's, the majority of the cacao production came from Central and South America; however, by the 1960's, production in West and Central Africa expanded to produce more than 70% of global production (Duguma et al.; Wood and Lass, 2008).

By far the majority of cocoa is produced on small farms (Rice and Greenberg, 2000). In Africa, smallholder farmers usually manage farms of less than five hectares. In contrast, Brazil has many large (>1000 hectare) plantations, but smaller farms are more common making the

Brazilian average about 28 hectares. Larger farms are common in Malaysia, where plantations of more than 40 hectares account for nearly 90% of production (Rice and Greenberg, 2000). This has been possible in Malaysia and a few other countries because of investment in plantations by governments, as increasing the size of farms often comes with production costs that are untenable for many farmers (Laird et al., 1996). On the contrary, small scale farming has been shown to have lower production cost per unit output and higher yield per unit area, as farmers are able to manage more intimately their land and thereby maximize pod production (Laird et al., 1996; Rice and Greenberg, 2000).

Physiology and taxonomy

Cacao is a semi-deciduous tree, which grows to be 5-10 meters tall. New vegetative growth appears in bursts called flushes, with leaves reaching mature lengths up to 40 cm. Flowers develop from floral cushions along the truck and branches, and are naturally pollinated by midges (Glendinning, 1972). These pollinators play a vital role in cacao reproduction as many varieties exhibit self-incompatibility (Knight and Rogers, 1955). Young pods develop weeks after pollination, and their development continues for 5-8 months until they reach maturity, at which point they are generally oblong, vary in size and color, and on average weigh 400 g. (Glendinning, 1972). Pods contain 20-50 seeds, which also range in color from off-white to dark purple or brown, and which have a dried mass of about 40 g. Within the seeds, cotyledons can be white or purple, depending on the genotype.

Theobroma cacao is one of 22 species in the *Theobroma* genus of the *Malvaceae* family, all of which are native to the American tropics (Cuatrecasas, 1964). A hybridization barrier has prevented interspecific crosses from being used to introduce new genetic material into cacao breeding programs (Zhang et al., 2011), but biotechnological advances may offer a means of overcoming the barrier (Silva et al., 2004). The majority of research on the genus, other than on cacao, has focused on *Theobroma grandiflorum (Alves et al., 2007; Zhang et al., 2011),* also known as cupuaçu, which is also cultivated in Amazonian jungles and is used to make liquid beverages from the sweet pulp found in its fruit.

Genetics and genomics

Until recently, cacao germplasm was described as belonging to one of several groups. These included the Criollo group and the Forastero group (Cheeseman, 1944), believed to have evolved in Central and South America (Cuatrecasas, 1964), respectively, and the Trinitario group, which were hybrids of the former two (Cheeseman, 1944). Subgroups within the Forastero varieties were studied, the most important of which was the Amelonado group, from which the majority of germplasm was taken to establish the crop in Africa and Asia (Wood, 1991). The Ecuadorian Nacional type was also considered a distinct variety (Lerceteau et al., 1997; Motamayor et al., 2003). A more recent analysis of >1200 genotypes using >100 microsatellite markers identified 10 genetic groups: Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañon, Nacional, Nanay, and Purus (Motamayor et al., 2008). The center of origin for all cacao is now believed to be the Amazon Basin, and the various genetic groups are believed to have diverged because of now eroded ridges in the Amazon region (Motamayor et al., 2008).

The genome of two cacao genotypes, a Criollo genotype (Argout et al., 2011) and an Amelonado type, the widely-produced Matina 1-6 (Motamayor et al., 2013), were determined by whole genome sequencing approaches, enabling a nearly full description of its content and structure. The tree is diploid and has ten chromosomes made up of ~445 million bases. Both genome size and predicted gene number are low to middling among dicots. Annotation of the two genome sequences predicts ~29,000 genes, accounting for ~100 Mb of sequence. The percentage of the Criollo genome (~35%) made up of transposable elements was significantly lower than that of the Matina genome (42%), with Criollo being toward the low end of the spectrum for sequenced plant genomes and Matina having a more typical value. Comparison of the genomes revealed large syntenic regions, as expected comparing two individuals of the same species; but differences were detected, such as 12 orthologous regions resembling chromosomal translocations (Motamayor et al., 2013). Compared to other dicots, cacao has average family sizes for a variety of defense genes, including receptor-like kinases, nucleotide-binding leucine-rich repeats (Argout et al., 2011) and the PR families (discussed below and in Chapter 3).

1.2. Pathogens of cacao

Populations of cacao around the world are host to a variety of pathogens, which severely limit the productivity of plantations and smallholder farms. Losses reach 30-40% of total yield, annually (Wood and Lass, 2008). The major diseases of cacao were recently reviewed (Bailey and Meinhardt, 2016), and are described in brief below.

Phytophthora spp. – Black pod rot

Phytophthora is a genus of oomycete stramenopiles, best known for the family member *Phytophthora infestans*, the causal agent of the Irish potato famine (Kroon et al., 2011). Several species within the genus are pathogens of cacao, and annual losses to black pod rot lead to a more than 30% reduction of pod yield and death of 10% of trees, globally (Guest, 2007; Hebbar, 2007). Symptoms of infection are formation of dark external lesions on pods as internal tissues, including seeds, shrivel. The disease can also affect stems and leaves of young plants, causing seedling blight, which can be devastating to cacao nurseries (Hebbar, 2007; Acebo-Guerrero et al., 2012). Resistance to *Phytophthora* is uncommon in commercial cacao varieties, but QTL for resistance have been identified using several genotypes (Risterucci et al., 2003; Guest, 2007; Lanaud et al., 2009).

The majority of black pod rot is caused by four *Phytophthora* species. *P. palmivora* has a global distribution and individually accounts for more than 20% yield loss (Erwin and Ribeiro, 1996; Guest, 2007). *P. megakarya,* which is only found in West and Central Africa, is the most destructive pathogen affecting the small farms in these regions, occasionally destroying all pods on a farm (Opoku et al., 2000; Guest, 2007). Two other common species *P. capsici* and *P. citrophthora* are also geographically limited, but these to South America (Erwin and Ribeiro, 1996). *P. capsici* is a common pathogen of peppers and solanaceous plants, and proliferates in warm wet seasons (Hausbeck and Lamour, 2004), while *P. citrophthora,* which commonly affects citrus trees, grows in cooler conditions (Erwin and Ribeiro, 1996).

***Moniliophthora roreri* – Frosty pod rot**

Frosty pod rot, caused by the basidiomycete fungus *Moniliophthora roreri*, accounts for ~5% of annual cocoa loss (Evans et al., 2003; Phillips-Mora and Wilkinson, 2007). Frosty pod and witches broom are caused by related fungi with poorly understood life cycles, which led to several taxonomic revisions (Aime and Phillips-Mora, 2005). Frosty pod is a recognizable disease, with lesions growing characteristic white mycelia several days after infection, and as this occurs, internal tissues degrade into a watery mass (Phillips-Mora and Wilkinson, 2007). The disease is believed to have originally been pathogenic to only *Theobroma gileri*, and to have expanded from its center of origin in Ecuador or Colombia in the 1950's (Holliday, 1971). There are now confirmed cases as far north as Mexico (Phillips-Mora et al., 2006) and the disease has caused severe losses in Peru (Evans et al., 1998). Its windborne dispersal makes its continued spread a source of concern, particularly if human activity carried spores to cocoa-producing countries in Africa.

***Moniliophthora perniciosa* – Witches' broom**

Witches' broom, caused by another basidiomycete, *Moniliophthora perniciosa*, is also native to the Amazon basin and is believed to be the only specialized cacao pathogen to co-evolve with its host (Grande et al., 1952). It is believed that this co-evolution produced cacao varieties with genetic resistance to witches' broom. Expeditions were carried out to collect and establish lines of resistant germplasm, the most famous example being the Pound collection (Pound, 1943), and these plants are widely used as parents in breeding programs (Purdy and Schmidt, 1996). While the plant pathogen interaction is difficult to study in the lab, previous work in the Guiltinan-Maximova Lab used *Solanum lycopersicum* as a model species to study the mechanisms of infection, and this analysis established similar symptoms of infection in the two plant species (Marelli, 2008).

M. perniciosa spores are spread by wind can penetrate and infect a variety of cacao tissues, but most often affect new growth, particularly shoots (Purdy and Schmidt, 1996). Both

witches' broom and frosty pod rot exhibit hemibiotrophic pathogenesis strategies, with witches' broom generally having a longer parasitic stage (Evans, 2016). Infection triggers loss of apical dominance, after which the plant produces dense vegetative clusters, the characteristic 'brooms' (Griffith et al., 2003). Several weeks after infection, infected shoots necrose and act as a source of new inoculum, causing rapid dispersal and a dramatic, up to 90% loss of productivity in infected plantations (Griffith et al., 2003). Severe infections can lead to death of the tree. A major outbreak of witches' broom in the Bahia region of Brazil occurred in 1989, ultimately causing Brazil to transition from the world's third largest cacao producing nation to a net importer (Meinhardt et al., 2008). An extensive phytosanitation program was carried out to eradicate or prevent the spread of the disease (Pereira et al., 1996), but the fungus remains prevalent throughout South and Central America.

***Oncobasidium theobromae* – Vascular streak dieback (VSD)**

Oncobasidium theobromae, is a relatively recently described basidiomycete disease of cacao (Talbot and Keane, 1971), affecting farms in plantations in Papua New Guinea, Indonesia, Malaysia, and Southeast Asia (Samuels et al., 2012). As cacao was only relatively recently introduced to the region, it is assumed that the pathogen has other native hosts, but they have not been identified (Keane and Prior, 1991). As with the other basidiomycetes, spores are produced in humid conditions which are then dispersed by wind. It tends to infect new growth on mature trees and seedlings plants near mature trees (Guest and Keane, 2007). Symptoms of infection do not develop for 3-5 months, during which time the pathogen spreads outward from the site of infection through the xylem. Symptoms after the switch from parasitic stage to necrotrophic stage include chlorosis and leaf abscission, with necrotic blotches being noted in more recent surveys of infected tissue (Guest and Keane, 2007). Recent work has mapped QTL and identified molecular markers linked to resistance to VSD (Epaina, 2014).

Cacao swollen shoot virus (CSSV)

CSSV is a double stranded DNA virus from the *Badnavirus* genus (Brunt, 1970; Muller et al., 2008), and it is known to be transmitted by mealybugs (Dufour, 1991). The disease was first described in 1940 (Posnette, 1940) after it was identified in farms in Ghana, but it can now be found in Togo, Sierra Leone, Ivory Coast, and Nigeria (Dzahini-Obiatey et al., 2010). Symptoms of infection include chlorosis and defoliation, with wilting of entire trees occurring in severe cases (Dzahini-Obiatey et al., 2010). Extensive measures have been taken to eliminate the spread of CSSV in Ghana, a program that at the time was described as the most costly initiative for prevention of a viral plant epidemic (Thresh and Owusu, 1986). Its spread poses an increased threat to cocoa production in West Africa, but breeding efforts are underway to improve resistance of African varieties (Gutiérrez et al., 2016).

Improving resistance to pathogens

The Pound collection, defined on an expedition from the late 1930's into the 1940's identified wild germplasm with broad spectrum resistance to a variety of cacao's diseases (Pound, 1943). Perhaps the most well-known of these genotypes is Scavina 6. Breeding programs incorporated this germplasm, usually crossing the resistant individuals with other trees with desirable quality traits to produce superior varieties (Lopes et al., 2011). A recent survey of genetic diversity found that the Upper Amazon, through Ecuador, Peru, Colombia, and Brazil, was highly diverse, likely poorly sampled on previous expeditions, and therefore underrepresented in germplasm collections around the world (Thomas et al., 2012; Zhang and Motilal, 2016).

Cacao breeding efforts have led to the development of QTL maps for resistance to some of the most severe pathogens, including *Moniliophthora perniciosa* (Queiroz et al., 2003; Brown et al., 2005; Faleiro et al., 2006), *Phytophthora spp.* (Clement et al., 2003; Risterucci et al., 2003; Brown et al., 2007), and *Moniliophthora roreri* (frosty pod rot) (Brown et al., 2007). A subsequent meta-analysis overlaid these QTL to combine them where multiple experiments

associated a trait with similar genomic loci (Lanaud et al., 2009). Most recently, a QTL mapping experiment for identifying witches' broom resistance associated regions was performed using a higher density SNP-based genetic map. The identified QTL were then searched for putative defense genes, and expression data from transcriptomic experiments were referenced to identify genes responsive to witches' broom which may be conferring resistance (Royaert et al., 2016). In many cases, resistant haplotypes were found to have been derived from Scavina 6. Problematically, the resistant parent (like Scavina 6) in many of these studies was characterized as resistant more than 70 years ago, and recent field evaluation has shown decreased effectiveness of the same germplasm against various diseases. New germplasm collections are taking place, and it is imperative that these individuals be incorporated into breeding programs, both to identify new genetic mechanisms of pathogen tolerance and to increase genetic diversity on farms around the world (Zhang and Motilal, 2016).

Other strategies for preventing losses to cacao pathogens include chemical treatments (Gockowski et al., 2010), biologic application (Ten Hoopen and Krauss, 2016), and phytosanitation (Medeiros et al., 2010). While each can be helpful in reducing losses, labor input, cost, and availability of chemicals often makes them untenable options for smallholder farmers. A recent study showed the treatment of leaves with glycerol induced the defense response and improved *Phytophthora* tolerance (Zhang et al., 2015). While identifying and developing low cost chemical treatments that could reduce losses is a useful and promising strategy, incorporating more genetic sources of resistance into breeding programs would be the most reliable and sustainable means of improving yield and stabilizing the livelihood of cacao farmers globally (Gutiérrez et al., 2016).

1.3. Plant-pathogen interactions and the defense response

Pathogen recognition overview

A plant's ability to recognize pathogens, and to distinguish them from the thousands of other microbes in its environment, is essential to its survival. The current model is based on the

ongoing evolutionary arms-race in which pathogens evolve to avoid and suppress plant immunity, and plants adapt, enabling recognition (Jones and Dangl, 2006; Pieterse and Van Wees, 2015). In the current model, extracellular pattern-recognition receptors (PRRs) evolved in plants and recognize microbial proteins and other extracellular structures called pathogen- or microbe-associated molecular patterns (PAMPs and MAMPs). The molecular events triggered by these interactions are called pattern-triggered immunity (PTI). The molecules on both the plant and pathogen sides of this process are often highly conserved (Dodds and Rathjen, 2010). Successful pathogens, often those that have co-evolved with their host, are by definition capable of suppressing the defense response triggered by PTI using effector proteins which are secreted into the plant cell (Jones and Dangl, 2006; Bigeard et al., 2015). Plants, however, have evolved sets of intracellular proteins, Resistance (R) genes, which recognize pathogen effects and lead to activation of a second wave of defense termed effector triggered immunity (ETI). Pathogen effectors and plant R genes are highly diverse within and between species (Dodds and Rathjen, 2010). This secondary burst of defenses is generally stronger than that triggered by PTI. Mutations can introduce variation that prevents plants from detecting pathogen effects, so diversifying selection can favor existence of multiple forms of R genes within populations, increasing the likelihood of plants recognizing effector variants (Jones and Dangl, 2006).

PAMP triggered immunity – Basal, broad spectrum defense

Broadly, PTI is considered to be a general defense response against non-adapted plant pathogens (Dodds and Rathjen, 2010). While the downstream molecular events triggered by PTI are less dramatic than those triggered by ETI, suppression of PTI has been shown to make plants susceptible to normally nonpathogenic strains of microbes (Li et al., 2005; Zipfel, 2009). Therefore, the induced activities appear to be sufficient for preventing colonization by the majority of microbes interacting with a plant.

Plant PTI has similarities to innate immunity in animals; both kingdoms rely on a set of genetically encoded factors for initial perception of potential invaders (Boller and Felix, 2009; Spoel and Dong, 2012). In fact, animals and plants are capable of detecting some of the same

microbial structures, like Flagellin, albeit through recognition of different parts of those structures (Gómez-Gómez and Boller, 2000; Hayashi et al., 2001). In plants, PRRs are often members of the receptor-like Serine/Threonine kinase family, which has more than 600 members in Arabidopsis, accounting for more than 2% of Arabidopsis' genes (Shiu and Bleecker, 2003). RLKs, like FLS2 and its orthologs, which are known to detect Flagellin (Gómez-Gómez and Boller, 2000), generally have extracellular protein-protein interaction domains, a transmembrane region, and an intracellular structure that acts in signal transduction (Dardick and Ronald, 2006). In the case of FLS2, the extracellular region is a highly conserved leucine-rich repeat, a structure known to allow interaction with a variety of ligands (Kobe and Kajava, 2001). In Arabidopsis, FLS2 was expressed in all tested tissues (leaves, flowers, stems, and roots) and was not induced by Flagellin treatment, suggesting its role and those of other PRRs are a constitutive, basal element in defense (Gómez-Gómez and Boller, 2000). Recognition of the ligand triggers activity of the kinase domain, beginning signal transduction, often through MAPKs (Nühse et al., 2000), and ion flux and oxidative burst are also often detected (Apel and Hirt, 2004; Boller and Felix, 2009).

As a result of PTI being triggered, signal transduction through a MAPK cascade achieves several functions. Production of ethylene, a well-studied plant defense hormone (Xu et al., 1994; Yang et al., 2015), begins within minutes of ligand binding by the PRR (Spanu et al., 1994). WRKY proteins, a family of plant-specific, stress-related transcription factors (Eulgem et al., 2000; Pandey and Somssich, 2009), are induced and trigger further gene induction (Asai et al., 2002). Ultimately, PTI activation led to induction of ~1000 genes in Arabidopsis, including R genes, genes encoding peptides with direct anti-microbial activities, and more RLKs (Zipfel et al., 2004).

Effector proteins suppress basal defenses

Effector proteins have been the subject of much research in bacteria (Alfano and Collmer, 2004; Deslandes and Rivas, 2012), fungi (Stergiopoulos and Wit, 2009; Rafiqi et al., 2012), and oomycetes (Wawra et al., 2012; Petre and Kamoun, 2014). Bacterial pathogens use type three secretions systems to inject their effector proteins into the plant cell (Cornelis and

Gijsegem, 2000). The first 15 amino acids of the effector protein itself usually contains a targeting motif recognized by the type three secretion system, which has expedited bioinformatic prediction of bacterial effectors (Anderson et al., 1999; Cornelis and Gijsegem, 2000). Fungal effectors are secreted to the interface of the pathogen and host through an endoplasmic reticulum-based secretory system (Koeck et al., 2011). However, while conserved motifs have been found in fungal effectors, like poly-cysteine repeats in poplar leaf rust (Hacquard et al., 2011), no definitive targeting signal has been identified (Rafiqi et al., 2012). In contrast, oomycete effectors are known to have short N-terminal targeting sequences of known motifs, including RXLR, LFLAK, and CHXC (Whisson et al., 2007; Jiang et al., 2008). These effectors are believed to enter the plant cell by interacting with phosphatidyl-inositol 3 phosphate lipid rafts, followed by endocytosis of the complex (Kale and Tyler, 2011). Genomes of various pathogens have been found to encode hundreds to over a thousand predicted effector proteins (Jiang et al., 2008; Deslandes and Rivas, 2012; Rafiqi et al., 2012). Effector proteins target a wide variety of proteins related to the PTI process (Macho and Zipfel, 2015): some have been identified that target PRRs directly (Göhre et al., 2008), others that target components necessary for PRR production (Fu et al., 2007), some that target proteins that interact with RLKs (Zhang et al., 2010), and still others that target the downstream MAPK signaling cascade (Zhang et al., 2007).

Effector triggered immunity – Adapted defenses for co-evolved pathogens

ETI, as mentioned above, is triggered by recognition of a pathogen effector protein, which is usually capable of suppressing PTI, by plant R proteins. The intracellular detection of effector proteins is carried out by the nucleotide-binding/leucine-rich repeat (NLR) superfamily of proteins, which begin local and systemic responses (Jones and Dangl, 2006). Both direct and indirect interactions of NLRs and effectors have been described. In the direct model, the LRR domain of the NLR has been shown using in vitro assays to specifically bind to certain effector structures, and specificity of the LRR sequence was shown to be critical for binding (Krasileva et al., 2010; Ravensdale et al., 2012). In the indirect model, effector proteins interact with a plant molecule that is ‘guarded’ by an NLR. The N terminal domain (a Toll/Interleukin Repeat (in dicots

only) or a Coiled-Coil (Jacob et al., 2013)) of the NLR directly binds the guarded plant molecule, and thereby it is able to detect pathogen manipulation (Mackey et al., 2003). Adoption of the indirect sensing strategy has been suggested to be evolutionarily beneficial as it requires the plant to bind only its own proteins, rather than more rapidly evolving pathogen effector proteins (Mukhtar et al., 2011). Cases have also been described where NLRs act in homo- (Bernoux et al., 2011) or hetero- (Sohn et al., 2014) dimers to successfully bind effectors and trigger downstream signaling.

After perception of effectors by NLRs, signal transduction can be carried out through a variety of mechanisms (Cui et al., 2015). There are a variety of cases where NLRs translocate from the cytoplasm to the nucleus to activate downstream targets (Heidrich et al., 2012). These interact directly with transcription factors including WRKY family members and Myb family members, in some cases activating nucleotide binding capacity of transcription factors and in others and de-repressing their inhibitory capacity (Shen et al., 2007; Chang et al., 2013; Padmanabhan et al., 2013). Other NLRs have never been detected inside the nucleus, and the way they trigger downstream signaling is not clear. Some are believed to recruit transcription factors to the plasma membrane (Holt III et al., 2002), and others are believed to interact with MAPKs, triggering signal transduction beginning at the membrane (Li et al., 2012). Most commonly, NLRs interact with one of two well-known signaling components, depending on their structure. Coiled-coil NLRs interact with Non-race-specific Disease Resistance 1 (NDR1), which while membrane bound interacts with other transcription factors to trigger signaling (Kim et al., 2005). In contract, Toll/Interleukin Repeat NLRs often interact with Enhanced Disease Susceptibility 1 (EDS1) which can move between the nucleus and cytoplasm, and forms signaling complexes with other proteins (Feys et al., 2005; Rietz et al., 2011). However, there are exceptions, as EDS1 has also been shown to interact with some coiled-coil NLRs (Venugopal et al., 2009). Ultimately the various signaling mechanisms trigger local and systemic reactions to combat infection.

Hypersensitive response

At the site of infection, signaling cascades lead to an upregulation of defense proteins and a reactive oxygen species (ROS) burst that leads to cell death, preventing spread of the pathogen. Types of ROS include superoxide, hydrogen peroxide, and hydroxide molecules (Torres, 2010). What was first described as 'hypersensitiveness' (Stakman, 1915) is now called the Hypersensitive Response (HR), and is defined as resistance-associated cell death appearing at the site of infection (Coll et al., 2011). ROS production by chloroplasts and mitochondria (Van Aken and Van Breusegem, 2015) and NADPH oxidases (Marino et al., 2012) plays a vital role in establishing HR. Forward genetic screens identified Lesion Simulating Disease Resistance 1, a negative regulator of cell death (Jabs et al., 1996), which has several interacting partners believed to be environmental sensors (Li et al., 2013). Interactions with light-sensing, hypoxia-sensing, and cold-sensing machinery make LSD1 a major regulator of cellular ROS response.

Early models of gene-for-gene resistance suggested that HR was a conserved response after recognition of an effector by an R gene (Jia et al., 2000). While induction of hypersensitive response (HR) is often taken as the indicator of ETI, there are a number of mechanisms by which the two processes can become decoupled (Gassmann, 2005; Coll et al., 2010). This may be because induction of cell death requires a greater accumulation of signaling output (Cui et al., 2015).

Phytohormones and the systemic response

The signaling cascade in ETI involves activation of plant hormone signaling pathways, which propagate the signal systemically. Table 1.1 summarizes some of the components and functions of five defense-related hormone signaling pathways.

Table 1.1. Summary of phytohormones, their regulation, and their role in resistance.

Phytohormone	Synthesis and Structure	Receptor(s) and Interactors	Defense-Related Targets	Homeostasis and Regulation	References
Salicylic Acid (SA)	Phenolic. Synthesized from shikimate intermediate, chorismate. Produced in chloroplasts.	NPR3 and NPR4 bind SA, activate NPR1, a master transcriptional co-regulator, which interacts with TGA TFs.	PR genes, WRKY TFs, MAPK signaling, ROS production. Inhibits JA signaling.	NPR1 stability regulated by SA-mediated interactions with NPR3 and NPR4 to modulate defense.	(Vlot et al., 2009; Fu et al., 2012; Pieterse et al., 2012; Fu and Dong, 2013)
Jasmonic Acid (JA)	Lipid-derived. Generated through oxylipin pathway after release of membrane α-linolenic acid.	Bound by COI1, as part of an SCF E3 ligase complex, binds JA, and de-represses TFs by inactivating JAZ.	PDF family, PR genes associated with necrotroph defense, toxins and anti-nutritive compounds active against herbivores. Inhibits SA signaling.	Transcription factor feedback loop regulates JAZ expression, down-regulating the pathway.	(Memelink, 2009; Pieterse et al., 2012; Song et al., 2014; Yang et al., 2015)
Ethylene (ET)	Hydrocarbon. Synthesized from methionine through Yang cycle.	Bound by several membrane-bound receptors that have His kinase activity. Trigger activation of EIN2 and EIN3, activating TFs.	Coordinates with JA signaling, activating wound and necrotroph defenses. Also regulates ROS production through PTI feedback.	EIN3 stability regulated by proteasome, feedback loops inactivate ET and JA production. Also the ripening hormone.	(Guo and Ecker, 2003, 2004; Bari and Jones, 2009; McManus, 2012; Zipfel, 2013)
Brassinoster-oids (BRs)	Polyhydroxylated diterpenoids. Synthesized by terpenoid pathway in plastids.	Bound by the RLK BRI1, de-repressing signaling pathway and activating TFs BES1 and BZR1.	Negative feedback on PTI. Increase ROS and antioxidant production, activate WRKY TFs. Enhance SA signals, inhibit JA signals.	Concentrations regulated through feedback regulation of BR and sterol synthesis and degradation.	(Zullo and Adam, 2002; Tanaka et al., 2005; Robert-Seilaniantz et al., 2011; De Bruyne et al., 2014)
Abscisic Acid (ABA)	Isoprenoid, 15-C weak acid. Synthesized through the MEP pathway. Originally thought to be leaf specific, now known to be	Bound by soluble PYR/PYL proteins, de-represses SNF1-related kinases. Signal transduction leads to activation of	Primarily controls leaf abscission. Negative transcriptional regulation of SA, JA, and ROS production. Believed to control shifts	Catabolized to phaseic acids when concentration is too high. Negative feedback from SA and JA pathways	(Anderson et al., 2004; Asselbergh et al., 2008; Fan et al., 2009; Robert-Seilaniantz et al., 2011; Finkelstein, 2013)

	produced in all tissues.	ABI3, ABI4, ABI5 TFs.	between SA and JA pathway activation.		
Gibberellic Acid (GA)	Tetracyclic diterpenoids. Synthesized by terpenoid pathway in plastids.	Bound by GID1, which degrades DELLAs, negative regulators of growth.	Primarily involved in growth promotion. But, DELLA proteins interact with SA and JA pathways, and activate ROS detoxification pathways.	Enzymatic control of bioactive GAs. Feedback inhibition of GA synthesis.	(Tanaka et al., 2006; Yang et al., 2008; Bari and Jones, 2009; Robert-Seilaniantz et al., 2011; De Bruyne et al., 2014)
Cytokinins (CKs)	Adenine derivatives and phenylurea compounds.	Bound by AHK2-4, triggers transduction cascade activating ARR which interacts with TGA TFs.	Have early and late responses, initially enhancing then suppressing SA pathway. Also differentially synergize and antagonize auxin pathway, affecting growth. Can suppress PTI and ETI.	Enzymatic control of bioactive CKs. Feedback inhibition of CK synthesis.	(Bari and Jones, 2009; Frébort et al., 2011; Robert-Seilaniantz et al., 2011; Naseem et al., 2015)

The two most studied defense hormones are salicylic acid (SA) and jasmonic acid (JA). SA is considered the master regulatory hormone of systemic acquired resistance and defense against biotrophs and hemibiotrophs (Vlot et al., 2009), whereas jasmonic acid (in coordination with ethylene) regulates defense against necrotrophic pathogens and herbivores (Browse, 2009). Signal transduction of each of the two hormones' pathways are known to have antagonistic action on the other (Beckers and Spoel, 2006; Yang et al., 2015). One trend between regulation of homeostasis in SA, JA, and ET pathways has been linked to SCF E3 Ubiquitin Ligase-mediated degradation of members of hormone receptor complexes (Guo and Ecker, 2003; Fu et al., 2012), and consideration of JA and ET receptor models motivated discovery of the SA receptors, NPR3 and NPR4. Several other hormones play roles in defense, modulating action of SA and JA pathways and participating in feedback regulation of PTI and ETI (Bari and Jones, 2009; Robert-Seilaniantz et al., 2011; Yang et al., 2015). Because the hormones themselves or modified version of them are soluble molecules or transport machinery exists, the hormones serve to prime defenses in distal tissues, promoting immunity beyond the site of

infection (Bari and Jones, 2009). The pathways lead to activation of transcription factors that regulate production of a variety of antimicrobial and anti-herbivore proteins, as well as increased callose deposition and lignification of cell walls, and increased ROS production (Bari and Jones, 2009; Pieterse et al., 2012).

Induced defenses against microbes

The signaling cascades described above lead to induction of a wide variety of chemicals and proteins with anti-microbial functions. One broad category of these genes are the Pathogenesis-Related (PR) genes, which are 17 families of genes encoding proteins with functions related to degradation of pathogen cell walls and membranes, protein inhibition and degradation, direct chemical toxicity, and regulation of cellular redox (van Loon and van Strien, 1999; van Loon et al., 2006). Individual PR genes are often used as markers for defense induction of the SA and JA signaling pathways. These families are discussed in detail in Chapter 3.

Other classes of genes have direct or indirect anti-microbial activity, but are not among the canonical PR gene families. One of these families is the polygalacturonase inhibiting protein (PGIPs). Plant pathogens secrete enzymes, including polygalacturonases (Idnurm and Howlett, 2001), to cleave plant cell wall components, and accordingly, plants produce PGIPs to inhibit this activity (De Lorenzo and Ferrari, 2002; Howell and Davis, 2005). A wide variety of small secreted peptides also have direct antimicrobial action (Tavormina et al., 2015). Many of these are formed by post-translational modification of inactive precursors.

Another group of molecules known to be induced by biotic stress are flavonoids, polyphenolic secondary metabolites that contribute pigmentation to plant tissue (Falcone Ferreyra et al., 2012). They often act as chemical signals in repelling or attracting insects and pathogens. They can play a protective mechanism, as they are able to scavenge ROS and bind and chelate ROS producing enzymes (Williams et al., 2004; Agati et al., 2012). Infection in soybean resulted in increased transcription of specific branches of flavonoid synthesis, including isoflavones and isoflavonones, and decreased transcription of anthocyanin synthesis pathway

members (Zou et al., 2005). It is assumed that this differential response prioritizes production of ROS scavenging flavonoids over those with strict roles in pigmentation and photosynthesis during infections (Samac and Graham, 2007). Increased phenylpropanoid synthesis was also linked to specific R-gene dependent resistance mechanisms (Torregrosa et al., 2004; Subramanian et al., 2005).

Durability of defense and immune memory

ETI and hormone signaling can activate the defense response for days to weeks, depending on severity of pathogen stress and its persistence in the environment (Pieterse et al., 2012; Fu and Dong, 2013). After biotic stress, changes in methylation of regions of the genome containing defense genes have been detected, which likely represses or de-represses branches of immunity more important in fending off the pathogen's reappearance (Dowen et al., 2012). For example, treatment with pathogen and an SA analog led to accumulation of histone modifications in promoters of WRKY transcription factors in distal tissues, and these were associated with altered expression after subsequent stress (Jaskiewicz et al., 2011). In the absence of pathogen challenge, chromatin remodeling proteins and DNA repair machinery have also been linked to decreased expression of PR genes, likely recruited to promoters through interaction with transcription factors and subsequently affecting local epigenetic tags (Song et al., 2011). DNA methylation and histone modifications can be heritable in plants, leading to heritable changes in defense gene expression (Heard and Martienssen, 2014). Evidence suggests that transgenerational modifications have similar effects as those caused by histone modifications within generations, leading to enhanced basal expression of defense genes and more rapid induction when pathogens are detected (Slaughter et al., 2012; Balmer et al., 2015). While extremely important, the study of immune memory in plants is a relatively new field. Further elucidation of processes at the intersection of epigenetics, defense, and heritability will be vital to improving plant breeding programs.

Cacao and molecular studies of defense

While plant defense is an extremely active area of study in many model and crop plants, studies investigating molecular interactions of plants and pathogens in cacao are sparse. Several studies have focused on functions of endogenous cacao defense genes. Stable overexpression of a class I chitinase was shown to inhibit growth of the fungus *Colletotrichum gloeosporioides* (Maximova et al., 2006), and transient overexpression of the same gene in leaves inhibited growth of *Phytophthora tropicalis* (Fister et al., 2016). Cacao NPR1 was characterized, it was shown to complement Arabidopsis npr1 mutants (Shi et al., 2010), and its transient overexpression of cacao NPR1 was also shown to enhance resistance to *Phytophthora* infection (Fister et al., 2015). A purified recombinant β-1,3-1,4 glucanase (Britto et al., 2013) and a purified recombinant PR-4 family chitinase (Pereira Menezes et al., 2014), both encoded by cacao, were both shown to have antifungal activity. Other studies have explored expression of exogenous proteins in cacao tissue. Stable overexpression of synthetic antimicrobial peptides also reduced disease symptoms after inoculation of leaves with two *Phytophthora* species (Mejia et al., 2012). Stable and transient expression of non-plant PI3P binding proteins in cacao improved resistance to fungal and oomycete pathogens, likely by blocking effector entry into cells (Helliwell et al., 2016). Several large transcriptomic experiments have been carried out to study cacao's defense pathways. Measuring the effect of salicylic acid treatment on two cacao varieties revealed genotype specificity in their responses (Fister et al., 2015). Transcriptomic changes resulting from treatment of cacao with endophytic fungi have been studied to improve understanding of how application of biologics regulates defense (Mejía et al., 2014). Gene regulation in response to witches' broom (Teixeira et al., 2014), *Phytophthora palmivora*, and *Colletotrichum theobromicola* have also been examined. While these large experiments have described trends in gene regulation, and a few genes' functions have been validated, little is known about specific protein interaction mechanisms in cacao. For example, young cacao plants infected with witches' broom showed increased expression of RLKs and NLRs (Teixeira et al., 2014), but no direct interaction of a cacao R gene with an effector from any of its pathogens has been described. Accordingly, the conclusions created from studies in model species motivate the molecular research performed in cacao.

1.4. Evolution and plant defense mechanisms

Plant genomes and their evolution

Understanding plant genome structure and organization is integral for developing strategies to study defense processes. The ability to sequence genomes and transcriptomes provided a new means of studying structural and functional genetics. Strategies for sequencing plant genomes have themselves evolved over the past two decades, as have the goals for performing genome sequencing (Bolger et al., 2014). Next Generation Sequencing strategies dramatically reduced the cost and time required to sequence a genome, allowing resequencing projects which focus on sequencing of hundreds to thousands of individuals from a species. The data produced allow higher resolution QTL mapping as the sequencing projects identify thousands of SNPs. Genome resequencing in crops has allowed for novel insights into loci controlling the defense response (Whiteman and Jander, 2010), abiotic stresses (Huang et al., 2009), plant maturation and flowering (Xia et al., 2012), all of which can greatly benefit productivity.

The availability of genome sequence data has revolutionized approaches for plant evolutionary and comparative –omics analyses, and the new data have emphasized the role of duplication events in plant evolutionary history. Phylogenetic data indicate that at least two whole genome polyploidization events occurred in early in land plant evolution, one predating seed plant divergence and another predating the divergence of monocots and dicots (Jiao et al., 2011), with more duplications occurring in specific lineages of monocots (Tang et al., 2010) and dicots (Barker et al., 2009). These large scale duplications not only increase genome size, but also enable functional diversification of gene families by relieving selective constraints (Lynch and Conery, 2000). While plant genome size ranges from ~63MB to nearly 150GB, evolutionary trends have been detected that explain gene and regulatory conservation across the plant kingdom (Dodsworth et al., 2015).

The vast differences in genome size are largely accounted for by transposable elements and other repetitive sequences; however, there remains a roughly two-fold range in the number

of predicted genes in sequenced plant species (Salse, 2012). Some of the variability in gene count is attributed to generation time (Sterck et al., 2007). With Arabidopsis, which has one of the lowest numbers of annotated genes, being an annual, a plant contributes gametes only to its generation, where longer-lived species may retain additional copies of genes because older, but still reproductively viable, individuals act as reservoirs for genetic redundancy (Van de Peer et al., 2009). However, this model does not explain variation in gene count among trees. Speciation events, which can involve dramatic changes in gene content, and lineage-specific segmental duplications, often driven by transposable elements, polymerase slippage, or unequal crossing-over (Freeling, 2009), also contribute to differences in gene count (Rabinowicz et al., 2005; Wendel et al., 2016). Immediately after duplications, the presence of two copies of a gene can allow mutations to occur for one copy without the same detrimental phenotypic effects seen after mutation of the parent sequence. Molecular evolutionary theory and *in silico* models built from the data of more than a dozen sequenced genomes have shed light on two processes controlling 'diploidization' of paleopolyploid genomes; sub-genome dominance and neofunctionalization (Barker et al., 2012; Salse, 2012; Wendel et al., 2016). Sub-genome dominance describes the tendency of one genome in a polyploid to retain functional versions of genes while the other accumulates mutations and is deleted, occasionally in large blocks (Schnable et al., 2011). Neofunctionalization is the process of mutations affecting the structure of a duplicated gene's regulatory sequences or coding sequence, thereby altering the protein's direct functionality or the gene's expression dynamics (Barker et al., 2012). Often this allows one paralog to retain the ancestral function while the other develops a new specialization. Another possibility is that one paralog simply accumulates mutations, making it a non-functional pseudogene. Certain classes of genes have been shown to tolerate the different types of duplication events with differential success. Genes associated with environmental responses, including biotic defense, often have more members occurring in locally duplicated blocks, while genes in metabolic pathways and those involved in regulatory processes often have more members surviving whole genome duplications (Rizzon et al., 2006). Therefore, stress response genes are often found to be physically linked. However, different genes encoding proteins in metabolic pathways have been shown, through an unknown mechanism, to physically cluster and become co-regulated (Chae et al., 2014). The genomic dynamics governing these processes remain an active area of study.

Plant-pathogen coevolution

The gene dosage theory describes the model whereby duplication of genes producing molecules which act in precise stoichiometry with other molecules would be deleterious, as disruption of the stoichiometry may inhibit the process (Birchler and Veitia, 2007). This model has been invoked to explain why certain functional classes of genes are more likely to persist after expansion by whole genome duplication or by local, segmental duplication (Sterck et al., 2007). Interestingly, abiotic and biotic stress response genes often have high retention rates after any duplication, implying that expansion of these families is favored to allow adaptation to a changing environment (Casneuf et al., 2006). R genes, PR family members, and other defense genes were also shown to frequently persist in tandem arrays, indicating continued evolutionary tolerance for expansion of the families (Cannon et al., 2004).

R genes in particular have been the focus of a great deal of evolutionary analyses, and are found to be extremely variable both in that they have many single nucleotide polymorphisms and expression dynamics often vary between individuals (Karasov et al., 2014). R genes often exist in clusters with other related genes. The repetitive nature of these regions makes polymerases more prone to slippage and increases the likelihood of recombination, both of which increase the likelihood of mutations altering sequences (Michelmore and Meyers, 1998; Wicker et al., 2007). This positive feedback creates more variation, which becomes beneficial as it increases the likelihood of a new variant being created that will be able to recognize effectors, which are also encoded in gene clusters, making their genomic regions also hypermutagenic (Raffaele and Kamoun, 2012). Consequently, R genes and other defense genes often show signatures of diversifying selection, whereby multiple haplotypes are favored in populations as this increases the likelihood of members of the population being able to bind variants of a fungal effector protein (McDowell et al., 1998). While variability is favored in both the plant and the pathogen populations, there is evidence that the possible amount of variability is limited. R genes from multiple species, when transformed into rice, were able to confer resistance to rice blast (Yang et al., 2013). This led to the proposal of a model describing ‘constrained divergence,’ according to which only a limited set of evolutionary pathways are

available for effectors and R genes. Application of biotechnological approaches therefore enables trans-specific conferral of resistance in transgenic plants.

R genes and effectors are not the only interacting molecules affected by co-evolution of the host and pathogen. Enzymes with direct roles in degradation of the other individual are also affected. For example, pathogenic plant cell wall-degrading enzymes show signatures of both diversifying and purifying selection (Brunner et al., 2013). The authors suggest that gene under purifying selection have highly constrained structures that allow optimized activity on cell wall substrates, whereas those under diversifying selection are detected by plant proteins. Similarly, plant chitinases show positive and negative selection in their chitin binding sites, likely enhancing substrate specificity and avoiding detection by pathogenic inhibitory proteins, respectively (Bishop et al., 2000). Consequently, these inhibitory proteins, like polygalacturonase inhibitors, show signatures of diversifying selection, allowing recognition of variability in wall-degrading enzymes (Misas-Villamil and van der Hoorn, 2008).

Several models have been proposed through which genetic variation, particularly in defense genes, can be maintained within a species. One model, frequency dependent selection, describes a scenario where the strength of selection for a given allele is inversely proportional to the frequency of the allele, so that over time, the allele's frequency oscillates (Tellier and Brown, 2007). Local adaptation can lead to different alleles dominating in sub-populations of a species in cases where the sub-populations are responding to different pressures (North et al., 2011). Finally, heterozygote advantage can be beneficial, for instance allowing one individual to harbor two R gene haplotypes capable of recognizing two different effector variants (Sellis et al., 2011). Determining which, if any, of these patterns is occurring is difficult and can be further muddied by population structure (Moeller et al., 2007).

1.5 Dissertation Overview

The plant defense response has been an intensively researched field for several decades. Every subheading of this literature review has been the subject of at least one review article or textbook. Nonetheless the surface has only been scratched, especially with regard to

applying the canon to improving crop species. Integration of the wealth of knowledge already created is essential for designing new experiments and breeding programs to improve crops like cacao.

The core question this dissertation attempts to answer is a deceptively simple one: what genes are most important in cacao's defense response? The breadth of this literature review belies the underlying complexity of this problem. The defense response is highly nuanced, with differentially responsive genes acting against a variety of pathogen. Further, the distribution of cacao germplasm around the world is heterogeneous, and as a consequence, only certain genotypes interact with certain pathogens. The history of these interactions likely altered the response within some populations, which may or may not have been incorporated into breeding programs. To make the problem manageable, the chapters focus on sub-questions that address several of the most important points for understanding defense in cacao.

One challenge within exploring cacao's defense response is definitional: what components does cacao have in terms of gene family size and activity of members, and how do these components compare to other species. While the publication of the Criollo genome presented an overview of cacao R genes, the induced defenses, including the PR families, were not explicitly defined. Chapter 2 is a bioinformatic identification of PR gene family members in cacao, and it includes a structural comparison of these gene families to those of several monocots and dicots. Within we also describe the transcriptomic response of the gene families to two cacao pathogens in order to identify which members of these gene families are responsive in leaf tissue.

Genotype specificity of the defense response is also a challenge for studying a crop plant. Chapter 3 focuses on this question and presents another transcriptomic analysis, the effect of treatment of two genotypes with salicylic acid, an important defense hormone. It focuses on two widely studied genotypes, a model disease-tolerant variety, Scavina 6 (Sca6), and a model highly susceptible variety, Imperial College Selection 1 (ICS1). Both are often used in breeding programs, Sca6 to introduce resistance alleles, and ICS1 to improve flavor quality traits.

While studying differential defense induction in two genotypes is useful, it defines only two possible reactions, and without sequence data, the underlying genetic mechanisms remain obscure. Consideration of processes at a finer resolution will be required to identify key defense components in cacao; for while a given gene may be important in defense, there may be haplotypes of that gene with significant effects in pathogen recognition, while other haplotypes may be non-functional. Therefore, it is essential to explore the genetic diversity within candidate defense genes in entire populations of cacao to explore the extent of variation that exists. Chapter 4 is an evaluation of genetic diversity using three defense genes and cacao plants representing three geographically distinct cacao populations. This type of analysis can identify loci under selection, thereby indicating which defense genes are likely to interact directly with cacao's pathogens.

Integral to functional analysis of defense genes is having a protocol for screening the effect of gene overexpression or knockdown. The Guiltinan-Maximova Lab has developed a protocol for transient transformation of cacao leaf tissue and subsequent pathogen inoculation for this purpose. While the technique is applied within Chapter 3, Chapter 5 presents our highly optimized protocol in full, exploring variable transformation success in a wide array of genotypes and different tissue stages. Chapter 5 also describes our detached leaf pathogen inoculation assay and presents preliminary data showing variability in basal defense between genotypes.

One goal of this dissertation is to review the literature on cacao, plant defense, and crop improvement methods in order to define a strategy for defense gene prioritization and functional analysis in cacao. This strategy is described in Appendix A. Genome and transcriptome sequencing, leveraged with QTL maps and comparative genomics, offer a wealth of data that can be used to prioritize genes for further study. While the scheme is only one route for defense gene prioritization, it is a mean of filtering the thousands of genes involved in defense to choose several candidates which may be critical for cacao immunity.

Finally, Chapter 7 offers a retrospective on promising aspects and shortcomings of the methods applied and considers future experiments that are vital to furthering the improvement of cacao. Crop plants like cacao are increasingly amenable to genomic and transcriptomic

analyses. Possible directions for future experiments probing cacao's defense response are discussed.

Acebo-Guerrero Y, Hernández-Rodríguez A, Heydrich-Pérez M, El Jaziri M, Hernández-Lauzardo AN (2012) Management of black pod rot in cacao (Theobroma cacao L.): a review. Fruits **67:** 41-48

Agati G, Azzarello E, Pollastri S, Tattini M (2012) Flavonoids as antioxidants in plants: location and functional significance. Plant Science **196:** 67-76

Aime MC, Phillips-Mora W (2005) The causal agents of witches' broom and frosty pod rot of cacao (chocolate, Theobroma cacao) form a new lineage of Marasmiaceae. Mycologia **97:** 1012-1022

Alfano JR, Collmer A (2004) Type III secretion system effector proteins: double agents in bacterial disease and plant defense. Annual Review of Phytopathology **42:** 385-414

Alves RM, Sebbenn AM, Artero AS, Clement C, Figueira A (2007) High levels of genetic divergence and inbreeding in populations of cupuassu (Theobroma grandiflorum). Tree Genetics & Genomes **3:** 289-298

Anderson DM, Fouts DE, Collmer A, Schneewind O (1999) Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals. Proceedings of the National Academy of Sciences **96:** 12839-12843

Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, Maclean DJ, Ebert PR, Kazan K (2004) Antagonistic Interaction between Abscisic Acid and Jasmonate-Ethylene Signaling Pathways Modulates Defense Gene Expression and Disease Resistance in Arabidopsis. The Plant Cell **16:** 3460-3479

Apel K, Hirt H (2004) REACTIVE OXYGEN SPECIES: Metabolism, Oxidative Stress, and Signal Transduction. Annual Review of Plant Biology **55:** 373-399

Argout X, Salse J, Aury J, Guiltinan M, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova S (2011) The genome of Theobroma cacao. Nat Genet **43**

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W-L, Gomez-Gomez L (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. Nature **415**

Asselbergh B, De Vleesschauwer D, Höfte M (2008) Global Switches and Fine-Tuning—ABA Modulates Plant Pathogen Defense. Molecular Plant-Microbe Interactions **21:** 709-719

Bailey BA, Meinhardt LW (2016) Cacao Diseases: A History of Old Enemies and New Encounters. Springer

Balmer A, Pastor V, Gamir J, Flors V, Mauch-Mani B (2015) The ‘prime-ome’: towards a holistic approach to priming. Trends in Plant Science **20:** 443-452

Bari R, Jones JG (2009) Role of plant hormones in plant defense responses. Plant Molecular Biology **69:** 473-488

Barker M, Baute G, Liu S-L (2012) Duplications and Turnover in Plant Genomes. *In* JF Wendel, J Greilhuber, J Dolezel, IJ Leitch, eds, Plant Genome Diversity Volume 1. Springer Vienna, pp 155-169

Barker MS, Vogel H, Schranz ME (2009) Paleopolyploidy in the Brassicales: Analyses of the Cleome Transcriptome Elucidate the History of Genome Duplications in Arabidopsis and Other Brassicales. Genome Biology and Evolution **1:** 391-399

Beckers GJM, Spoel SH (2006) Fine-Tuning Plant Defense Signalling: Salicylate versus Jasmonate. Plant biol (Stuttg) **8:** 1-10

Bergmann JF (1969) THE DISTRIBUTION OF CACAO CULTIVATION IN PRE-COLUMBIAN AMERICA. Annals of the Association of American Geographers **59:** 85-96

Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis Jeffrey G, Kobe B, Dodds Peter N (2011) Structural and Functional Analysis of a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling, and Autoregulation. Cell host & microbe **9:** 200-211

Bigeard J, Colcombet J, Hirt H (2015) Signaling Mechanisms in Pattern-Triggered Immunity (PTI). Molecular Plant **8:** 521-539

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. The Plant Cell **19:** 395-402

Bishop JG, Dean AM, Mitchell-Olds T (2000) Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. Proceedings of the National Academy of Sciences of the United States of America **97:** 5322-5327

Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KF (2014) Plant genome sequencing—applications for crop improvement. Current Opinion in Biotechnology **26:** 31-37

Boller T, Felix G (2009) A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. Annual Review of Plant Biology **60:** 379-406

Britto DS, Pirovani CP, Andrade BS, dos Santos TP, Pungartnik C, Cascardo JCM, Micheli F, Gesteira AS (2013) Recombinant β-1,3-1,4-glucanase from Theobroma cacao impairs Moniliophthora perniciosa mycelial growth. Molecular Biology Reports **40:** 5417-5427

Brown JS, Phillips-Mora W, Power EJ, Krol C, Cervantes-Martínez C, Motamayor JC, Schnell RJ (2007) Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao L*. **v. 47, no. 5, p. 1851-1858**

Brown JS, Schnell R, Motamayor J, Lopes U, Kuhn DN, Borrone JW (2005) Resistance gene mapping for witches' broom disease in Theobroma cacao L. in an F2 population using SSR markers and candidate genes. Journal of the American Society for Horticultural Science **130:** 366-373

Browse J (2009) Jasmonate Passes Muster: A Receptor and Targets for the Defense Hormone. Annual Review of Plant Biology **60:** 183-205

Brunner PC, Torriani SFF, Croll D, Stukenbrock EH, McDonald BA (2013) Coevolution and Life Cycle Specialization of Plant Cell Wall Degrading Enzymes in a Hemibiotrophic Pathogen. Molecular Biology and Evolution

Brunt AA (1970) Cacao swollen shoot virus. C. M. I. /A. A. B. Descr. Plant Viruses **10:** 4

Cannon S, Mitra A, Baumgarten A, Young N, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biology **4:** 10

Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y (2006) Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. Genome Biology **7:** R13

Chae L, Kim T, Nilo-Poyanco R, Rhee SY (2014) Genomic Signatures of Specialized Metabolism in Plants. Science **344:** 510-513

Chang C, Yu D, Jiao J, Jing S, Schulze-Lefert P, Shen Q-H (2013) Barley MLA Immune Receptors Directly Interfere with Antagonistically Acting Transcription Factors to Initiate Disease Resistance Signaling. The Plant Cell **25:** 1158-1173

Cheeseman EE (1944) Notes on the nomenclature, classification, and possible relationships of cocoa populations. Tropical Agriculture **21:** 144-159

Clement D, Risterucci AM, Motamayor JC, N'Goran J, Lanaud C (2003) Mapping QTL for yield components, vigor, and resistance to Phytophthora palmivora in Theobroma cacao L. Genome **46:** 204-212

Coll NS, Epple P, Dangl JL (2011) Programmed cell death in the plant immune system. Cell Death Differ **18:** 1247-1256

Coll NS, Vercammen D, Smidler A, Clover C, Van Breusegem F, Dangl JL, Epple P (2010) Arabidopsis Type I Metacaspases Control Cell Death. Science **330:** 1393-1397

Cornelis GR, Gijsegem FV (2000) Assembly and Function of Type III Secretory Systems. Annual Review of Microbiology **54:** 735-774

Cuatrecasas J (1964) Cacao and its allies. A taxonomic revision of the genus Theobroma. Contrib. US Nat. Herb **35:** 379-605

Cui H, Tsuda K, Parker JE (2015) Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annual Review of Plant Biology **66:** 487-511

Dardick C, Ronald P (2006) Plant and Animal Pathogen Recognition Receptors Signal through Non-RD Kinases. PLoS Pathogens **2:** e2

De Bruyne L, Höfte M, De Vleesschauwer D (2014) Connecting Growth and Defense: The Emerging Roles of Brassinosteroids and Gibberellins in Plant Innate Immunity. Molecular Plant **7:** 943-959

De Lorenzo G, Ferrari S (2002) Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. Current Opinion in Plant Biology **5:** 295-299

Deslandes L, Rivas S (2012) Catch me if you can: bacterial effectors and plant targets. Trends in Plant Science **17:** 644-655

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. Nat Rev Genet **11**

Dodsworth S, Leitch AR, Leitch IJ (2015) Genome size diversity in angiosperms and its influence on gene space. Current Opinion in Genetics & Development **35:** 73-78

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences **109:** E2183–E2191

Dufour B (1991) The role and importance of different insect species in the ecology of CSSV (cacao swollen shoot virus) in Togo. Café, cacao, thé **35:** 197-204

Duguma B, Gockowski J, Bakala J Smallholder Cacao (Theobroma cacao Linn.) cultivation in agroforestry systems of West and Central Africa: challenges and opportunities. Agroforestry Systems **51:** 177-188

Dzahini-Obiatey H, Domfeh O, Amoah F (2010) Over seventy years of a viral disease of cocoa in Ghana: from researchers' perspective. African Journal of Agricultural Research **5:** 476-485

Epaina P (2014) Identification of Molecular Markers and Quantitative Trait Loci linked to resistance to Vascular Streak Dieback and Phytophthora pod rot of Cacao Theobroma cacao L).

Erwin DC, Ribeiro OK (1996) Phytophthora diseases worldwide. American Phytopathological Society (APS Press), St. Paul, Minnesota

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends in Plant Science **5:** 199-206

Evans H, Krauss U, Rios Rutz R, Zecevich Acosta T, Arevalo-Gardini E (1998) Cocoa in Peru. Cocoa Growers' Bulletin**:** 7-22

Evans HC (2016) Witches' Broom Disease (Moniliophthora perniciosa): History and Biology. *In* AB Bailey, WL Meinhardt, eds, Cacao Diseases: A History of Old Enemies and New Encounters. Springer International Publishing, Cham, pp 137-177

Evans HC, Holmes KA, Reid AP (2003) Phylogeny of the frosty pod rot pathogen of cocoa. Plant Pathology **52:** 476-485

Falcone Ferreyra ML, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Frontiers in Plant Science **3:** 222

Faleiro FG, Queiroz VT, Lopes UV, Guimarães CT, Pires JL, Yamada MM, Araújo IS, Pereira MG, Schnell R, de Souza Filho GA (2006) Mapping QTLs for witches' broom (Crinipellis perniciosa) resistance in cacao (Theobroma cacao L.). Euphytica **149:** 227-235

Fan J, Hill L, Crooks C, Doerner P, Lamb C (2009) Abscisic Acid Has a Key Role in Modulating Diverse Plant-Pathogen Interactions. Plant Physiology **150:** 1750-1761

Feys BJ, Wiermer M, Bhat RA, Moisan LJ, Medina-Escobar N, Neu C (2005) Arabidopsis SENESCENCE-ASSOCIATED GENE101 stabilizes and signals within an ENHANCED DISEASE SUSCEPTIBILITY1 complex in plant innate immunity. Plant Cell **17**

Finkelstein R (2013) Abscisic Acid Synthesis and Response. The Arabidopsis Book / American Society of Plant Biologists **11:** e0166

Fister AS, O'Neil ST, Shi Z, Zhang Y, Tyler BM, Guiltinan MJ, Maximova SN (2015) Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. Journal of Experimental Botany **66:** 6245-6258

Fister AS, Shi Z, Zhang Y, Helliwell EE, Maximova SN, Guiltinan MJ (2016) Protocol: transient expression system for functional genomics in the tropical tree Theobroma cacao L. Plant Methods **12:** 1-13

Frébort I, Kowalska M, Hluska T, Frébortová J, Galuszka P (2011) Evolution of cytokinin biosynthesis and degradation. Journal of Experimental Botany **62:** 2431-2452

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology **60:** 433-453

Fu ZQ, Dong X (2013) Systemic Acquired Resistance: Turning Local Infection into Global Defense. Annual Review of Plant Biology **64:** 839-863

Fu ZQ, Guo M, Jeong B-r, Tian F, Elthon TE, Cerny RL, Staiger D, Alfano JR (2007) A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. Nature **447:** 284-288

Fu ZQ, Yan S, Saleh A, Wang W, Ruble J, Oka N, Mohan R, Spoel SH, Tada Y, Zheng N, Dong X (2012) NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. Nature **486:** 228-232

Gassmann W (2005) Natural Variation in the Arabidopsis Response to the Avirulence Gene hopPsyA Uncouples the Hypersensitive Response from Disease Resistance. Molecular Plant-Microbe Interactions **18:** 1054-1060

Glendinning DR (1972) NATURAL POLLINATION OF COCOA. New Phytologist **71:** 719-729

Gockowski J, Tchatat M, Dondjang J-P, Hietet G, Fouda T (2010) An Empirical Analysis of the Biodiversity and Economic Returns to Cocoa Agroforests in Southern Cameroon. Journal of Sustainable Forestry **29:** 638-670

Göhre V, Spallek T, Häweker H, Mersmann S, Mentzel T, Boller T, de Torres M, Mansfield JW, Robatzek S (2008) Plant Pattern-Recognition Receptor FLS2 Is Directed for Degradation by the Bacterial Ubiquitin Ligase AvrPtoB. Current Biology **18:** 1824-1832

Gómez-Gómez L, Boller T (2000) FLS2: An LRR Receptor–like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. Molecular Cell **5:** 1003-1011

Grande B, Colonial o, Holliday P (1952) Witches' broom disease of cacao : (Marasmius perniciosus Stahel). H.M. Stationery office, London

Griffith GW, Nicholson J, Nenninger A, Birch RN, Hedger JN (2003) Witches' brooms and frosty pods: two major pathogens of cacao. New Zealand Journal of Botany **41:** 423-435

Guest D (2007) Black pod: Diverse pathogens with a global impact on cocoa yield. Phytopathology **97:** 1650-1653

Guest D, Keane P (2007) Vascular-streak dieback: A new encounter disease of cacao in Papua New Guinea and Southeast Asia caused by the obligate basidiomycete Oncobasidium theobromae. Phytopathology **97:** 1654-1657

Guo H, Ecker JR (2003) Plant Responses to Ethylene Gas Are Mediated by SCFEBF1/EBF2-Dependent Proteolysis of EIN3 Transcription Factor. Cell **115:** 667-677

Guo H, Ecker JR (2004) The ethylene signaling pathway: new insights. Current Opinion in Plant Biology **7:** 40-49

Gutiérrez OA, Campbell AS, Phillips-Mora W (2016) Breeding for Disease Resistance in Cacao. *In* AB Bailey, WL Meinhardt, eds, Cacao Diseases: A History of Old Enemies and New Encounters. Springer International Publishing, Cham, pp 567-609

Hacquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, Legué V, Kohler A, Tanguay P, Petre B, Frey P, Van de Peer Y, Rouzé P, Martin F, Hamelin RC, Duplessis S (2011) A Comprehensive Analysis of Genes Encoding Small Secreted Proteins Identifies Candidate Effectors in Melampsora larici-populina (Poplar Leaf Rust). Molecular Plant-Microbe Interactions **25:** 279-293

Hausbeck MK, Lamour KH (2004) Phytophthora capsici on vegetable crops: research progress and management challenges. Plant Disease **88:** 1292-1303

Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, Eng JK, Akira S, Underhill DM, Aderem A (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature **410:** 1099-1103

Heard E, Martienssen RA (2014) Transgenerational Epigenetic Inheritance: myths and mechanisms. Cell **157:** 95-109

Hebbar PK (2007) Cacao Diseases: A Global Perspective from an Industry Point of View. Phytopathology **97:** 1658-1663

Heidrich K, Blanvillain-Baufumé S, Parker JE (2012) Molecular and spatial constraints on NB-LRR receptor signaling. Current Opinion in Plant Biology **15:** 385-391

Helliwell EE, Vega-Arreguín J, Shi Z, Bailey B, Xiao S, Maximova SN, Tyler BM, Guiltinan MJ (2016) Enhanced resistance in Theobroma cacao against oomycete and fungal

pathogens by secretion of phosphatidylinositol-3-phosphate-binding proteins. Plant Biotechnology Journal **14:** 875-886

Holliday P (1971) Some tropical plant pathogenic fungi of limited distribution. Review of Plant Pathology **50:** 337-348

Holt III BF, Boyes DC, Ellerström M, Siefers N, Wiig A, Kauffman S, Grant MR, Dangl JL (2002) An Evolutionarily Conserved Mediator of Plant Disease Resistance Gene Function Is Required for Normal Arabidopsis Development. Developmental Cell **2:** 807-817

Howell JT, Davis M (2005) Plant defense mechanisms against fungal pathogens: polygalacturonase inhibitor proteins. Canadian journal of plant pathology **27:** 5-15

Huang X-Y, Chao D-Y, Gao J-P, Zhu M-Z, Shi M, Lin H-X (2009) A previously unknown zinc finger protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. Genes & Development **23:** 1805-1817

Idnurm A, Howlett BJ (2001) Pathogenicity genes of phytopathogenic fungi. Molecular Plant Pathology **2:** 241-255

Jabs T, Dietrich RA, Dangl JL (1996) Initiation of Runaway Cell Death in an Arabidopsis Mutant by Extracellular Superoxide. Science **273:** 1853-1856

Jacob F, Vernaldi S, Maekawa T (2013) Evolution and conservation of plant NLR functions. Frontiers in Immunology **4**

Jaskiewicz M, Conrath U, Peterhänsel C (2011) Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. EMBO reports **12:** 50-55

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. The EMBO Journal **19:** 4004-4014

Jiang RH, Tripathy S, Govers F, Tyler BM (2008) RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members. Proceedings of the National Academy of Sciences **105:** 4874-4879

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature **473:** 97-100

Jones JDG, Dangl JL (2006) The plant immune system. Nature **444:** 323-329

Kale SD, Tyler BM (2011) Entry of oomycete and fungal effectors into plant and animal host cells. Cellular Microbiology **13:** 1839-1848

Karasov TL, Horton MW, Bergelson J (2014) Genomic variability as a driver of plant–pathogen coevolution? Current Opinion in Plant Biology **18:** 24-30

Keane PJ, Prior C (1991) Vascular-streak dieback of cocoa,

Kim MG, da Cunha L, McFall AJ, Belkhadir Y, DebRoy S, Dangl JL, Mackey D (2005) Two Pseudomonas syringae Type III Effectors Inhibit RIN4-Regulated Basal Defense in Arabidopsis. Cell **121:** 749-759

Knight R, Rogers HH (1955) Incompatibility in Theobroma cacao. Heredity **9:** 69-77

Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. Current Opinion in Structural Biology **11:** 725-732

Koeck M, Hardham AR, Dodds PN (2011) The role of effectors of biotrophic and hemibiotrophic fungi in infection. Cellular Microbiology **13:** 1849-1857

Krasileva KV, Dahlbeck D, Staskawicz BJ (2010) Activation of an Arabidopsis Resistance Protein Is Specified by the in Planta Association of Its Leucine-Rich Repeat Domain with the Cognate Oomycete Effector. The Plant Cell **22:** 2444-2458

Kroon LPNM, Brouwer H, de Cock AWAM, Govers F (2011) The Genus Phytophthora Anno 2012. Phytopathology **102:** 348-364

Laird SA, Obialor C, Skinner EA (1996) An introductory handbook to cocoa certification: a feasibility study and regional profile of West Africa. Rainforest Alliance

Lanaud C, Fouet O, Clément D, Boccara M, Risterucci AM, Surujdeo-Maharaj S, Legavre T, Argout X (2009) A meta–QTL analysis of disease resistance traits of Theobroma cacao L. Molecular Breeding **24:** 361-374

Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997) Evaluation of the extent of genetic variability among Theobroma cacao accessions using RAPD and RFLP markers. Theoretical and Applied Genetics **95:** 10-19

Li G, Meng X, Wang R, Mao G, Han L, Liu Y, Zhang S (2012) Dual-Level Regulation of ACC Synthase Activity by MPK3/MPK6 Cascade and Its Downstream WRKY Transcription Factor during Ethylene Induction in Arabidopsis. PLoS Genet **8:** e1002767

Li X, Lin H, Zhang W, Zou Y, Zhang J, Tang X, Zhou J-M (2005) Flagellin induces innate immunity in nonhost interactions that is suppressed by Pseudomonas syringae effectors. Proceedings of the National Academy of Sciences of the United States of America **102:** 12990-12995

Li Y, Chen L, Mu J, Zuo J (2013) LESION SIMULATING DISEASE1 Interacts with Catalases to Regulate Hypersensitive Cell Death in Arabidopsis. Plant Physiology **163:** 1059-1070

Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP (2011) Cacao breeding in Bahia, Brazil: strategies and results. Crop Breeding and Applied Biotechnology **11:** 73-81

Lynch M, Conery J (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151 - 1155

Macho AP, Zipfel C (2015) Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Current Opinion in Microbiology **23:** 14-22

Mackey D, Belkhadir Y, Alonso JM, Ecker JR, Dangl JL (2003) Arabidopsis RIN4 Is a Target of the Type III Virulence Effector AvrRpt2 and Modulates RPS2-Mediated Resistance. Cell **112:** 379-389

Marelli JP (2008) *Solanum lycopersicum* as a model system to study pathogenicity mechanisms of *Moniliophthora perniciosa*, the causal agent of witches' broom disease of *Theobroma cacao*. Pennsylvania State University, Pennsylvania State University

Marino D, Dunand C, Puppo A, Pauly N (2012) A burst of plant NADPH oxidases. Trends in Plant Science **17:** 9-15

Maximova S, Marelli J-P, Young A, Pishak S, Verica J, Guiltinan M (2006) Over-expression of a cacao class I chitinase gene in *Theobroma cacao L*. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224:** 740-749

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL (1998) Intragenic Recombination and Diversifying Selection Contribute to the Evolution of Downy Mildew Resistance at the RPP8 Locus of Arabidopsis. The Plant Cell **10:** 1861-1874

McManus MT (2012) Annual Plant Reviews, The Plant Hormone Ethylene. *In*, Ed 1. Wiley, Hoboken

Medeiros F, Pomella A, de Souza J, Niella G, Valle R, Bateman R, Fravel D, Vinyard B, Hebbar P (2010) A novel, integrated method for management of witches' broom disease in Cacao in Bahia, Brazil. Crop Protection **29:** 704-711

Meinhardt LW, Rincones J, Bailey BA, Aime MC, Griffith GW, Zhang D, Pereira GAG (2008) Moniliophthora perniciosa, the causal agent of witches' broom disease of cacao: what's new from this old foe? Molecular Plant Pathology **9:** 577-588

Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S (2012) Expression of designed antimicrobial peptides in Theobroma cacao L. trees reduces leaf necrosis caused by Phytophthora spp. Small Wonders Pept Dis Control **1905**

Mejía LC, Herre EA, Sparks JP, Winter K, García MN, Van Bael SA, Stitt J, Shi Z, Zhang Y, Guiltinan MJ, Maximova SN (2014) Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5**

Memelink J (2009) Regulation of gene expression by jasmonate hormones. Phytochemistry **70:** 1560-1570

Michelmore R, Meyers B (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res **8:** 1113 - 1130

Misas-Villamil JC, van der Hoorn RAL (2008) Enzyme–inhibitor interactions at the plant–pathogen interface. Current Opinion in Plant Biology **11:** 380-388

Moeller DA, Tenaillon MI, Tiffin P (2007) Population Structure and Its Effects on Patterns of Nucleotide Polymorphism in Teosinte (Zea mays ssp. parviglumis). Genetics **176:** 1799-1809

Motamayor J, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley S, Zheng P, Utro F, Royaert S (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biol **14**

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao L*). PLoS ONE **3:** e3311

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity **91:** 322-330

Mukhtar MS, Carvunis A-R, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, Pevzner SJ, Donovan SE, Ghamsari L, Santhanam B, Romero V, Poulin MM, Gebreab F, Gutierrez BJ, Tam S, Monachello D, Boxem M, Harbort CJ, McDonald N, Gai L, Chen H, He Y, Vandenhaute J, Roth FP, Hill DE, Ecker JR, Vidal M, Beynon J, Braun P, Dangl JL (2011) Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. Science **333:** 596-601

Muller E, Rao G, Khurana S, Lenardon S (2008) Cacao swollen shoot virus. Characterization, diagnosis & management of plant viruses. Volume 1: industrial crops**:** 423-444

Naseem M, Kaltdorf M, Dandekar T (2015) The nexus between growth and defense signalling: auxin and cytokinin modulate plant immune response pathways. Journal of Experimental Botany

North A, Pennanen J, Ovaskainen O, Laine A-L (2011) LOCAL ADAPTATION IN A CHANGING WORLD: THE ROLES OF GENE-FLOW, MUTATION, AND SEXUAL REPRODUCTION. Evolution **65:** 79-89

Nühse TS, Peck SC, Hirt H, Boller T (2000) Microbial Elicitors Induce Activation and Dual Phosphorylation of the Arabidopsis thaliana MAPK 6. Journal of Biological Chemistry **275:** 7521-7526

Opoku I, Appiah A, Akrofi A, Owusu G (2000) Phytophthora megakarya: a potential threat to the cocoa industry in Ghana. Ghana Journal of Agricultural Science **33:** 237-248

Padmanabhan MS, Ma S, Burch-Smith TM, Czymmek K, Huijser P, Dinesh-Kumar SP (2013) Novel Positive Regulatory Role for the SPL6 Transcription Factor in the N TIR-NB-LRR Receptor-Mediated Plant Innate Immunity. PLoS Pathog **9:** e1003235

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiology **150:** 1648-1655

Pereira JL, de Almeida LCC, Santos SM (1996) Witches' broom disease of cocoa in Bahia: attempts at eradication and containment. Crop Protection **15:** 743-752

Pereira Menezes S, de Andrade Silva E, Matos Lima E, Oliveira de Sousa A, Silva Andrade B, Santos Lima Lemos L, Peres Gramacho K, da Silva Gesteira A, Pirovani C, Micheli F (2014) The pathogenesis-related protein PR-4b from *Theobroma cacao* presents RNase activity, Ca2+ and Mg2+ dependent-DNase activity and antifungal action on *Moniliophthora perniciosa*. BMC Plant Biology **14:** 161

Petre B, Kamoun S (2014) How Do Filamentous Pathogens Deliver Effector Proteins into Plant Cells? PLoS Biology **12:** e1001801

Phillips-Mora W, Wilkinson MJ (2007) Frosty pod of cacao: a disease with a limited geographic range but unlimited potential for damage. Phytopathology **97:** 1644-1647

Phillips-Mora W, Coutiño A, Ortiz C, López A, Hernández J, Aime M (2006) First report of Moniliophthora roreri causing frosty pod rot (moniliasis disease) of cocoa in Mexico. Plant Pathology **55:** 584-584

Pieterse CMJ, Does DVd, Zamioudis C, Leon-Reyes A, Wees SCMV (2012) Hormonal Modulation of Plant Immunity. Annual Review of Cell and Developmental Biology **28:** 489-521

Pieterse JCM, Van Wees MSC (2015) Induced Disease Resistance. *In* B Lugtenberg, ed, Principles of Plant-Microbe Interactions: Microbes for Sustainable Agriculture. Springer International Publishing, Cham, pp 123-133

Ploetz R (2016) The Impact of Diseases on Cacao Production: A Global Overview. *In* Cacao Diseases. Springer, pp 33-59

Posnette A (1940) Transmission of'swollen shoot'disease of Cacao. Tropical Agriculture, Trinidad and Tobago **17**

Pound FJ (1943) Cacao and witches' broom disease (Marasmius perniciosa). Report on a recent visit to the Amazon territory of Peru, September, 1942 - February, 1943. Government Print, Port of Spain

Purdy LH, Schmidt RA (1996) STATUS OF CACAO WITCHES' BROOM: biology, epidemiology, and management. Annual Review of Phytopathology **34:** 573-594

Queiroz V, Guimarães C, Anhert D, Schuster I, Daher R, Pereira M, Miranda V, Loguercio L, Barros E, Moreira M (2003) Identification of a major QTL in cocoa (Theobroma cacao L.) associated with resistance to witches' broom disease. Plant Breeding **122:** 268-272

Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA (2005) Differential methylation of genes and repeats in land plants. Genome Research **15:** 1431-1440

Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Micro **10:** 417-430

Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN (2012) Challenges and progress towards understanding the role of effectors in plant–fungal interactions. Current Opinion in Plant Biology **15:** 477-482

Ravensdale M, Bernoux M, Ve T, Kobe B, Thrall PH, Ellis JG, Dodds PN (2012) Intramolecular Interaction Influences Binding of the Flax L5 and L6 Resistance Proteins to their AvrL567 Ligands. PLoS Pathog **8:** e1003004

Rice RA, Greenberg R (2000) Cacao Cultivation and the Conservation of Biological Diversity. AMBIO: A Journal of the Human Environment **29:** 167-173

Rietz S, Stamm A, Malonek S, Wagner S, Becker D, Medina-Escobar N, Corina Vlot A, Feys BJ, Niefind K, Parker JE (2011) Different roles of Enhanced Disease Susceptibility1 (EDS1) bound to and dissociated from Phytoalexin Deficient4 (PAD4) in Arabidopsis immunity. New Phytologist **191:** 107-119

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theoretical and Applied Genetics **108:** 168-174

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol **2:** e115

Robert-Seilaniantz A, Grant M, Jones JDG (2011) Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. Annual Review of Phytopathology **49:** 317-343

Royaert S, Jansen J, da Silva DV, de Jesus Branco SM, Livingstone Iii DS, Mustiga G, Marelli J-P, Araújo IS, Corrêa RX, Motamayor JC (2016) Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of Theobroma cacao L. in Brazil. BMC Genomics **17:** 1-16

Salse J (2012) In silico archeogenomics unveils modern plant genome organisation, regulation and evolution. Current Opinion in Plant Biology **15:** 122-130

Samac DA, Graham MA (2007) Recent Advances in Legume-Microbe Interactions: Recognition, Defense Response, and Symbiosis from a Genomic Perspective. Plant Physiology **144:** 582-587

Samuels GJ, Ismaiel A, Rosmana A, Junaid M, Guest D, Mcmahon P, Keane P, Purwantara A, Lambert S, Rodriguez-Carres M (2012) Vascular Streak Dieback of cacao in Southeast Asia and Melanesia: in planta detection of the pathogen and a new taxonomy. Fungal biology **116:** 11-23

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences **108:** 4069-4074

Sellis D, Callahan BJ, Petrov DA, Messer PW (2011) Heterozygote advantage as a natural consequence of adaptation in diploids. Proceedings of the National Academy of Sciences **108:** 20666-20671

Shen Q-H, Saijo Y, Mauch S, Biskup C, Bieri S, Keller B, Seki H, Ülker B, Somssich IE, Schulze-Lefert P (2007) Nuclear Activity of MLA Immune Receptors Links Isolate-Specific and Basal Disease-Resistance Responses. Science **315:** 1098-1103

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan M (2010) Functional analysis of the *Theobroma cacao* NPR1 gene in arabidopsis. BMC Plant Biology **10:** 248

Shiu S-H, Bleecker AB (2003) Expansion of the Receptor-Like Kinase/Pelle Gene Family and Receptor-Like Proteins in Arabidopsis. Plant Physiology **132:** 530-543

Silva CRS, Venturieri GA, Figueira A (2004) Description of Amazonian Theobroma L. collections, species identification, and characterization of interspecific hybrids. Acta Botanica Brasilica **18:** 333-341

Slaughter A, Daniel X, Flors V, Luna E, Hohn B, Mauch-Mani B (2012) Descendants of Primed Arabidopsis Plants Exhibit Resistance to Biotic Stress. Plant Physiology **158:** 835-843

Sohn KH, Segonzac C, Rallapalli G, Sarris PF, Woo JY, Williams SJ, Newman TE, Paek KH, Kobe B, Jones JDG (2014) The Nuclear Immune Receptor *RPS4* Is Required for *RRS1* ^{SLH1}-Dependent Constitutive Defense Activation in *Arabidopsis thaliana*. PLoS Genet **10:** e1004655

Song J, Durrant WE, Wang S, Yan S, Tan EH, Dong X (2011) DNA Repair Proteins Are Directly Involved in Regulation of Gene Expression during Plant Immune Response. Cell host & microbe **9:** 115-124

Song S, Qi T, Wasternack C, Xie D (2014) Jasmonate signaling and crosstalk with gibberellin and ethylene. Current Opinion in Plant Biology **21:** 112-119

Spanu P, Grosskopf DG, Felix G, Boller T (1994) The apparent turnover of 1-aminocyclopropane-1-carboxylate synthase in tomato cells is regulated by protein phosphorylation and dephosphorylation. Plant Physiology **106:** 529-535

Spoel SH, Dong X (2012) How do plants achieve immunity? Defense without specialized immune cells. Nature Reviews Immunolology **12:** 89-100

Stakman E (1915) Relation between Puccinia graminis and plants highly resistant to its attack. J Agric Res **4:** 193-200

Sterck L, Rombauts S, Vandepoele K, Rouzé P, Van de Peer Y (2007) How many genes are there in plants (… and why are they there)? Current Opinion in Plant Biology **10:** 199-203

Stergiopoulos I, Wit PJGMd (2009) Fungal Effector Proteins. Annual Review of Phytopathology **47:** 233-263

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA Interference of Soybean Isoflavone Synthase Genes Leads to Silencing in Tissues Distal to the Transformation Site and to Enhanced Susceptibility to Phytophthora sojae. Plant Physiology **137:** 1345-1353

Talbot P, Keane P (1971) Oncobasidium: A new genus of tulasnelloid fungi. Australian Journal of Botany **19:** 203-206

Tanaka K, Asami T, Yoshida S, Nakamura Y, Matsuo T, Okamoto S (2005) Brassinosteroid Homeostasis in Arabidopsis Is Ensured by Feedback Expressions of Multiple Genes Involved in Its Metabolism. Plant Physiology **138:** 1117-1125

Tanaka N, Matsuoka M, Kitano H, Asano T, Kaku H, Komatsu S (2006) gid1, a gibberellin-insensitive dwarf mutant, shows altered regulation of probenazole-inducible protein (PBZ1) in response to cold stress and pathogen attack. Plant, Cell & Environment **29:** 619-631

Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proceedings of the National Academy of Sciences **107:** 472-477

Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BPA (2015) The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. The Plant Cell **27:** 2095-2118

Teixeira PJPL, Thomazella DPdT, Reis O, do Prado PFV, do Rio MCS, Fiorin GL, José J, Costa GGL, Negri VA, Mondego JMC, Mieczkowski P, Pereira GAG (2014) High-resolution

transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. The Plant Cell **26:** 4245-4269

Tellier A, Brown JKM (2007) Stability of genetic polymorphism in host–parasite interactions. Proceedings of the Royal Society of London B: Biological Sciences **274:** 809-817

Ten Hoopen G, Krauss U (2016) Biological Control of Cacao Diseases. *In* Cacao Diseases. Springer, pp 511-566

Thomas E, van Zonneveld M, Loo J, Hodgkin T, Galluzzi G, van Etten J (2012) Present Spatial Diversity Patterns of <italic>Theobroma cacao</italic> L. in the Neotropics Reflect Genetic Differentiation in Pleistocene Refugia Followed by Human-Influenced Dispersal. PLoS ONE **7:** e47676

Thresh JM, Owusu GK (1986) The control of cocoa swollen shoot disease in Ghana: an evaluation of eradication procedures. Crop Protection **5:** 41-52

Torregrosa C, Cluzet S, Fournier J, Huguet T, Gamas P, Prospéri J-M, Esquerré-Tugayé M-T, Dumas B, Jacquet C (2004) Cytological, Genetic, and Molecular Analysis to Characterize Compatible and Incompatible Interactions Between Medicago truncatula and Colletotrichum trifolii. Molecular Plant-Microbe Interactions **17:** 909-920

Torres MA (2010) ROS in biotic interactions. Physiologia Plantarum **138:** 414-429

Van Aken O, Van Breusegem F (2015) Licensed to Kill: Mitochondria, Chloroplasts, and Cell Death. Trends in Plant Science **20:** 754-766

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: a tale of duplications. Trends in Plant Science **14:** 680-688

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology **44:** 135-162

van Loon LC, van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiological and Molecular Plant Pathology **55:** 85-97

Venugopal SC, Jeong R-D, Mandal MK, Zhu S, Chandra-Shekara AC, Xia Y, Hersh M, Stromberg AJ, Navarre D, Kachroo A, Kachroo P (2009) Enhanced Disease Susceptibility 1 and Salicylic Acid Act Redundantly to Regulate Resistance Gene-Mediated Signaling. PLoS Genet **5:** e1000545

Vlot AC, Dempsey DMA, Klessig DF (2009) Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47:** 177-206

Wawra S, Belmonte R, Löbach L, Saraiva M, Willems A, van West P (2012) Secretion, delivery and function of oomycete effector proteins. Current Opinion in Microbiology **15:** 685-691

Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. Genome Biology **17:** 1

Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, Armstrong MR, Grouffaud S, van West P, Chapman S, Hein I, Toth IK, Pritchard L, Birch PRJ (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature **450:** 115-118

Whiteman NK, Jander G (2010) Genome-Enabled Research on the Ecology of Plant-Insect Interactions. Plant Physiology **154:** 475-478

Wicker T, Yahiaoui N, Keller B (2007) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. The Plant Journal **51:** 631-641

Williams RJ, Spencer JPE, Rice-Evans C (2004) Flavonoids: antioxidants or signalling molecules? Free Radical Biology and Medicine **36:** 838-849

Wood G (1991) A history of early cocoa introductions. Cocoa Growers' Bulletin (United Kingdom)

Wood GAR, Lass R (2008) Cocoa. Longman Scientific & Technical, copublished by John Wiley & Sons, New York

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lü S, Wu H, Tabata S, Harada K (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proceedings of the National Academy of Sciences **109:** E2155–E2164

Xu Y, Chang P-FL, Liu D, Narasimhan ML, Raghothama KG, Hasegawa PM, Bressan RA (1994) Plant defense genes are synergistically induced by ethylene and methyl jasmonate. The Plant Cell **6:** 1077-1085

Yang D-L, Li Q, Deng Y-W, Lou Y-G, Wang M-Y, Zhou G-X, Zhang Y-Y, He Z-H (2008) Altered Disease Development in the eui Mutants and Eui Overexpressors Indicates that Gibberellins Negatively Regulate Rice Basal Disease Resistance. Molecular Plant **1:** 528-537

Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J-Q, Hartl DL, Tian D (2013) Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proceedings of the National Academy of Sciences **110:** 18572-18577

Yang Y-X, Ahammed GJ, Wu C, Fan S-y, Zhou Y-H (2015) Crosstalk among jasmonate, salicylate and ethylene signaling pathways in plant disease and immune responses. Current Protein & Peptide Science **16:** 450-461

Zhang D, Figueira A, Motilal L, Lachenaud P, Meinhardt LW (2011) Theobroma. *In* C Kole, ed, Wild Crop Relatives: Genomic and Breeding Resources: Plantation and Ornamental Crops. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 277-296

Zhang D, Motilal L (2016) Origin, Dispersal, and Current Global Distribution of Cacao Genetic Diversity. *In* Cacao Diseases. Springer, pp 3-31

Zhang J, Li W, Xiang T, Liu Z, Laluk K, Ding X, Zou Y, Gao M, Zhang X, Chen S, Mengiste T, Zhang Y, Zhou J-M (2010) Receptor-like Cytoplasmic Kinases Integrate Signaling from Multiple Plant Immune Receptors and Are Targeted by a Pseudomonas syringae Effector. Cell host & microbe **7:** 290-301

Zhang J, Shao F, Li Y, Cui H, Chen L, Li H, Zou Y, Long C, Lan L, Chai J, Chen S, Tang X, Zhou J-M (2007) A Pseudomonas syringae Effector Inactivates MAPKs to Suppress PAMP-Induced Immunity in Plants. Cell host & microbe **1:** 175-185

Zhang Y, Smith P, Maximova SN, Guiltinan MJ (2015) Application of glycerol as a foliar spray activates the defense response and enhances disease resistance of *Theobroma cacao*. Molecular Plant Pathology **16:** 27-37

Zipfel C (2009) Early molecular events in PAMP-triggered immunity. Current Opinion in Plant Biology **12:** 414-420

Zipfel C (2013) Combined roles of ethylene and endogenous peptides in regulating plant immunity and growth. Proceedings of the National Academy of Sciences **110:** 5748-5749

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JD, Felix G, Boller T (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature **428:** 764-767

Zou J, Rodriguez-Zas S, Aldea M, Li M, Zhu J, Gonzalez DO, Vodkin LO, DeLucia E, Clough SJ (2005) Expression Profiling Soybean Response to Pseudomonas syringae Reveals New

Defense-Related Genes and Rapid HR-Specific Downregulation of Photosynthesis. Molecular Plant-Microbe Interactions **18:** 1161-1174

Zullo MAT, Adam G (2002) Brassinosteroid phytohormones: structure, bioactivity and applications. Brazilian Journal of Plant Physiology **14:** 143-181

Chapter 2: *Theobroma cacao L*. Pathogenesis-Related Gene Tandem Array Members Show Diverse Expression Dynamics in Response to Pathogen Colonization

Published as:

Fister AS, Mejia LC, Zhang Y, Herre EA, Maximova SN, Guiltinan MJ (2016) *Theobroma cacao L*. pathogenesis-related gene tandem array members show diverse expression dynamics in response to pathogen colonization. BMC Genomics 17: 1-16.

Abstract

The Pathogenesis-Related (PR) group of proteins are operationally defined as polypeptides that increase in concentration in plant tissues upon contact with a pathogen. To date, 17 classes of highly divergent proteins have been described that act through multiple mechanisms of pathogen resistance. Characterizing these families in cacao, an economically important tree crop, and comparing the families to those in other species, is an important step in understanding cacao's immune response.

Using publically available resources, all members of the 17 recognized Pathogenesis-Related gene families in the genome of *Theobroma cacao* were identified and annotated resulting in a set of ~350 members in both published cacao genomes. Approximately 50% of these genes are organized in tandem arrays scattered throughout the genome. This feature was observed in five additional plant taxa (3 dicots and 2 monocots), suggesting that tandem duplication has played an important role in the evolution of the PR genes in higher plants. Expression profiling captured the dynamics and complexity of PR gene expression at basal levels and after induction by two cacao pathogens (the oomycete, *Phytophthora palmivora*, and the fungus, *Colletotrichum theobromicola*), identifying specific genes within families that are more responsive to pathogen challenge. Subsequent qRT-PCR validated the induction of several PR-1, PR-3, and PR-4 family members, with greater than 1000-fold induction detected for specific genes.

We describe candidate genes that are likely to be involved in cacao's defense against *Phytophthora* and *Colletotrichum* infection and could be potentially useful for marker-assisted selection for breeding of disease resistant cacao varieties. The data presented here, along with existing cacao –omics resources, will enable targeted functional genetic screening of defense genes likely to play critical functions in cacao's defense against its pathogens.

Background

Plant-microbe interactions leading to pathogenesis or resistance rely on a complex series of interactions between host and microbial molecules. The process begins when plant membrane-bound pattern recognition receptors (PRRs) detect microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) (Macho and Zipfel, 2014), or intracellular R genes bind secreted microbial effector proteins (Dangl and Jones, 2001) (Jones and Dangl, 2006; Kliebenstein, 2014). Recognition of pathogen presence activates multiple signal transduction cascades, including several interacting phytohormone signaling systems (Yang et al., 2015), which organize local and systemic responses to the infection including the activation of genes encoding antimicrobial proteins and enzymes involved in the synthesis of secondary metabolites with antimicrobial activities (Alvarez, 2000; Durrant and Dong, 2004; Jones and Dangl, 2006; Vlot et al., 2009; Fu and Dong, 2013). Ultimately, the plant's survival hinges on its ability to rapidly produce peptides and chemicals with antimicrobial properties. Understanding this process is integral to breeding for or engineering more resistant plant cultivars, a dire need for improved global food security and sustainable agriculture.

Pathogenesis-Related (PR) proteins, or as they have more recently been called, inducible defense-related proteins, have long been studied with regard to their importance in plant immunity (van Loon and van Strien, 1999; van Loon et al., 2006). The 17 families of genes that fall under the broad 'PR' classification encode a group of proteins with various antimicrobial properties and that were originally identified because certain family members show strong induction in response to biotic stress associated with activation of systemic acquired resistance signaling (van Loon and van Strien, 1999). Table 1 summarizes the roles of the 17 most

commonly acknowledged PR families based on extensive work in a variety of species. Overall, the PR families encode a diverse array of proteins involved in pathogen defense though multiple mechanisms.

A better understanding of the defense response in crop plants is integral to increasing the sustainability of food and feed production. Cacao production around the world is severely inhibited by cacao's susceptibility to pathogens, with roughly 40% of the crop lost annually, accounting for a multi-billion dollar loss of cocoa trade and chocolate industry annually (Guiltinan et al., 2008). Two high-quality cacao genome sequences have been acquired, that of the fine-flavor Belizean Criollo genotype (Argout et al., 2011) and the widely-cultivated Matina genotype (Motamayor et al., 2013). These resources enable new genome-wide strategies for characterizing the cacao defense response. To date, a handful of cacao PR genes have been studied, providing strong evidence that they play important roles in the response of cacao plants to pathogen infection. Application of glycerol to cacao leaves was recently found to promote defense and induce PR genes, likely through a fatty-acid-related signaling pathway (Zhang et al., 2015). The PR-1s of cacao were recently identified, with at least one showing induction by *Moniliopthora perniciosa*, the causal agent of cacao's witches broom disease (Teixeira et al., 2013). Specific members of the PR-3 (Maximova et al., 2006; Fister et al., 2016),PR-4 (Pereira Menezes et al., 2014), and PR-10 (Pungartnik et al., 2008; Menezes et al., 2012) families have also been the subject of functional characterization, focusing on enzymatic properties and roles in defense. The results of a recent RNA-seq study measuring induction of genes by witches' broom revealed that PR gene expression was elevated in infected tissues, but their induction (and induction of other known defense-related genes) was not sufficient to halt disease progression (Teixeira et al., 2014). A study by our group used a microarray to measure the effect of salicylic acid treatment on two cacao genotypes (Fister et al., 2015). Notably we found that PR gene induction levels differed between two contrasting genotypes, and surprisingly that more PR family members were induced in the more susceptible variety, ICS1, indicating that PR induction is only one piece of a successful defense response. Previously generated EST libraries (Gesteira et al., 2007; Argout et al., 2008) and focused gene expression measurements (Pereira Menezes et al., 2014; Fister et al., 2015) have begun to characterize genotype specificity of the defense response in cacao, but much more work is required to characterize defense

mechanisms across the described cacao populations (Motamayor et al., 2008). Much more work is required to characterize the tissue specificity, induction, and function of these genes in cacao to understand and harness their potential for combating the diversity of cacao pathogens.

With the goal of better understanding the evolution, structure, and expression dynamics of the cacao PR gene families, we carried out a comprehensive annotation and analysis of all PR gene families and characterized their genomic organization and expression in response to pathogens. Using a comparative genomics approach, we found that in cacao and in five other diverse plant species (*Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa, Populus trichocarpa, and Vitis vinifera*), PR gene family sizes are similar and members are often physically clustered in tandem arrays, with more than half of the family members existing in these arrays. Analyzing existing EST databases, we found support for expression of 62% of the *T. cacao* PR genes and identified many with expression limited to a specific tissue. Using a whole-genome microarray, we also identified PR gene family members induced by two major cacao pathogens, *Phytophthora palmivora* (Guest, 2007; Ploetz, 2007) and *Colletotrichum theobromicola* (Rojas et al., 2010), the causal agents of black pod rot and anthracnose, respectively. Comparing our new dataset to existing cacao transcriptomic analyses, we identified several PR genes strongly induced by multiple pathogens and treatments, suggesting potential roles as broad-spectrum defense response genes.

Table 2.1- Summary of PR gene families and their functions

PR Gene Class	Common Name	Function	References
PR-1	None (CAP/SCP superfamily)	Unknown.	(van Loon and van Strien, 1999; van Loon et al., 2006; Cantacessi et al., 2009)
PR-2	β-1,3-glucanase	Aid in cell wall degradation.	(van Loon and van Strien, 1999; van Loon et al., 2006; Balasubramanian et al., 2012)
PR-3	Chitinase – type I, II, IV, V, VI, VII	Aid in cell wall degradation.	(Brunner et al., 1998; van Loon and van Strien, 1999; van Loon et al., 2006; Grover, 2012)

PR-4	Chitinase - Hevein-like	Aid in cell wall degradation. May have RNase and DNase activity.	(Brunner et al., 1998; van Loon and van Strien, 1999; Caporale et al., 2004; van Loon et al., 2006; Grover, 2012; Lu et al., 2012; Pereira Menezes et al., 2014)
PR-5	Thaumatin-like	Degrade pathogen membranes.	(van Loon and van Strien, 1999; van Loon et al., 2006; Sels et al., 2008; Liu et al., 2010; Petre et al., 2011)
PR-6	Proteinase-inhibitor	Inhibit proteolysis by herbivorous insects.	(van Loon and van Strien, 1999; van Loon et al., 2006; Sels et al., 2008; Mithöfer and Boland, 2012)
PR-7	Endoproteinase	Aid in cell wall degradation.	(Tornero et al., 1996; van Loon and van Strien, 1999; van Loon et al., 2006)
PR-8	Chitinase - type III	Aid in cell wall degradation. May have lysozymal activity.	(Terwisscha van Scheltinga et al., 1996; Brunner et al., 1998; van Loon and van Strien, 1999; van Loon et al., 2006; Grover, 2012)
PR-9	Peroxidase	Regulate reactive oxygen species concentration, contribute to cell wall lignification.	(van Loon and van Strien, 1999; Passardi et al., 2004; van Loon et al., 2006)
PR-10	Ribonuclease-like	Degrade RNA, may degrade viruses.	(Walter et al., 1996; van Loon and van Strien, 1999; Park et al., 2004; van Loon et al., 2006)
PR-11	Chitinase - type I	Aid in cell wall degradation.	(Brunner et al., 1998; van Loon and van Strien, 1999; van Loon et al., 2006; Grover, 2012)
PR-12	Defensin	Degrade fungal membranes.	(van Loon and van Strien, 1999; van Loon et al., 2006; Stotz et al., 2009)
PR-13	Thionin	Directly permeabilize lipid bilayers.	(van Loon and van Strien, 1999; Stec, 2006; van Loon et al., 2006)
PR-14	Lipid-transfer Protein	Degrade pathogen membranes, mechanism unclear.	(van Loon and van Strien, 1999; van Loon et al., 2006;

			Yeats and Rose, 2008)
PR-15	Germin / Oxalate Oxidase	Regulate reactive oxygen species production.	(van Loon et al., 2006; Dunwell et al., 2008; Davidson et al., 2009)
PR-16	Germin-like / Oxalate Oxidase-like	Regulate reactive oxygen species production, catalyze monosaccharides.	(van Loon et al., 2006; Dunwell et al., 2008; Davidson et al., 2009)
PR-17	Putative Zinc-metalloproteinase	Proteinase function probable, mechanism unclear.	(Christensen et al., 2002; van Loon et al., 2006)

Materials and Methods

Theobroma cacao PR Gene Identification and Filtration

Amino acid sequences for the type members of each PR gene family (Supplemental Table S1) were used as queries to search the Criollo genome database using BLASTp (cutoff E < $1e^{-5}$, BLOSUM62 matrix) (Altschul et al., 1990). Using this strategy, we identified putative genes in 15 of the 17 known plant PR protein classes. PR-13s were not identified in the Criollo genome (they are specific to monocots and a subset of dicots (Stec, 2006)), and PR-15s are also considered to be monocot specific, although the BLASTp search finds them in the Criollo genome because of their homology with PR-16s (Dunwell et al., 2008). Next, a custom Python (python.org) (Cock et al., 2009) script (PRAminoacidgetterASF) was used to extract protein IDs from the BLASTp output and use them to extract the peptide sequences available in the Criollo cacao genome database.

The list of amino acid sequences was uploaded to the NCBI Batch Web CD-Search Tool (v3.13) (Marchler-Bauer et al., 2015) with an e-value cutoff of 0.01. Another script (PRdomainsorterASF) was used to sort the output of the CD-Search with gene IDs and BLASTp e-values of putative PR genes. Polypeptides were manually curated for the presence of domains used in Wanderly-Nogueira *et al.* (Wanderley-Nogueira et al., 2012) to classify each family. For the PR-6 family, we used presence of the “potato-inhibitor family domain” (pfam00280) to screen putative cacao PR genes, as it is the only domain found in the type member sequence.

Putative PR genes missing the characteristic domains were removed, and the remaining genes are listed in Supplemental Table S2.

This process was repeated for the Matina cacao genome (Motamayor et al., 2013). In order to compare PR gene distribution in the genomes, a third python script was used to retrieve positional information from the Criollo and Matina GFF files (PRstartstopfinderASF). This data was plotted in Fig. 1 (Criollo) and Supplemental Fig. S1 (Matina) using the R packages ggplot2 (Wickham, 2009) and ggbio (Yin et al., 2012), and gene positional information is also included in Supplemental Table S2 (Criollo) and Supplemental Table S3 (Matina). All python scripts are available on the Guiltinan-Maximova Lab website (http://plantscience.psu.edu/research/labs/guiltinan/protocols/ bioinformatic-scripts).

PR gene identification in other plant species

Using the same type member queries, BLASTp searches were against predicted polypeptide sequences downloaded from Phytozome v10.3 (Goodstein *et al*., 2012) from the *Arabidopsis thaliana* (TAIR10), *Brachypodium distachyon* (v3.1), *Oryza sativa* (v7.0), *Populus trichocarpa* (v3.0), and *Vitis vinifera* (Genoscope 12X) genomes using the same parameters. The procedure described above was used to curate, use CD-Search, and organize PR genes in order to count the number of genes per class. Tandem arrays were manually identified using JBrowse (Skinner et al., 2009) in Phytozome v10.3 (Goodstein et al., 2012). For all species, the PR-15 and PR-16 lists were largely redundant because of homology of the families, but PR-15s are monocot specific and should therefore only be present in *Brachypodium distachyon* and *Oryza sativa*. Therefore, for plotting gene family sizes in Fig. 2, these two families were combined. Gene IDs and BLASTp e-values for identified genes for these species are listed in Supplemental Tables S4-S8.

Plant growth, infection, and RNA extraction

Seeds from open pollinated *T. cacao* mother trees, accession UF12, were collected from a plantation in Charagre, Bocas del Toro province, Panama. The seeds were surface sterilized by immersing them in 0.5% sodium hypochlorite for three minutes and rinsed with sterile water before being placed for germination in plastic trays with soil (2:1 mixture of clay rich soil from

Barro Colorado Island, Panama and rinsed river sand) and incubated in Percival growth chambers. One-month-old seedlings were transplanted to individual pots (600 ml volume) containing the same soil mixture and kept in the growth chambers. Germination of seeds and seedling growth was done in growth chambers (model I35LL, 115 volts, 1/4 Hp, series: 8503122.16, Percival Scientific, Inc., Perry IA) with 12/12 h light/dark photoperiod and temperatures of 30°C and 26°C respectively (Mejía et al., 2014).

Two month old seedlings, with approximately six leaves each, were spray-inoculated with conidia of *Colletotrichum theobromicola* or zoospores of *Phytophthora palmivora*. Conidia of *C. theobromicola* were produced using the same methods as in (Mejía et al., 2014) for production of other species of *Colletotrichum* and zoospores were produced as in (Mejía et al., 2008). Whole seedlings were sprayed either with pathogen inoculum (*P. palmivora* isolate PTP zoospores at 5 x 10^4 per ml or *C. theobromicola* isolate ER08-11 conidia at 2 x 10^7 per ml) or sterile distilled water (controls) and then placed back into the growing chamber, but only leaves in stage C (Mejia et al., 2012) at the time of inoculation were considered as a target for the experiment. Pathogens *C. theobromicola* and *P. palmivora* were re-isolated from lesions developed in inoculated samples, which was interpreted as confirmation of successful colonization of plants by the pathogens. Samples were harvested from 72 h post-inoculation for RNA extraction, and tissue at this time point was used to re-isolate pathogen, which was considered as a measure of successful inoculation. Leaves sprayed with water remained healthy, did not develop lesions, and no pathogens were re-isolated from them. Representative photographs of infected and control leaves are shown in Supplemental Fig. S5. Four seedlings received each treatment, and five leaf samples were collected from each group of four seedlings. Each biological replicate consisted of a single individual leaf. Target leaves were cut with scissors from the plant, immediately weighed, and placed in RNAlater solution in borosilicate vials following manufacturer's instructions (Applied Biosystems/Ambion, Austin, TX). Vials containing samples were shipped to PSU on dry ice where RNA extractions were performed using a previously described protocol (Verica et al., 2004). Total RNA sample concentration and purity was assessed using a NanoDrop spectrophotometer and RNA quality was determined using an Agilent Bioanalzyer.

Building PR-1, PR-3 and PR-4 Phylogenies

To construct phylogenies, nucleotide sequences of family members for PR-1, PR-3, and PR-4 from the Criollo genome and primary transcripts from Arabidopsis (TAIR10) (Lamesch et al., 2012) were aligned using the MUSCLE (Edgar, 2004) translational alignment function in Geneious (Drummond et al., 2012) with eight iterations. Alignments were manually curated. No adjustments were made to the PR-1 or PR-3 families, but Tc05_g027340 was removed from the PR-4 alignment as it appears to have annotation errors in intron prediction. Maximum likelihood trees were generated in Geneious using a RAxML (Stamatakis, 2014) plugin.

Microarray Analysis

Transcriptomic analysis was performed using a whole-genome Roche NimbleGen custom oligo expression array (platform GPL18356), which was previously described in (Maximova et al., 2014). Probe labeling, hybridization, and detection were performed at the Penn State Genomics Core Facility, and the statistical analysis of the microarray data were performed as previously described (Maximova et al., 2014). Briefly, the Bioconductor package (Gentleman et al., 2004) was used in R to perform quality control checks and calculate normalized expression values using the RMA procedure. Normalized expression values were plotted to ensure all replicates for a given treatment had similar expression patterns. These data are available on GEO (GSE73804). In calculating fold induction, probes with mean $\log_2$ expression values across all probes less than 6 were removed. The LIMMA package (Smyth, 2004; Smyth, 2005) was then used to calculate fold induction on a per-probe basis and to calculate a Bayesian moderated test statistic for each comparison (pathogen-treatments relative to water-treatment). A Benjamini-Hochberg multiple testing correction (Benjamini and Hochberg, 1995) was then applied. Probes with Benjamini-Hochberg $p < 0.05$ were considered significant. In identifying individual PR genes with statistically significant differential regulation, any gene with multiple probes showing statistically significant change had fold change recalculated by averaging across all significant probes.

cDNA Synthesis and qRT-PCR validation of microarray

One microgram of RNA from each of the five samples from each treatment were reverse transcribed by M-MuLV Reverse Transcriptase (New England Biolabs, Ipswich, MA, USA) with oligo-$(dT)_{15}$ primers to obtain cDNA. To create highly specific primers for PR gene family members, nucleotide sequences for the PR-1, PR-3, PR-4, and PR-10 families were aligned using MUSCLE (Edgar, 2004) in Geneious (Drummond et al., 2012). qRT-PCR primers were designed to target bases that differentiate family members. Primer sequences are listed in Supplemental Table S15. qRT-qPCR was performed in a total reaction volume of 10 µL containing 4 µL of diluted cDNA (1:8), 5 µL of SYBR Green PCR Master Mix (TaKaRa, Mountain View, CA, USA), 0.2 µL of Rox and 0.4 µL of each 5 µM primer. Each reaction was performed on each of the five samples per treatment in technical duplicate using the Applied Biosystem Step One Plus Realtime PCR System (Nutley, NJ, USA) with the following program: 15 min at 94 °C, 40 cycles of 15 s at 94 °C, 20 s at 60 °C, and 40 s at 72 °C. The specificity of the primer pair was verified by dissociation curve.

Data normalization, a statistical randomization test, and relative pathogen-treated vs. water-treated expression ratios were computed using REST [64]. Fold changes with *p-values* less than 0.05 were considered significant.

Results

Identification of Cacao PR Gene Families

Using the Criollo cacao genome database (cocoagendb.cirad.fr/) (Argout et al., 2011), we developed a strategy for PR gene identification using the family type members described in van Loon *et al* (van Loon et al., 2006). This bioinformatics approach resulted in a total of 359 PR genes identified in the Criollo genome, and size of the families in the Criollo genome is listed in Table 2-2. Graphic representation of the genomic organization of these genes and the chromosomal positions of each of these loci is included in Fig. 1 and detailed information including gene IDs and chromosomal positions is provided in Supplemental Table S2. The

Table 2.2 – Summary of PR gene families in the *Theobroma cacao* Criollo genome

Common Name	Conserved Domain	Number of Peptides in Family	Best BLASTp hit
PR-1 CAP domain protein	SCP (smart00198)	14	3.00E-53
PR-2 β-1,3-glucanase	glyco hydro 17 (pfam00332)	43	7.00E-102
PR-3 Chitinase Class I, II, IV, VII	chitinase glyco hydro 19 (cd00325)	11	3.00E-79
PR-4 Chitinase - Hevein-like	barwin (pfam00967)	8	3.00E-49
PR-5 Thaumatin-like	thaumatin (pfam00314)	30	5.00E-72
PR-6 Proteinase-inhibitor	potato inhibitor family (pfam00280)	8	5.00E-11
PR-7 Endoproteinase	PA subtilisin like (cd02120)	54	0
PR-8 Chitinase Class III	GH18 hevamine XipI class III (cd02877)	14	2.00E-91
PR-9 Peroxidase	secretory peroxidase (cd00693)	81	4.00E-113
PR-10 Ribonuclease-like	Bet v1 (pfam00407)	23	3.00E-48
PR-11 Chitinase class V	GH18 plant chitinase class v (cd02879)	11	3.00E-116
PR-12 Defensin	gamma-thionin (pfam00304)	3	7.00E-10
PR-13	thionin (pfam00321)	0	NA

Thionin			
PR-14 Lipid-transfer Protein	nsLTP1 (cd01960)	16	6.00E-19
PR-15 Germin / Oxalate Oxidase	Two cupin 1 (pfam00190) domains	0	NA
PR-16 Germin-like / Oxalate Oxidase-like	Two cupin 1 (pfam00190) domains	38	2.00E-52
PR-17 Unknown	BSP (pfam04450)	5	7.00E-90
	Total	359 loci (38 unassembled)	

process of gene identification was repeated for the Matina cacao genome (Motamayor et al., 2008). The Matina PR chromosomal distribution is plotted in Supplemental Fig S1 and Matina gene IDs and their positions are listed in Supplemental Table S3. Overall, the family sizes and genomic organization of the gene families in the two genomes was similar, however we observed some differences that could be the result of either chromosomal rearrangements or assembly errors. For the subsequent analysis, we focused on the genes identified in the Criollo genome assembly.

In order to determine whether PR family sizes in cacao were similar to those in other species, we next applied the PR gene identification pipeline to the *Arabidopsis thaliana* (Lamesch et al., 2012)*, Brachypodium distachyon* (International Brachypodium Inititative, 2010)*, Populus trichocarpa* (Tuskan et al., 2006), *Oryza sativa* (Yu et al., 2002)*,* and *Vitis vinifera* (Jaillon et al., 2007) genomes. PR genes identified in these species are listed in Supplemental Tables S4-S8. We found that in these species as in cacao, PR genes typically existed as families rather than as single genes, with a notable exception being that our strategy only identified one PR-4, PR-8,

and PR-10 gene in the Arabidopsis genome. The size of gene families in cacao correlated well (R^2 > .85, $p < 0.001$) with PR family sizes in the other species (Fig. 2). Family sizes in cacao were typical of those in the other dicots, with no major species-specific family expansions or reductions. We also noticed trends of family conservation across the plant genomes; PR-11s were not found in the monocots (*Brachypodium distachyon* and *Oryza sativa*) surveyed, PR-12s were only in Arabidopsis and cacao, and PR-13s were found only in the monocots and Arabidopsis. The largest size disparity was in the PR-9s, where the two monocots had ~150 members while the dicots had less than 100 members.

Organization of PR gene families into tandem arrays

Criollo gene IDs indicate their order on chromosomes, where the first gene on chromosome 1 is Tc01_g000010, the second Tc01_g000020, etc. We noticed that many of the cacao PR genes were clustered with other members of the same family. To quantify this phenomenon, we defined a tandem array as any two or more genes of the same family that are located within 10 genes of one another (Rizzon et al., 2006; Lyons and Freeling, 2008). Using this parameter, we identified 46 PR tandem arrays containing a total 181 genes, distributed across all chromosomes (Fig. 1 and Supplemental Table S2). The number of genes within each tandem array ranged from two to sixteen across the families. The largest tandem arrays were a group of PR-10s on chromosome 4 (Chr4PR-10.6, 15 members), a group of PR-16s on chromosome 5 (Chr5PR-16.3, 14 members), a group of PR-11s on chromosome 9 (Chr9PR-11.1, 9 members), and a group of PR-9s on chromosome 2 (Chr2PR-9.5, 9 members). Next, using JBrowse (Skinner et al., 2009) we manually identified tandem arrays for each of the additional five species surveyed. We found that tandem arrays were very common across PR gene families in the diverse plant taxa surveyed (Supplemental Table S9), with more than half of the genes for most classes existing in tandem arrays. Proportions of PR family members found in tandem arrays, particularly among dicots, were also similar.

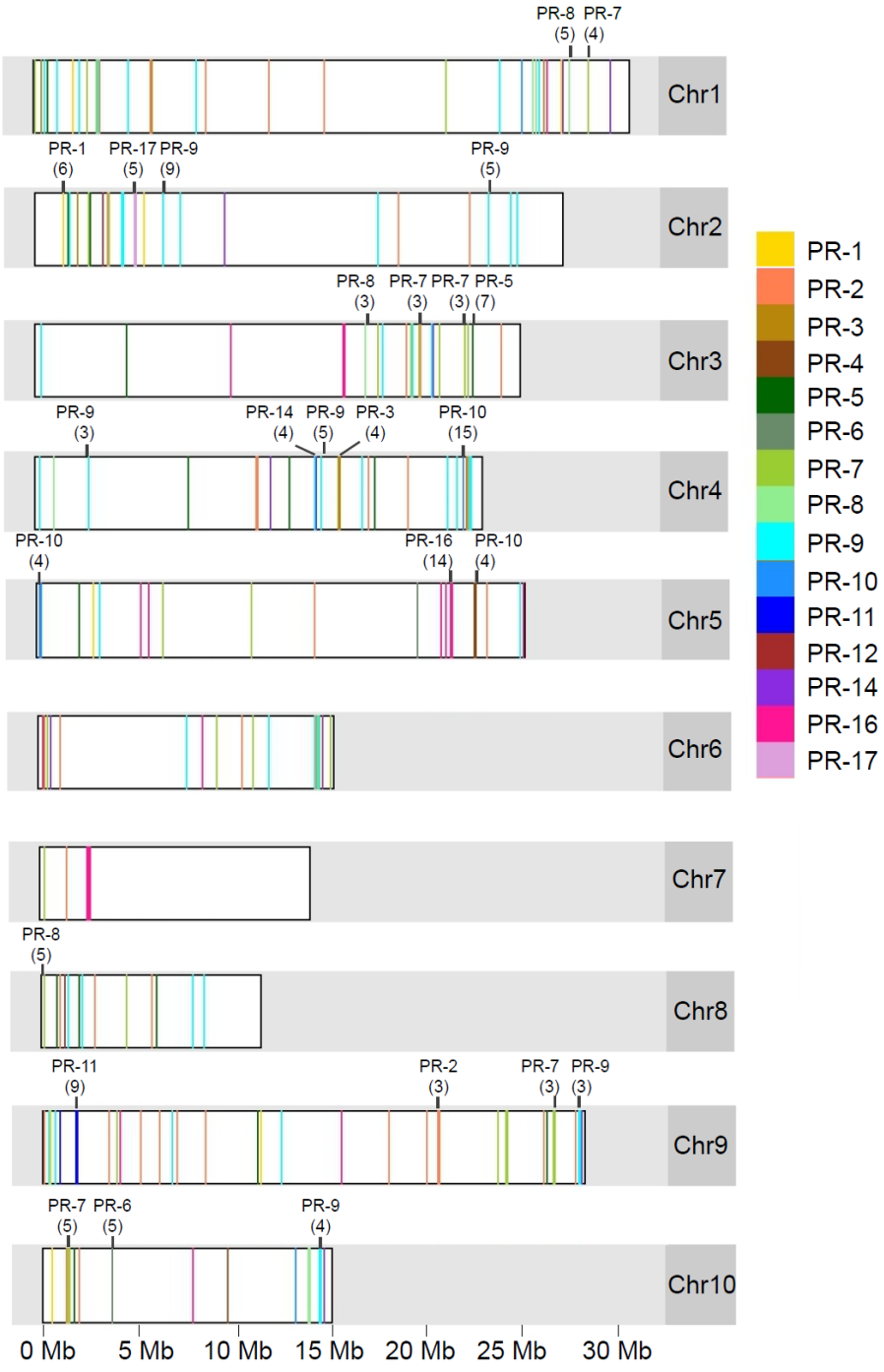


Figure 2.1. Karyogram depicting position of PR genes along the length of chromosomes based on the Criollo genome sequence. Tandem arrays are labelled above the chromosomes with gene family and number of genes in the array in parentheses. Length of chromosomes is shown in Mb. Due to resolution of the image lines representing nearby genes partially overlap.

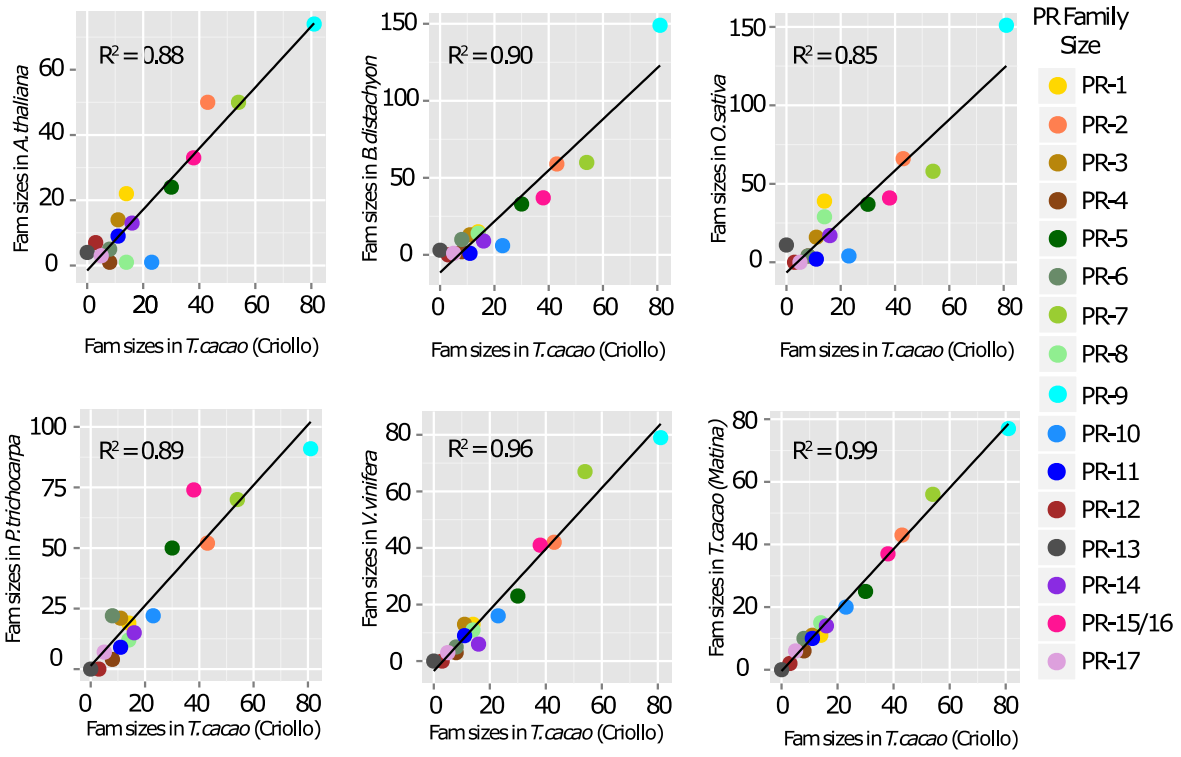


Figure 2.2 - Scatterplots comparing PR gene family size in the in the Criollo *T. cacao* genome to five plant species and the Matina *T. cacao* genome.

To investigate this phenomenon, we created maximum-likelihood trees for the PR-3 family (Fig. 3), the PR-1 family (Supplemental Figure S2, and the PR-4 family (Supplemental Figure S3), which include the gene family members from cacao and *Arabidopsis thaliana*. The phylogeny has several well-supported nodes indicating multiple PR-3 family members existed when Arabidopsis and cacao diverged. Further, the support for the tree suggests that there are three clades within the family. Cacao has tandem arrays in both clades B and C. Bootstrap support in clade B, interestingly, suggests that Tc01_g000770 is more closely related to Tc01_g010350 than it is to its tandem array members, Tc01_g000800. This suggests that in this scenario, a duplication led to the formation of an additional chitinase gene at the distal end of chromosome 1 after the tandem array had formed. Clade C contains tandem arrays of cacao and

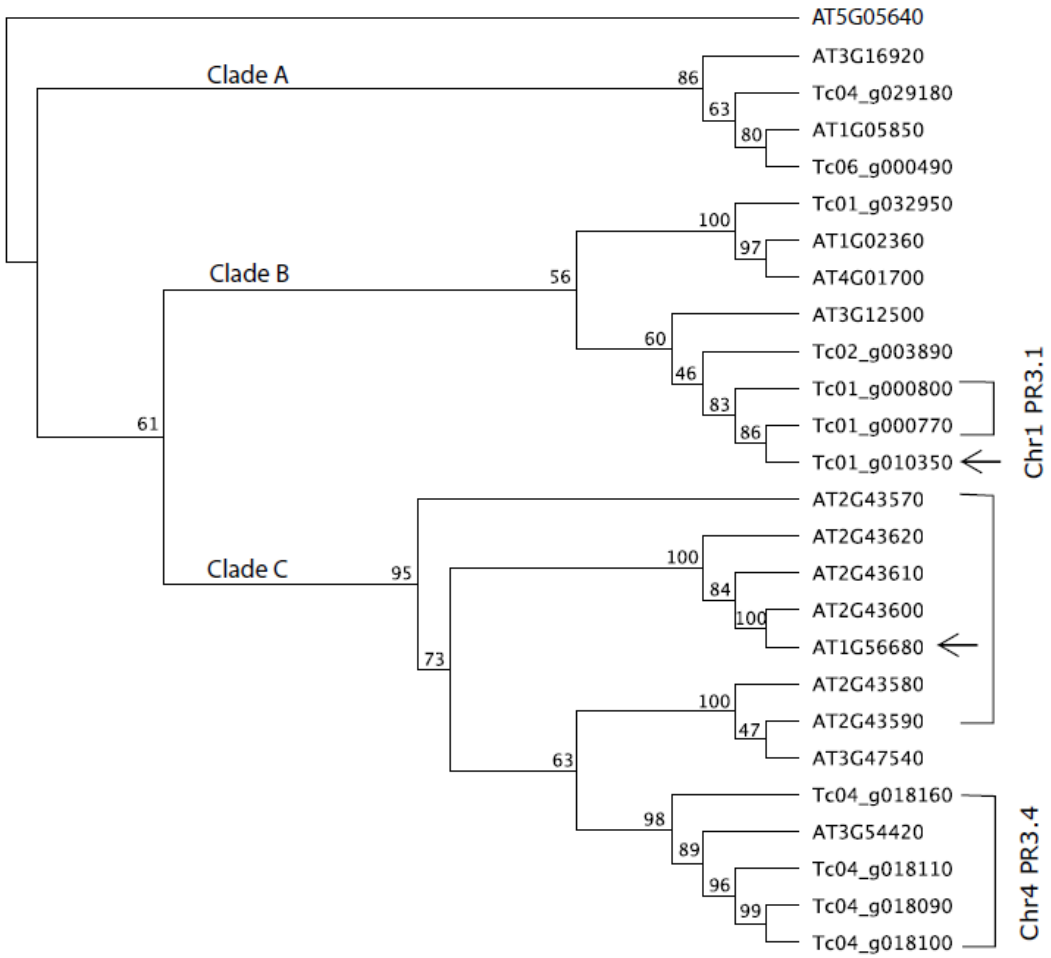


Figure 2.3 - Maximum-likelihood phylogeny of Criollo and Arabidopsis PR-3 family members. Node labels represent bootstrap support from 100 replicates. Brackets denote members of tandem arrays. Arrows indicate cases where non-tandem array members group most-closely with a tandem array member. Branch lengths represent genetic distance in substitutions per site. AT5G05460, a cytosolic beta-endo-N-acetyglucosaminidase and member of the chitinase superfamily, was included as an outgroup.

Arabidopsis genes. The branch support suggests that members of the Arabidopsis tandem array have continually expanded and diverged over evolutionary time, with strong support for array members split between three subclades. AT1G56690 presents another likely case of a recent non-local duplication, this one to a different chromosome. A fourth subclade contains the four members of the cacao tandem array on chromosome 4, none of which have been involved in recent duplications to other chromosomes. Examination of the PR-1 and PR-4 phylogenies also show evidence for expansion of gene families over evolutionary time locally, distally on chromosomes, and across chromosomes. Supplemental Tables S10-S12 include matrices of percentage identity for these three PR families, and further demonstrate that tandem array members are often, but not always, most closely related to one another.

Activation of cacao PR gene expression by pathogen colonization

To further our understanding of PR gene expression in cacao, we measured global gene expression after treating plants with two pathogens, *P. palmivora* and *C. theobromicola*. Fig. 4 A – B show scatterplots of $\log_2$ normalized expression for *P. palmivora* and *C. theobromicola* treatment, respectively, compared to water treatment for all probes corresponding to PR genes on a whole genome microarray, revealing that normalized expression values detected by the microarray reflect transcript abundance ranging from very low to very high (Supplemental Table S13) in all treatments. As expected, a similar trend was noted when analyzing all probes on the microarray (Supplemental Figure S4). For both pathogens, the majority of PR gene probes revealed constitutive expression across treatments, a large number of genes being up-regulated in pathogen-treated samples, and only a few examples of PR gene down-regulation. A total of 67 PR genes were induced by *P. palmivora* and 45 were induced by *C. theobromicola* (Benjamini-Hochberg-corrected $p < 0.05$ (Benjamini and Hochberg, 1995)) (Table 3). Of the two pathogen treatments, *P. palmivora* had a stronger effect in that in generally induced more genes per

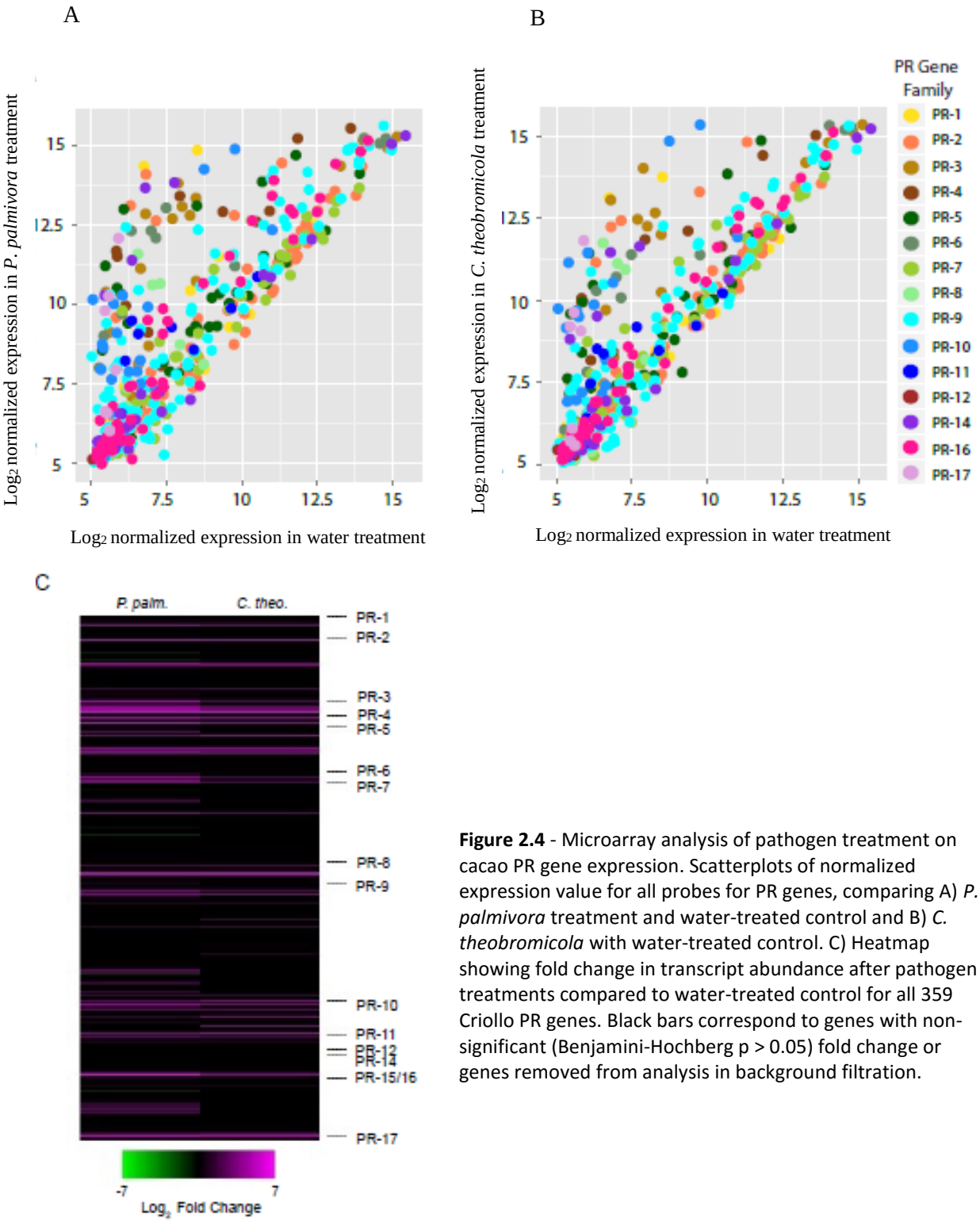


Figure 2.4 - Microarray analysis of pathogen treatment on cacao PR gene expression. Scatterplots of normalized expression value for all probes for PR genes, comparing A) *P. palmivora* treatment and water-treated control and B) *C. theobromicola* with water-treated control. C) Heatmap showing fold change in transcript abundance after pathogen treatments compared to water-treated control for all 359 Criollo PR genes. Black bars correspond to genes with non-significant (Benjamini-Hochberg $p > 0.05$) fold change or genes removed from analysis in background filtration.

family and the increase in transcript abundance relative to water-treated samples was greater (Fig. 4C, Supplemental Table S14). One exception was the PR-10s: while more of the PR-10 genes were induced by *P. palmivora*, those induced by both pathogens were equally or more strongly induced by *C. theobromicola.* A single PR-10 gene (Tc04_g028940) was strongly induced by *C. theobromicola* (log_2 3.6- fold increase) but not induced by *P. palmivora*. For both pathogens, statistically significant PR gene down-regulation was rare, as only 7 genes (2 PR-2s, 3 PR-7s, 1 PR-9, and 1 PR-16) were repressed by *P. palmivora* and none were by *C. theobromicola*. There

Table 2.3 - Regulation of Criollo PR genes as detected by microarray. Counts of up- and down-regulated genes represent the number of genes with Benjamini-Hochberg $p < 0.05$.

		P. palmivora		*C. theobromicola*	
	Number removed in background filtration (Average Log_2 Normalized Expression <6)	Up-regulated	Down-regulated	Up-regulated	Down-regulated
PR-1	7/14	1/14	0/14	1/14	0/14
PR-2	11/43	5/43	2/43	4/43	0/43
PR-3	1/11	8/11	0/11	5/11	0/11
PR-4	1/8	3/8	0/8	3/8	0/8
PR-5	6/30	6/30	0/30	5/30	0/30
PR-6	2/8	5/8	0/8	2/8	0/8
PR-7	21/54	2/54	3/54	1/54	0/54
PR-8	9/14	2/14	0/14	2/14	0/14
PR-9	26/81	12/81	1/81	7/81	0/81
PR-10	13/23	8/23	0/23	6/23	0/23
PR-11	5/11	3/11	0/11	3/11	0/11
PR-12	3/3	0/3	0/3	0/3	0/3
PR-14	3/16	2/16	0/16	2/16	0/16
PR-16	16/38	7/38	1/38	1/38	0/38
PR-17	2/5	3/5	0/5	3/5	0/5
Total	126/359	67/359	7/359	45/359	0/359

was also significant overlap in genes differentially regulated by the two pathogens. Forty-two PR genes were affected by both treatments, 32 were uniquely affected by *P. palmivora*, and 3 were unique to *C. theobromicola*. A large set of PR genes (159 in P. palmivora-treated samples and 188 in C. theobromicola-treated samples) were found to be expressed at similar levels in water and in pathogen treated tissues, suggesting that these genes may encode a set of proteins involved in basal defense in cacao, or they could be specifically induced in other tissues.

qRT-PCR validation of microarray results

To support the findings of our microarray analysis, we performed qRT-PCR on select genes from four families. Because family members, and tandem array members in particular, often have high similarity, with this analysis we sought to verify the specificity of microarray probes, as well as to confirm induction of genes of interest. Our analysis included 30 genes: 14 PR-1s, 6 PR-3s, 7 PR-4s and 3 PR-10s (Table 4). Primer sequences for qRT-PCR are listed in Supplemental Table S15. Generally, the qRT-PCR results verified the induction of genes with statistically significant induction detected on the microarray, although the degree of induction was often underestimated by microarray measurement, as is often observed. By designing highly specific qRT-PCR primers, we were able to verify induction of multiple gene family members, and even tandem array members, in the PR-3 and PR-4 families. Members of a single array showed induction ranging from ~20-fold to 5,000-fold. Of the tested PR-10s, all verified the trend of equally strong induction by both pathogens or greater induction by *C. theobromicola*.

Discussion

The role of PR genes in mediating resistance to disease has been well studied in a wide variety of model and crop plant species (van Loon et al., 2006; Campos et al., 2007; Sels et al., 2008; Wanderley-Nogueira et al., 2012). These proteins are grouped together based on their increased accumulation in response to activation of systemic acquired resistance pathways and their roles in plant defense. Our analysis of the PR gene families of *T. cacao* resulted in the

Table 2.4 - Validation of PR gene induction by qRT-PCR. N.S. indicates p-value was not significant. Genes shown as induced by microarray had BH p-values < 0.05. Inductions detected by qRT-PCR were calculated using REST software (Pfaffl et al., 2002) and represent the average of five pathogen-treated samples compared to five water-treated samples relative to TcTub1 (Tc06_g000360). Transcripts were considered undetected if the average Ct value across all treatments was greater than 35.

		P. palmivora Treatment		*C. theobromicola* Treatment	
	Gene ID	Microarray Fold Induction	qRT-PCR Fold Induction	Microarray Fold Induction	qRT-PCR Fold Induction
PR-1s	Tc01_g003940	N.S.	Transcript not detected	N.S.	Transcript not detected
	Tc01_g034430	N.S.	N.S.	N.S.	N.S.
	Tc02_g002380	N.S.	N.S.	N.S.	N.S.
	Tc02_g002390	N.S.	N.S.	N.S.	N.S.
	Tc02_g002400	N.S.	N.S.	N.S.	8.3 (p=.001)
	Tc02_g002410	125.4	763 (p<.001)	91.3	55.7 (p<.001)
	Tc02_g002420	N.S.	N.S.	N.S.	N.S.
	Tc02_g002430	N.S.	N.S.	N.S.	N.S.
	Tc02_g010380	N.S.	N.S.	N.S.	N.S.
	Tc05_g005530	N.S.	N.S.	N.S.	N.S.
	Tc09_g000720	N.S.	N.S.	N.S.	N.S.
	Tc09_g016580	N.S.	N.S.	N.S.	N.S.
	Tc09_g016590	N.S.	N.S.	N.S.	N.S.
	Tc10_g000980	N.S.	Transcript not detected	N.S.	Transcript not detected
PR-3s	Tc01_g000770	33.6	70.2 (p < .001)	18.8	13.8 (p = .01)
	Tc02_g003890	27.1	Transcript not detected	22.31	Transcript not detected
	Tc04_g018100	22.5	5086.0 (p < .001)	8.3	36.7 (p = .019)
	Tc04_g018110	29.2	763.2 (p < .000)	11.5	13.7 (p = .041)
	Tc04_g018160	63.6	158.4 (p = .003)	73	65.6 (p = .001)
	Tc06_g000490	N.S.	3.4 (p = .016)	N.S	N.S.
PR-4s	Tc00_g012980	N.S.	N.S.	N.S.	N.S.
	Tc05_g027210	24.9	1027.7 (p < .001)	14.9	22.7 (p = .01)
	Tc05_g027220	11.1	258.9 (p = .001)	6.7	N.S.
	Tc05_g027230	N.S.	164.1 (p = .011)	N.S.	N.S.
	Tc05_g027250	N.S.	N.S.	N.S.	N.S.

	Tc05_g027320	53.4	29.3 (p = .009)	17.8	8.9 (p = .001)
	Tc10_g011130	N.S.	61.5 (p < .001)	N.S.	N.S.
PR-10s	Tc01_g031100	39.7	28.0 (p = .019)	57.2	32.3 (p = .002)
	Tc04_g028780	25.5	32.9 (p = .027)	25.5	41.6 (p = .004)
	Tc04_g028860	6.02	24.3 (p = .038)	53.4	96.8 (p = .001)

identification of multigene families for 15 families of PR proteins. These gene families include about 350 genes that are distributed throughout the genome. About 50% of the cacao PR genes are found in arrays of tandemly duplicated genes, and many family members, even within tandem arrays, exhibited varying levels of inducibility by pathogen treatment. The structure of the PR gene families of five other plant species shared these features with cacao, suggesting that PR tandem arrays are features highly conserved within most if not all higher plants. The high degree of correlation in family sizes suggests that similar evolutionary forces have likely acted on diverse plant genera, likely indicating that PR family expansions have been beneficial to land plant survival. This body of work provides strong evidence that gene duplication and neo-functionalization, particularly with regard to expression dynamics, have played major roles in shaping the genomics of the plant defense response.

Local duplications arise through various mechanisms including polymerase slippage, unequal crossing over, and transposon movement, and local duplications are known to contribute to eukaryotic evolution by increasing genetic diversity (Rizzon et al., 2006; Barker et al., 2012). Organization of PR genes into tandem arrays has been described for several plants and PR families, including PR-7s in tomato (Jordá et al., 1999), PR-10s in grape (Lebel et al., 2010), PR-12s in Arabidopsis (Silverstein et al., 2005), PR-1s in Arabidopsis and rice (van Loon et al., 2006), and PR-16s in rice (Manosalva et al., 2009). The physical clustering of PR-4s in cacao was also previously described (Pereira Menezes et al., 2014). Tandem duplications have also been shown to play a key role in evolution of Resistance (R) gene families (Leister, 2004) (Spoel and Dong, 2012) and they are particularly common in the NBS-LRR class of R genes, as well as in PR-1s, thaumatins, germins, and major latex proteins in Arabidopsis (Cannon et al., 2004). Here we demonstrate that this clustering is common across PR families. Correlation analysis of family size indicates that sizes are similar across diverse plant taxa, indicating that expanded family sizes are common and are likely selectively beneficial in higher plants. Our phylogenetic analysis

of the PR-1, PR-3, and PR-4 families suggests that the families have continually expanded both locally and inter-chromosomally over land plant evolution, although further investigation of expansions of certain sub-clades in different species is necessary to explain functional dynamics of family expansion.

Gene family expansions have a complicated interplay with expression dynamics. Employing our microarray analyses, we detected unique expression dynamics within groups of family members with very high percent identity. The data presented here suggest that in some cases single genes within tandem arrays are induced by a given pathogen, while in other tandem arrays two or more genes can be induced by the same stimulus. Large tandem arrays for PR-10s (Chr4PR-10.6, 15 members) and PR-16s (Chr5PR-16.3, 14 members) have members ranging from constitutive low expression to constitutive high expression, with a few showing inducibility by pathogens. Consequently, evolutionary dynamics of family members after a duplication event remain unclear, but several mechanisms are likely at play in a scenario-specific manner. First, selection could favor greater concentration of antimicrobial peptides produced in a given tissue, leading to multiple family members exhibiting similar protein structure and expression patterns. Our microarray analyses revealed several cases that could support this model; for example, four PR-3s that make up a tandem array were all induced by *P. palmivora*. Alternatively, mutations affecting nearby regulatory machinery or the coding sequence of the gene could result in new tissue specificity or binding/enzymatic activity of a protein. Our microarray dataset found that only one of six PR-1s in a tandem array was induced by pathogen, suggesting the others have alternative functions, tissue specificities, or are in the process of becoming pseudogenes. Evolutionary studies have revealed that products of small-scale duplications diverge in expression more rapidly than they do in terms of protein structure (Haberer et al., 2004), with age of paralogs correlating with their divergence in expression in Arabidopsis (Casneuf et al., 2006; Ganko et al., 2007) and rice (Li et al., 2009). For defense genes, divergence in expression patterns could be beneficial, decreasing metabolic burden associated with mounting a defense response in tissues distal to the site of infection. Further work, particularly RNA-seq experiments across a wide range of tissue types, would allow more comprehensive dissection of functional patterns associated with this gene organization. *In silico* promoter analysis may be a means of identifying a mechanism underlying expression dynamics of tandem arrays.

Teixeira *et al.* (Teixeira et al., 2014) previously reported the induction of more than 67 PR genes after infection of cacao plants with *Moniliophthora perniciosa*, but that the induction did not eliminate pathogen colonization. Similarly, the induction that we see here did not halt infection, but likely slowed the pathogens' progress. These transcriptomic experiments identify candidate genes that require functional characterization to better understand roles of PR proteins against the diversity of cacao's pathogens. The infection and microarray analysis we performed with oomycete (*P. palmivora*) and fungal (*C. theobromicola*) pathogens confirms the induction of 67 and 45 PR genes by the respective pathogen treatments, respectively. However, the majority of the PR genes had stable expression across treatments under our experimental conditions. Analysis of other tissues may reveal that a subset of those genes have tissue specificity in their basal expression and inducibility. The existence of PR family members with constitutively high expression could suggest that certain family members have evolved to act as a preliminary line of defense. For example, two PR-3s (Tc06_g000490 and Tc04_g029180) had very high expression in water treated samples. Constitutive high-level expression in leaves may allow the plant to begin degrading chitin of invading pathogens before PAMP or R-gene mediated signal transduction can elevate expression of induced defenses. Knockdown or deletion of these constitutive high-expressors followed by pathogen challenge resulting in increased susceptibility would demonstrate the role of basal defense components. Broadly, we saw a more dramatic defense response in samples infected with *P. palmivora* than in those infected with *C. theobromicola,* with more genes being up-regulated and their degree of induction being greater. The microarray and qRT-PCR analysis indicated that the PR-10 family deviates from this trend, with members showing equal or more dramatic induction by *C. theobromicola* than by *P. palmivora.* The PR-10 member Tc04_g028860 is particularly noteworthy, showing 96-fold induction by *C. theobromicola* treatment, about four times the induction by *P. palmivora* treatment. While it is possible that these differences reflect pathogen-specific responses, we cannot rule out the possibility that they result from different speeds with which the two pathogens colonize the host.

Induction of PR-1 genes is a hallmark of plant defense activation. While they belong to the well-studied Sperm Coating Protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS) group (Cantacessi et al., 2009), a sub-group of the Cysteine-rich secretory protein superfamily, little is known about

their biological function (Chalmers et al., 2008). Our analysis indicates that TcPR1-g (Tc10_g000980) that was previously reported to be induced in tissue infected with witches' broom (Teixeira et al., 2013), was not induced under our experimental conditions. This lack of induction by *P. palmivora* and *C. theobromicola* suggests that family member activation may differ for certain pathogens. Another example is the induction of the PR-1 Tc02_g002410, which was not induced by witches' broom, by *P. palmivora* and *C. theobromicola*. Our qRT-PCR experiment validated strong induction of only this gene (>700 fold by P. palmivora and > 50 fold by *C. theobromicola*), and confirmed low expression of Tc10_g000980 across all samples. The specificity of the reaction is interesting, but even more puzzling as the function of PR-1s in plants remains unclear.

PR-3 family member expression was also of particular interest because of our prior work with a class I chitinase (Tc02_g003890) (Maximova et al., 2006). Here we report induction of several other PR-3s. A tandem array on chromosome four (Chr4PR-3.4) was notable in that multiple members were found to be induced by both pathogens, suggesting that, in this case, proximity may be contributing to their co-expression, and that these proteins may act in a coordinated fashion to defend the plant against both of the tested pathogens. While chitin is significantly less abundant in the cell walls of oomycetes than fungi, and its function in oomycetes is not well understood, recent evidence suggests that chitin synthase enzymes are active in hyphal tips, where chitin may play a role in cell wall structure (Guerriero et al., 2010). Further, inhibition of these chitin synthases with nikkomycin Z led to bursting of hyphal tips and cell death. Accordingly, induction of chitinases in plants by oomycete treatment may reflect an important defense process, inhibition of hyphal tip growth.

Interestingly, our earlier work described that stable overexpression of Tc02_g003890, a class I chitinase, in transgenic cacao plants resulted in an increased resistance of leaves to *Colletotrichum gloeosporioides* (Maximova et al., 2006). The same gene was also upregulated in the highly disease-susceptible genotype ICS1 by treating leaves with salicylic acid (Fister et al., 2015), and we found that its transient overexpression in cacao leaves increases resistance to *P. capsici* (Fister et al., 2016). The qRT-PCR we performed here did not verify its induction by treatment with *P. palmivora* or *C. theobromicola*, suggesting that this gene may respond to SA but not these two pathogens. This result suggests that the underlying mechanisms of these

plant pathogen interactions are complex and that further research is necessary to unravel the specific mechanisms involved. One possibility is that the pathogens are able to suppress the mechanisms of SA induced gene expression via secretion of pathogen effector proteins as has been seen with other systems (Tanaka et al., 2015).

Cacao PR-4s were also recently identified (Pereira Menezes et al., 2014). Pereira-Menezes *et al.*'s (Pereira Menezes et al., 2014) work built upon an earlier EST database (Gesteira et al., 2007) by characterizing genotype specificity in the speed and level of induction of PR-4b (Tc05_g027210), which shows anti-fungal activity dependent on its RNase activity, in a resistant (TSH1188) and a susceptible (Catongo) genotype. Our microarray and qRT-PCR indicates that the gene was also induced by *P. palmivora* (more than 1000-fold and *C. theobromicola* (roughly 20-fold), showing one of the strongest inductions of the genes tested with qRT-PCR. Its induction by a variety of pathogens makes it a critical candidate for further study. Analyses similar to Pereira-Menezes *et al.*'s work across a broader background of genotypes are required to validate the importance of genes described here. Assaying the effect of over-expression or knockout of this gene would be useful for defining roles of single genes within these families.

We observed a few differences in organization when comparing two different varieties of cacao. The two varieties compared in this study are representatives of distinct genetic clusters that developed over *T. cacao*'s evolution and are thought to have diverged because of the presence of geological barriers (Motamayor et al., 2008). Consequently, it is possible that these two genotypes, having been subjected to different pathogens over their evolutionary history and having unique selective pressures applied by domestication after cultivation of cacao began, have undergone unique duplications or translocations altering gene organization. Indeed, our identification of PR genes in the two genomes may support this hypothesis, as gene counts within families differ for the two genomes, and while the positions of the genes are generally consistent, some chromosomal rearrangement appears to have occurred. It is possible however, that these are differences resulting from genome assembly strategies. Analysis of additional cacao genome sequences from other genetic groups (Motamayor et al., 2008) would help resolve these possibilities.

As induction of PR genes is a hallmark of the defense response in many plant species, their identification in cacao is critical to the study of cacao's defense response. Our finding that PR gene family size and organization into tandem arrays is consistent across diverse plant species suggests that the diverse expression patterns seen within families in other species are likely similar to those we have described in cacao. Therefore, this study lays a foundational knowledge of defense gene expression upon which functional molecular genetic approaches can be based. Genes identified here, once functionally verified, will be useful in breeding cacao cultivars with superior resistance to pathogens.

Conclusions

In this study we identified 359 PR genes in the cacao genome, and found that approximately half of these physically cluster into tandem arrays with other members of the same PR family. Physical clustering of PR genes into tandem arrays was also identified in five diverse plant species. Using a whole genome microarray and qRT-PCR to measure the induction of genes by two cacao pathogens, we identified which PR genes are induced in leaf tissue by pathogens, and we identified differences in basal expression within PR families. This work is critical in improving the understanding of the defense response in cacao, and it provides a list of key candidate defense genes that will be the focus of future molecular characterization.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Alvarez ME (2000) Salicylic acid in the machinery of hypersensitive cell death and disease resistance. Plant Molecular Biology **44:** 429-442

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci AM, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan MJ, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao L.* generated from various tissues and under various conditions. BMC Genomics **9:** 512

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JSS, Schuster SC, Carlson JE, Sallet E,

Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of *Theobroma cacao*. Nature Genetics **43:** 101-108

Argout X, Salse J, Aury J, Guiltinan M, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova S (2011) The genome of Theobroma cacao. Nat Genet **43**

Balasubramanian V, Vashisht D, Cletus J, Sakthivel N (2012) Plant β-1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. Biotechnology Letters **34:** 1983-1990

Barker M, Baute G, Liu S-L (2012) Duplications and Turnover in Plant Genomes. *In* JF Wendel, J Greilhuber, J Dolezel, IJ Leitch, eds, Plant Genome Diversity Volume 1. Springer Vienna, pp 155-169

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) **57:** 289-300

Brunner F, Stintzi A, Fritig B, Legrand M (1998) Substrate specificities of tobacco chitinases. The Plant Journal **14:** 225-234

Campos MA, Rosa DD, Teixeira JÉC, Targon MLPN, Souza AA, Paiva LV, Stach-Machado DR, Machado MA (2007) PR gene families of citrus: their organ specific-biotic and abiotic inducible expression profiles based on ESTs approach. Genetics and Molecular Biology **30:** 917-930

Cannon S, Mitra A, Baumgarten A, Young N, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biology **4:** 10

Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, Matthews JB, Loukas A, Hofmann A, Otranto D, Sternberg PW, Gasser RB (2009) A portrait of the "SCP/TAPS" proteins of eukaryotes — Developing a framework for fundamental research and biotechnological outcomes. Biotechnology Advances **27:** 376-388

Caporale C, Di Berardino I, Leonardi L, Bertini L, Cascone A, Buonocore V, Caruso C (2004) Wheat pathogenesis-related proteins of class 4 have ribonuclease activity. FEBS Letters **575:** 71-76

Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y (2006) Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. Genome Biology **7:** R13

Chalmers I, McArdle A, Coulson R, Wagner M, Schmid R, Hirai H, Hoffmann K (2008) Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the Schistosoma mansoni venom allergen-like (SmVAL) gene family. BMC Genomics **9:** 89

Christensen AB, Cho BHO, Næsby M, Gregersen PL, Brandt J, Madriz-Ordeñana K, Collinge DB, Thordal-Christensen H (2002) The molecular characterization of two barley proteins establishes the novel PR-17 family of pathogenesis-related proteins. Molecular Plant Pathology **3:** 135-144

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics **25:** 1422-1423

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defense responses to infection. Nature **411:** 826-833

Davidson RM, Reeves PA, Manosalva PM, Leach JE (2009) Germins: A diverse protein family important for crop improvement. Plant Science **177:** 499-510

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2012) Geneious v5.6, Available from http://www.geneious.com.

Dunwell JM, Gibbings JG, Mahmood T, Saqlan Naqvi SM (2008) Germin and Germin-like Proteins: Evolution, Structure, and Function. Critical Reviews in Plant Sciences **27:** 342-375

Durrant WE, Dong X (2004) Systemic acquired resistance. Annual Review of Phytopathology **42:** 185-209

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792-1797

Fister AS, O'Neil ST, Shi Z, Zhang Y, Tyler BM, Guiltinan MJ, Maximova SN (2015) Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. Journal of Experimental Botany **66:** 6245-6258

Fister AS, Shi Z, Zhang Y, Helliwell EE, Maximova SN, Guiltinan MJ (2016) Protocol: transient expression system for functional genomics in the tropical tree Theobroma cacao L. Plant Methods **12:** 1-13

Fu ZQ, Dong X (2013) Systemic Acquired Resistance: Turning Local Infection into Global Defense. Annual Review of Plant Biology **64:** 839-863

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in Arabidopsis. Molecular Biology and Evolution **24:** 2298-2309

Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biology **5:** R80

Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho KP, Schuster I, Macedo JN, Pereira GA, Cascardo JC (2007) Comparative analysis of expressed genes from cacao meristems infected by Moniliophthora perniciosa. Annals of Botany **100:** 129-140

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. (2007) Nature **449:** 463-467

Grover A (2012) Plant chitinases: genetic diversity and physiological roles. Critical Reviews in Plant Sciences **31:** 57-73

Guerriero G, Avino M, Zhou Q, Fugelstad J, Clergeot P-H, Bulone V (2010) Chitin synthases from *Saprolegnia* are involved in tip growth and represent a potential target for anti-oomycete drugs. PLoS Pathogens **6:** e1001070

Guest D (2007) Black pod: Diverse pathogens with a global impact on cocoa yield. Phytopathology **97:** 1650-1653

Guiltinan M, Verica J, Zhang D, Figueira A (2008) Genomics of Theobroma cacao, "The foods of the gods". *In* PM Moore, R Ming, eds, Genomics of tropical crop plants. Springer, New York

Haberer G, Hindemitt T, Meyers BC, Mayer KFX (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiology **136:** 3009-3022

International Brachypodium Inititative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763-768

Jones JDG, Dangl JL (2006) The plant immune system. Nature **444:** 323-329

Jordá L, Coego A, Conejero V, Vera P (1999) A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. Journal of Biological Chemistry **274:** 2360-2365

Kliebenstein DJ (2014) Orchestration of plant defense systems: genes to populations. Trends in Plant Science **19:** 250-255

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research **40:** D1202-D1210

Lebel S, Schellenbaum P, Walter B, Maillot P (2010) Characterisation of the *Vitis vinifera* PR10 multigene family. BMC Plant Biology **10:** 184

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics **20:** 116-122

Li Z, Zhang H, Ge S, Gu X, Gao G, Luo J (2009) Expression pattern divergence of duplicated genes in rice. BMC Bioinformatics **10:** S8

Liu J-J, Sturrock R, Ekramoddoullah AM (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Reports **29:** 419-436

Lu H-C, Lin J-H, Chua ACN, Chung T-Y, Tsai IC, Tzen JTC, Chou W-M (2012) Cloning and expression of pathogenesis-related protein 4 from jelly fig (*Ficus awkeotsang Makino*) achenes associated with ribonuclease, chitinase and anti-fungal activities. Plant Physiology and Biochemistry **56:** 1-13

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. The Plant Journal **53:** 661-673

Macho Alberto P, Zipfel C (2014) Plant PRRs and the activation of innate immune signaling. Molecular Cell **54:** 263-272

Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, Leung H, Leach JE (2009) A Germin-Like Protein Gene Family Functions as a Complex Quantitative Trait Locus Conferring Broad-Spectrum Disease Resistance in Rice. Plant Physiology **149:** 286-296

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. Nucleic Acids Research **43:** D222-D226

Maximova S, Florez S, Shen X, Niemenak N, Zhang Y, Curtis W, Guiltinan M (2014) Genome-wide analysis reveals divergent patterns of gene expression during zygotic and somatic

embryo maturation of *Theobroma cacao L*., the chocolate tree. BMC Plant Biology **14:** 185

Maximova S, Marelli J-P, Young A, Pishak S, Verica J, Guiltinan M (2006) Over-expression of a cacao class I chitinase gene in *Theobroma cacao L*. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224:** 740-749

Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S (2012) Expression of designed antimicrobial peptides in Theobroma cacao L. trees reduces leaf necrosis caused by Phytophthora spp. Small Wonders Pept Dis Control **1905**

Mejía LC, Herre EA, Sparks JP, Winter K, García MN, Van Bael SA, Stitt J, Shi Z, Zhang Y, Guiltinan MJ, Maximova SN (2014) Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5**

Mejía LC, Rojas EI, Maynard Z, Van Bael S, Arnold AE, Hebbar P, Samuels GJ, Robbins N, Herre EA (2008) Endophytic fungi as biocontrol agents of *Theobroma cacao* pathogens. Biological Control **46:** 4-14

Menezes SP, dos Santos JL, Cardoso THS, Pirovani CP, Micheli F, Noronha FSM, Alves AC, Faria AMC, da Silva Gesteira A (2012) Evaluation of the allergenicity potential of TcPR-10 protein from *Theobroma cacao*. PLoS ONE **7:** e37969

Mithöfer A, Boland W (2012) Plant defense against herbivores: chemical aspects. Annual Review of Plant Biology **63:** 431-450

Motamayor J, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley S, Zheng P, Utro F, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler B, Stack J, Feltus F, Mustiga G, Amores F, Phillips W, Marelli JP, May G, Shapiro H, Ma J, Bustamante C, Schnell R, Main D, Gilbert D, Parida L, Kuhn D (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology **14:** R53

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao L*). PLoS ONE **3:** e3311

Motamayor JC, Lachenaud P, Silva E Mota JW, Loor R, Kuhn DN, Brown JS (2008) Geographic and genetic population differentiation of the amazonian chocolate tree (Theobroma cacao L.). PLoS ONE **3**

Park C-J, Kim K-J, Shin R, Park JM, Shin Y-C, Paek K-H (2004) Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. The Plant Journal **37:** 186-198

Passardi F, Penel C, Dunand C (2004) Performing the paradoxical: how plant peroxidases modify the cell wall. Trends in Plant Science **9:** 534-540

Pereira Menezes S, de Andrade Silva E, Matos Lima E, Oliveira de Sousa A, Silva Andrade B, Santos Lima Lemos L, Peres Gramacho K, da Silva Gesteira A, Pirovani C, Micheli F (2014) The pathogenesis-related protein PR-4b from *Theobroma cacao* presents RNase activity, Ca2+ and Mg2+ dependent-DNase activity and antifungal action on *Moniliophthora perniciosa*. BMC Plant Biology **14:** 161

Petre B, Major I, Rouhier N, Duplessis S (2011) Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar. BMC Plant Biology **11:** 33

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research **30:** e36

Ploetz RC (2007) Cacao diseases: Important threats to chocolate production worldwide. Phytopathology **97:** 1634-1639

Pungartnik C, da Silva AC, de Melo SA, Gramacho KP, de Mattos Cascardo JC, Brendel M, Micheli F, da Silva Gesteira A (2008) High-affinity copper transport and Snq2 export permease of *Saccharomyces cerevisiae* modulate cytotoxicity of PR-10 from *Theobroma cacao*. Molecular Plant-Microbe Interactions **22:** 39-51

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol **2:** e115

Rojas EI, Rehner SA, Samuels GJ, Van Bael SA, Herre EA, Cannon P, Chen R, Pang J, Wang R, Zhang Y (2010) *Colletotrichum gloeosporioides sl* associated with *Theobroma cacao* and other plants in Panama: multilocus phylogenies distinguish host-associated pathogens from asymptomatic endophytes. Mycologia **102:** 1318-1338

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC (2008) Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiology and Biochemistry **46:** 941-950

Silverstein KAT, Graham MA, Paape TD, VandenBosch KA (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiology **138:** 600-610

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. Genome Research **19:** 1630-1638

Smyth G (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology **3**

Smyth GK (2005) limma: Linear Models for Microarray Data. *In* R Gentleman, V Carey, W Huber, R Irizarry, S Dudoit, eds, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer New York, pp 397-420

Spoel SH, Dong X (2012) How do plants achieve immunity? Defense without specialized immune cells. Nature Reviews Immunolology **12:** 89-100

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics

Stec B (2006) Plant thionins – the structural perspective. Cellular and Molecular Life Sciences CMLS **63:** 1370-1385

Stotz HU, Thomson JG, Wang Y (2009) Plant defensins: Defense, development and application. Plant Signaling & Behavior **4:** 1010-1012

Tanaka S, Han X, Kahmann R (2015) Microbial effectors target multiple steps in the salicylic acid production and signaling pathway. Frontiers in Plant Science **6:** 349

Teixeira PJPL, Costa GGL, Fiorin GL, Pereira GAG, Mondego JMC (2013) Novel receptor-like kinases in cacao contain PR-1 extracellular domains. Molecular Plant Pathology **14:** 602-609

Teixeira PJPL, Thomazella DPdT, Reis O, do Prado PFV, do Rio MCS, Fiorin GL, José J, Costa GGL, Negri VA, Mondego JMC, Mieczkowski P, Pereira GAG (2014) High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. The Plant Cell **26:** 4245-4269

Terwisscha van Scheltinga AC, Hennig M, Dijkstra BW (1996) The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and

structure motifs of glycosyl hydrolase family 18. Journal of Molecular Biology **262:** 243-257

Tornero P, Conejero V, Vera P (1996) Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases. Proceedings of the National Academy of Sciences of the United States of America **93:** 6332-6337

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology **44:** 135-162

van Loon LC, van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiological and Molecular Plant Pathology **55:** 85-97

Verica JA, Maximova SN, Strem MD, Carlson JE, Bailey BA, Guiltinan MJ (2004) Isolation of ESTs from cacao (*Theobroma cacao L.*) leaves treated with inducers of the defense response. Plant Cell Reports **23:** 404-413

Vlot AC, Dempsey DMA, Klessig DF (2009) Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47:** 177-206

Walter MH, Liu J-W, Wünn J, Hess D (1996) Bean ribonuclease-like pathogenesis-related protein genes (Ypr10) display complex patterns of developmental, dark-induced and exogenous-stimulus-dependent expression. European Journal of Biochemistry **239:** 281-293

Wanderley-Nogueira AC, Belarmino LC, Soares-Cavalcanti NdM, Bezerra-Neto JP, Kido EA, Pandolfi V, Abdelnoor RV, Binneck E, Carazzole MF, Benko-Iseppon AM (2012) An overall evaluation of the resistance (R) and pathogenesis-related (PR) superfamilies in soybean, as compared with *Medicago* and Arabidopsis. Genetics and Molecular Biology **35:** 260-271

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, New York

Yang Y-X, Ahammed GJ, Wu C, Fan S-y, Zhou Y-H (2015) Crosstalk among jasmonate, salicylate and ethylene signaling pathways in plant disease and immune responses. Current Protein & Peptide Science **16:** 450-461

Yeats TH, Rose JKC (2008) The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). Protein Science : A Publication of the Protein Society **17:** 191-198

Yin T, Cook D, Lawrence M (2012) ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biology **13:** R77

Yu J, Hu S, Wang J, Wong G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science **296:** 79 - 92

Zhang Y, Smith P, Maximova SN, Guiltinan MJ (2015) Application of glycerol as a foliar spray activates the defense response and enhances disease resistance of *Theobroma cacao*. Molecular Plant Pathology **16:** 27-37

Chapter 3:

Two *Theobroma cacao* Genotypes with Contrasting Pathogen Tolerance Show Aberrant Transcriptional and ROS Responses after Salicylic Acid Treatment

Published as:
Fister, A. S., O'Neil, S. T., Shi, Z., Zhang, Y., Tyler, B. M., Guiltinan, M. J., & Maximova, S. N. (2015). Two Theobroma cacao genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. *Journal of experimental botany*, erv334.

Abstract

Understanding the genetic basis of pathogen susceptibility in various crop plants is crucial to increasing the stability of food, feed, and fuel production. Varietal differences in defense responses provide insights into the mechanisms of resistance and are a key resource for plant breeders. To explore the role of salicylic acid in the regulation of defense in cacao, we demonstrated that SA treatment decreased susceptibility to a pod rot pathogen, *Phytophthora tropicalis* in two genotypes, Scavina 6 and Imperial College Selection 1, which differ in their resistance to several agriculturally important pathogens. Transient overexpression of TcNPR1, a major transcriptional regulator of the SA-dependent plant immune system, also increased pathogen tolerance in cacao leaves. To explore further the genetic basis of resistance in cacao, we used microarrays to measure gene expression profiles after salicylic acid (SA) treatment in these two cacao genotypes. The two genotypes displayed distinct transcriptional responses to SA. Unexpectedly, the expression profile of the susceptible genotype ICS1 included a larger number of pathogenesis-related genes that were induced by SA at 24 h after treatment, whereas genes encoding many chloroplast and mitochondrial proteins implicated in reactive oxygen species production were up-regulated in the resistant genotype, Sca6. Sca6 accumulated significantly more superoxide at 24 h after treatment of leaves with SA. These experiments

revealed critical insights regarding the molecular differences between cacao varieties, which will allow a better understanding of defense mechanisms to help guide breeding programmes.

Introduction

Theobroma cacao (cacao), the seeds of which are used to make chocolate, is an economically important crop providing income to small-scale farmers in tropical regions all over the world (Wood and Lass, 2008). However, ~30–40% of annual cacao production is lost to pathogens due to its very high disease susceptibility (Hebbar, 2007; Argout *et al.*, 2008). Cacao is the host to several diseases including witches' broom disease (WBD), caused by *Moniliophthora perniciosa* (Purdy and Schmidt, 1996), frosty pod rot caused by *Moniliophthora roreri* (Phillips-Mora and Wilkinson, 2007), and black pod rot caused by several *Phytophthora* species (Bailey *et al.*, 2005*a*). Two genotypes of cacao, Scavina 6 (Sca6) and Imperial College Selection 1 (ICS1), are of special importance to the study of cacao disease resistance because they differ in their tolerance to the above-mentioned pathogens; Sca6 is a more resistant variety and ICS1 is highly susceptible (Yamada and Lopes, 1999; Brown *et al.*, 2005; Faleiro *et al.*, 2006). Several quantitative trait loci (QTLs) have been mapped for resistance to WBD and black pod rot in Sca6; however, the mechanistic differences underlying the variation in susceptibility between these two varieties are still unclear (Risterucci *et al.*, 2003; Brown *et al.*, 2005; Faleiro *et al.*, 2006). A fuller understanding of the genes associated with susceptible and resistance responses would be extremely useful for cacao breeding programmes and the selection of new varieties with higher resistance.

Salicylic acid (SA) is considered to be the most important signalling hormone controlling the responses to biotrophic and hemibiotrophic pathogens in other plant species (Shah, 2003; Durrant and Dong, 2004; Loake and Grant, 2007; Vlot *et al.*, 2008; Fu and Dong, 2013). Hundreds of genes induced by SA have been isolated and characterized in the model plant *Arabidopsis thaliana* (Arabidopsis) (Cao *et al.*, 1994; Dong, 2004; Uquillas *et al.*, 2004; Grant and Lamb, 2006; Wang *et al.*, 2006; Lee *et al.*, 2007, 2009; Loake and Grant, 2007; Attaran *et al.*, 2009). In SA-treated unripe pepper fruit, 177 of 7900 cDNA clones exhibited more than 4-fold transcript accumulation (Lee *et al.*, 2009). In rice, microarray analysis identified SA-inducible WRKY

transcription factors involved in rice blast resistance (Shimono *et al.*, 2007). Aside from its role in pathogenesis-related (PR) gene induction, SA is also involved in the oxidative burst and hypersensitive response during pathogen attack (Alvarez, 2000; Torres *et al.*, 2006). SA synthesis and reactive oxygen species (ROS) production are believed to act in a positive feedback loop with each other and together lead to induction of programmed cell death (Overmyer *et al.*, 2003). The oxidative burst and cell death are known to involve ROS production both in the apoplast as well as in intracellular compartments, including mitochondria and chloroplasts (Vlot *et al.*, 2009; O'Brien *et al.*, 2012).

Recent studies have indicated that proteins in the non-expressor of pathogenesis-related (NPR) family are the receptors for SA (Fu *et al.*, 2012; Wu *et al.*, 2012). Two groups have demonstrated that NPR3 and NPR4 are members of a receptor complex for NPR1, mediating an interaction between it and CUL3 E3 ligase (Fu *et al.*, 2012), and that cysteine residues in NPR1 are necessary for direct interaction between the protein and SA (Wu *et al.*, 2012). Taken together, these results suggest that there may be some partial redundancy within this family enabling each to interact directly with SA. Even without its potential role as a direct receptor for SA, the importance of NPR1 in regulating the transcriptional changes associated with systemic acquired resistance has been a highly active area of research (Fu and Dong, 2013).

Recent evidence suggests that *T. cacao* also uses the SA-dependent pathway during defense responses (Borrone *et al.*, 2004; Bailey *et al.*, 2005*a*, *b*; Maximova *et al.*, 2006; Gesteira *et al.*, 2007), and PR genes are up-regulated in leaves after treatment with the SA analogue BTH (Verica *et al.*, 2004). Moreover, genes encoding cacao homologues of NPR1 (Tc09_g007660) and NPR3 (Tc06_g011480) can partially restore the Arabidopsis *npr1* and *npr3* mutant phenotypes, demonstrating the highly conserved nature of this signalling pathway (Shi *et al.*, 2010, 2013). In contrast, the transcriptional responses of *Theobroma cacao* cultivar 'Comum' to infection by WBD did not include significant changes in the transcription of genes in the SA pathway although there was activation of a variety of other genes implicated in defense responses and repression of photosynthesis (Teixeira *et al.*, 2014), implying that the SA pathway may not be the predominant mechanism of response to this particular pathogen or in this cultivar. To explore the mechanisms potentially responsible for genotype-specific differences in defense responses in cacao, we used a custom cacao microarray to evaluate differential gene expression in cacao leaves in response to SA treatment in the Sca6 and ICS1 genotypes. Our results

uncovered distinct differences between the two genotypes in accumulation of ROS in SA treated leaves that were consistent with specific differences in gene expression, suggesting that these mechanisms may play key roles in determining disease susceptibility in cacao.

Materials and methods

Leaf disk pathogen bioassay using cacao leaves

Sca6 genotype is known to be more resistant to a number of pathogens, and genotype ICS1 is considered to be highly susceptible (Risterucci *et al.*, 2003; Faleiro *et al.*, 2006). Thus we utilized these two genotypes to study the molecular mechanisms of defense response in cacao. A leaf inoculation assay was performed with *Phytophthora tropicalis* to verify and quantify the differences between ICS1 and Sca6 in their response to treatment with SA. Fully-expanded, light green, and supple leaves at developmental stage C (Mejia *et al.*, 2012) on greenhouse-grown trees of both genotypes were treated with 1 mM SA or water (as a control). Twenty-four hours after treatment, the leaves were harvested from the plants and inoculated with mycelial plugs of *Phytophthora tropicalis* as previously described (Mejia *et al.*, 2012). Eight leaf pieces from each genotype and each treatment were inoculated and photographs were taken 72 h post-inoculation with a 1/30s exposure time, aperture of f=5.6, using a Nikon D90 equipped with a Nikon AF-S NIKKOR DX 18–135 mm lens. Lesion sizes were measured using ImageJ. Average lesion sizes were calculated from 24 replicates and significance was determined by single factor ANOVA. As a complementary measurement of pathogen virulence, the relative amount of pathogen DNA was measured by determining the ratio of *Phytophthora* DNA to cacao DNA in infection zones by qPCR. Lesions were collected using a 2 cm diameter cork borer surrounding the inoculation site and genomic DNA was extracted using a Tissumizer (Tekmar, Mason, Ohio, USA) and DNeasy plant mini kit (Qiagen). Specific primers for *P. tropicalis Actin* (F: GACAACGGCTCCGGTATGTGCAAGG and R: GTCAGCACACCACGCTTGGACTG) and cacao *Actin*7 (Tc01g010900) (F: AGGTGGAGATCATTGAAGGAGGGT and R: ACCAGCGGTCATCACAAGTCACAA) genes were used as pathogen and host targets. qPCR was performed using an ABI 7300 (Applied Biosystems, Foster City, CA, USA) as previously described (Shi *et al.*, 2013). Differences between

genotypes and treatments were identified using Fisher's partial least-squares difference analysis.

Transient *Agrobacterium*-mediated transformation of cacao leaves

To create a T-DNA binary vector for overexpression of the *TcNPR1* coding sequence, plasmid pGZ12.0106 (GenBank: KP844566) was digested with restriction enzymes SpeI and HpaI and then was ligated to a DNA fragment containing the *TcNPR1* coding sequence isolated as a Spe I-Pvu II restriction fragment generated by digesting plasmid pGEM-TcNPR1 (Shi *et al.*, 2010) resulting in pGS12.0224 (GenBank: KP844565). The T-DNA region of the vector contains the modified CaMV-35S derivative, E12-Ω promoter (Mitsuhara *et al.*, 1996), which drives *TcNPR1, EGFP* (Clontech)*,* and *NPTII-A* (De Block *et al.*, 1984) transgenes, and these are followed by the 35S terminator. A second copy of the *NPTII* marker gene (*NPTII-B*), is flanked by the NOS promoter and terminator (Lichtenstein and Fuller, 1987). The vector map for pGS12.0224 was created in Geneious (Drummond *et al.*, 2012), and is shown in Supplementary Fig. S1. The pGS12.0224 vector and the control vector (pGH00.0126, GenBank: KF018696) were used to transiently transform cacao leaf tissue using *Agrobacterium tumefaciens* vacuum infiltration as previously described (Shi *et al.*, 2013). Forty-eight hours after infiltration, leaves were screened with a fluorescent stereo-microscope equipped with an EGFP filter system as previously described (Maximova *et al.*, 1998). Leaves exhibiting green fluorescence over 90% of their area were used in *P. tropicalis* infection assays. Inoculation was performed as described above. Disease impact was determined using lesion size analysis and qPCR as described above. Tissue surrounding the lesions was collected and used for RNA extractions and subsequent qRT-PCR to verify transgene expression. RNA from each sample were isolated as previously described (Verica *et al.*, 2004). qRT-PCR was performed using the Taqman ABI 7300 Sequence Detection System (Applied Biosystems Inc, Foster City, CA, USA). Primer and probe sequences for qRT-PCR were: *TcNPR1*: 5'—GTCACGTGCTGTCTGACCTTGT, 3'—TCACAGTTCATAATCTGGTCGAGC, Probe—TYCCGCGCTGTTCGGCAGT; *TcActin*: 5'—GATTCAGATGCCCAGAAGTCTTG, 3'—TCTCGTGGATTCCAGCAGCT, Probe—CCAGCCCTCGTTGTGGGAAAGG; *TcUbiquitin*: 5' —AGGCCTCAACTGGTTGCTGT, 3'—ACCGGCAAGACCATCACTCT, Probe—CGAGAGCAGCGACACCCATCGACA. The qRT-PCR normalization and analysis were performed using REST software (Pfaffl *et al.*, 2002) as previously described (Mejía *et al.*, 2014).

Plant treatment and RNA extraction

Three- to four-month-old cacao rooted cuttings of genotypes ICS1 and Sca6 were treated with 2 mM SA dissolved in water under greenhouse conditions as previously described (Swanson *et al.*, 2008). Plants treated with only water served as negative controls. Twenty-four hours after treatment, leaf samples from different developmental stages A, C and E (Mejia *et al.*, 2012) were collected. Three biological replicates were collected for each genotype. Thus, 36 samples were collected in total. RNAs from each sample were isolated as previously described (Verica *et al.*, 2004).

Microarray analysis

Roche Nimblegen oligonucleotide custom *T. cacao* gene expression 4×72 k (four arrays of 72 000 probes) were manufactured (*T. cacao* 17K microarray, design ID 7114 manufactured by Roche). Each array contained four probes of 50–60-mers in length for each of 17 247 unigenes. Three biological replicates were collected for each genotype, each treatment and each developmental stage, except for Stage A Sca6 and Stage E ICS1, both of which had only two replicates per treatment. Array design, RNA extraction protocol, hybridization procedures, scanning and data normalization protocols are described in Mejía *et al.* (2014). Data from the microarray experiment are available at the NCBI Gene Expression Omnibus (GEO: GPL18260).

Statistical analysis of microarray results

After $\log_2$ transformation of expression data, probe sets with mean $\log_2$ expression less than the background level of 8.0 were removed. Differential expression between genotypes on per-treatment and per-treatment+per-leaf-stage bases were assessed by general linear hypothesis (GLH) tests via the R multcomp package version 1.3 (Bretz *et al.*, 2010). Differential expression between treatments was assessed on a per-genotype+per-leaf-stage basis with two-sided student's *t*-tests on expression differences (testing whether mean differential expression equaled 0) between replicates paired by sampling day per group. Differential expression between treatments overall and on per-genotype bases were assessed with GLH tests. On the microarrays, a number of genes were represented by additional probe sets generated from 3' UTR regions. For those genes with multiple probe sets that passed expression cutoffs, a mixed linear model treating probe set as a random factor was used for the GLH tests rather than a simple linear model. All *P*-values were adjusted using the Benjamini-Hochberg procedure on a per-test basis (Benjamini and Hochberg, 1995).

Identification of *T. cacao* PR genes

PR genes in the *T. cacao* Criollo genome were identified according to the protocol described in Campos *et al.* (2007). The amino acid sequence for each PR gene type member was compared to a database of cacao polypeptide sequences using BLASTp with an e-value cutoff of $e<10^{-5}$.

GO enrichment analysis of microarray data

The probe set on the microarray was annotated with best hits from Blast searches of the Arabidopsis genome as previously described (Mejía *et al.*, 2014). All genes with available *A. thaliana* loci accession numbers were classified according to gene ontology (GO) terms using the tools for GO annotations at The Arabidopsis Information Resource (http://arabidopsis.org/tools/bulk/go/index.jsp). Gene Ontology enrichment analysis was performed using the Parametric Analysis of Gene Enrichment (Kim and Volsky, 2005) module on agriGO (http://bioinfo.cau.edu.cn/agriGO/) (Du *et al.*, 2010). For any given comparison, all genes on the microarray with annotated Arabidopsis best hits and statistically significant (Benjamini-Hochberg $P<0.05$) differential regulation were included in the analysis.

qRT-PCR measurement of selected genes

qRT-PCR analyses were performed on the same RNA samples produced for the microarray experiment. One microgram of RNA from each sample was reverse transcribed by M MuLV Reverse Transcriptase (New England Biolabs) using oligo-(dT)15 as a primer to generate cDNA. qRT-PCR was performed in 10 µl reactions, consisting of 5 µl of SYBR Green PCR Master Mix (Takara), 0.2 µl of Rox, and 0.4 µl of each primer, diluted to 5 µM, and 4 µl of cDNA. Each reaction was performed in technical duplicate using the Applied Biosystem Step One Plus Realtime PCR System (Roche) with the following programme: 15 min at 94°C, 40 cycles of 15 s at 94°C, 20 s at 60°C and 40 s at 72°C. The specificities of the primer pairs were examined using a dissociation curve and by visualization on 2% agarose gels. A cacao *Actin* gene was used as a reference (Shi *et al.*, 2010), and fold change was calculated using the $\Delta\Delta C_T$ method (Livak and Schmittgen, 2001).

Attached-leaf SA treatment peroxide and superoxide staining

To test whether accumulation of ROS differed between the two genotypes, we performed nitroblue tetrazolium (NBT) and 3,3'-diaminobenzidine (DAB) staining to quantify the accumulation of superoxide and peroxide, respectively. Greenhouse-grown mature (flowering) ICS1 and Sca6 trees were sprayed with 2 mM SA or with water as a control. Twenty-four hours after treatment, stage C leaves were removed from the trees, discs were punched out using a 1.5 cm diameter cork borer and randomized into groups of three for infiltration with either 1% NBT solution in 10 mM potassium phosphate buffer or 1 mg/ml DAB following a published protocol (Daudi *et al.*, 2012). Leaves were vacuum infiltrated for 5 min three times, with pressure reaching −23 in.-Hg. After infiltration, NBT-treated discs were bathed in NBT solution in darkness for 2 h on an orbital shaker (60 rpm) and DAB-treated samples were bathed in DAB solution in darkness for 8 h on an orbital shaker (60 rpm). After incubation, chlorophyll was bleached by soaking leaf discs in a 3:1:1 ethanol, glycerol, acetic acid mixture for 45 min with periodic vortexing for 5 s. Leaf discs were placed back in their original petri dishes, but sandwiched between the lid and the base to flatten them. Leaf discs were photographed as described above. Using ImageJ, NBT-stained areas were calculated by selecting blue tissue using a colour threshold, passing red=0–160, green=0–160, and blue=0–255. Area of staining was averaged across the three discs on a plate, and to reflect darker stains with a larger number, this value was multiplied by 255 minus the mean grey value within the area selected by the colour threshold. DAB staining also was detected by using a colour threshold to select brown tissue, passing red=0–160, green=0–150 and blue=0–90. The area of staining for each biological replicate was multiplied by 255 minus the measured mean grey value. Differences between genotypes and treatments were identified using Fisher's partial least-squares difference analysis.

Results

SA treatment enhances resistance to *P. tropicalis* in both genotypes

At the time of collection, the ICS1 and Sca6 leaves were very similar in appearance and texture, however at the end of the 24 h incubation, most of the leaf surfaces of ICS1 became chlorotic, while Sca6 leaves remained green (Fig. 1A–D). This difference was apparent until the end of the pathogen infection period. By 3 d post-inoculation, all SA and water treated leaves of both genotypes developed necrotic lesions at the sites of infection (Fig. 1A–D). Lesion areas were measured using ImageJ software (Schneider *et al.*, 2012). Symptoms were most severe in the water-treated ICS1 leaves. The average lesion areas of SA-treated ICS1 leaves was 20% smaller compared to the water treated ICS1 leaves, though the difference was not significant (Fig. 1E). Treatment of Sca6 leaves with SA resulted in a statistically significant reduction of lesion areas. The lesions of SA treated Sca6 leaves were ~60% smaller than those of Sca6 control leaves and 80% smaller than ICS1 controls. To assess pathogen growth we extracted genomic DNA from the lesions and performed qPCR with *P. tropicalis*—specific primers and primers for the cacao actin gene. SA treatment significantly ($P<0.05$) reduced pathogen growth in both genotypes (Fig. 1F). Pathogen biomass in control Sca6 leaves was similar to SA-treated ICS1, but was further reduced in SA-treated Sca6 tissue. These results confirmed that detached leaves of ICS1 are more susceptible to *P. tropicalis* than Sca6 at both basal and SA-induced states. Both genotypes demonstrated an SA response that resulted in decreased lesion size and pathogen growth, however, the SA effect was greater in Sca6 (4-fold reduction in pathogen biomass accumulation) than in ICS1 (1.7-fold) (Fig. 1E, F).

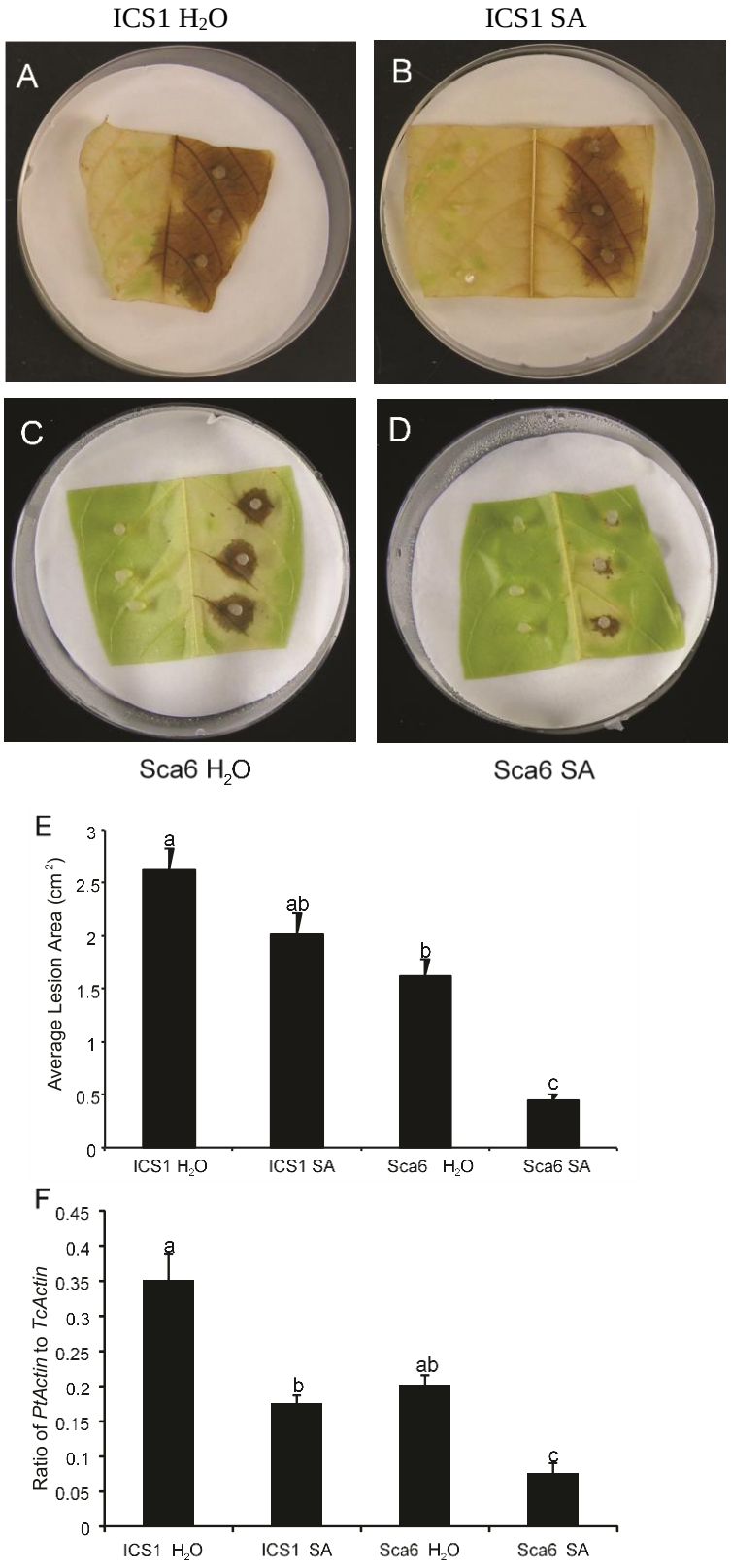


Figure 3.1 - Inoculation of Salicylic Acid (SA) pre-treated stage C leaves from ICS1 and Sca6 with *Phytophthora tropicalis*. Stage C leaves were inoculated with agar plugs containing *P. tropicalis* mycelium 24 hours after water or 1 mM SA treatment. Representative images of (**A**) water-treated ICS1 leaves, (**B**) SA-treated ICS1 leaves, (**C**) water-treated Sca6 leaves, and (**D**) SA-treated Sca6 leaves three days after inoculation. Scale bars represent 1 cm. **E.** Average lesion areas in replicate leaves were evaluated by ImageJ. Data represent means ± SE of treated leaves from 24 replicates per genotype. Letters above the bar chart show the significant differences ($p<0.05$) determined by Fisher's PLSD analysis. **F.** Relative pathogen biomass was measured by qPCR with DNA isolated 48 hrs after inoculation and is expressed as the ratio of *P. tropicalis actin* to cacao *actin*. Bars represent means ± SE of four biological replicates, each with three technical replicates. Letters above the bar show the significant differences ($p<0.05$) determined by Fisher's PLSD analysis.

Transient overexpression of TcNPR1 enhances resistance to *P. tropicalis*

The NPR1 protein is a key regulator of systemic acquired resistance in plants (Fu and Dong, 2013) and our group has previously demonstrated that expression of cacao TcNPR1 was able to partially restore the phenotype of an Arabidopsis *npr1* mutant (Shi *et al.*, 2010). To examine further the involvement of the systemic acquired resistance pathway in the defense response in cacao, we employed transient transformation of cacao Sca6 leaves followed by a pathogen infection assay (Mejia *et al.*, 2012; Shi *et al.*, 2013). Three days after *P. tropicalis* infection, lesions had formed on both control and *TcNPR1* transgenic leaf sections (Fig. 2A). Using qRT-PCR, we demonstrated that expression of *TcNPR1* was increased ~3-fold compared to tissue transformed with a vector control lacking the *TcNPR1* transgene (Fig. 2B). Average lesion area was significantly smaller in *TcNPR1* overexpressing leaves compared to the control vector (Fig. 2C). Additional quantification of *P. tropicalis* DNA in the lesions by qRT-PCR demonstrated that the growth of the pathogen was also significantly reduced in the transformed tissue (Fig. 2D). These results indicated that *TcNPR1* overexpression results in increased pathogen resistance in our *in vitro* assay and further implicates the SA pathway as a major mechanism of resistance in cacao.

Differential gene expression detected by microarray analysis

To study the responses of the two contrasting genotypes (Sca6 and ICS1) to SA treatment, we used a microarray to measure the transcript levels of over 17 000 cacao genes. By using a general linear model to assess expression differences across leaf stages, 436 and 601 genes were identified as being up- and down-regulated, respectively, in ICS1 (Benjamini-Hochberg (BH) adjusted $P<0.05$) in response to SA (Supplementary Table S1). Although a number of these genes had very low fold changes, they were statistically significant and thus were included in our subsequent analysis. In the Sca6 genotype, 490 and 447 (Supplementary Table S2) genes were detected as up- and down-regulated, respectively (BH-adjusted $P<0.05$) (Table 1). Of all significant genes regulated, 234 genes (Supplementary Table S3) had statistically significant differential regulation in both genotypes. The effect of SA on expression of these genes is plotted in Fig. 3, demonstrating that while many of the genes (173) were regulated in a

consistent manner between the genotypes (up- or down-regulated in both genotypes) others (61) responded differently between the genotypes.

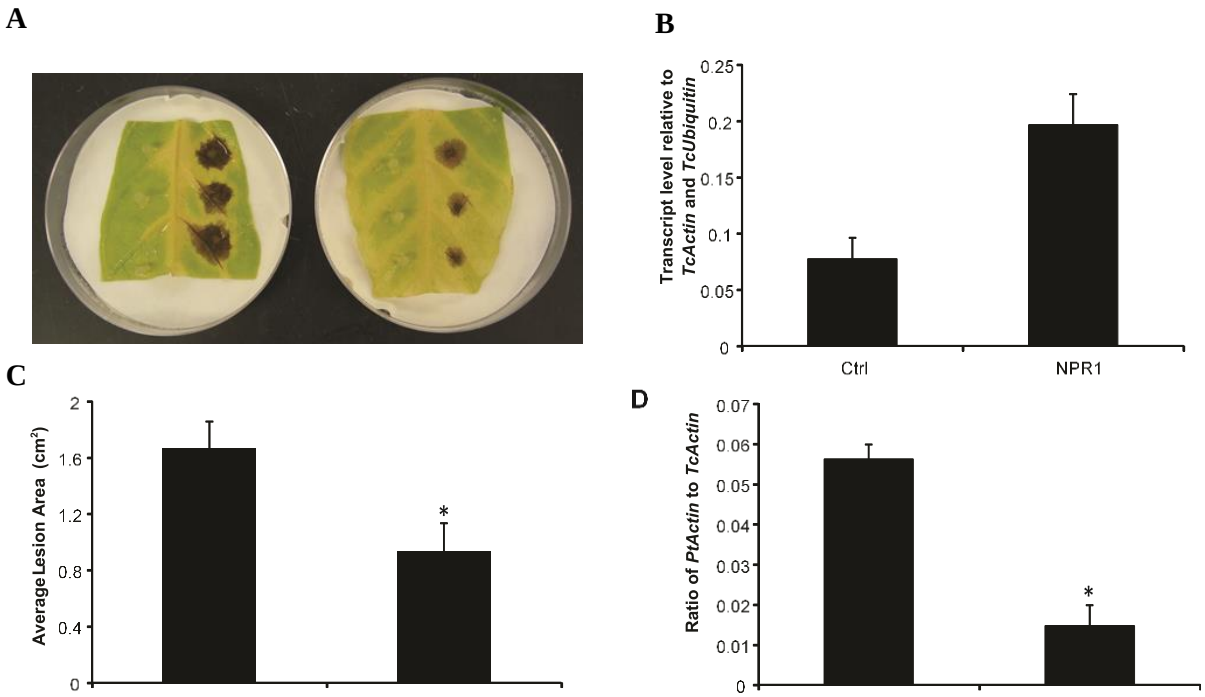


Figure 3.2 - Functional analysis of *TcNPR1*. **A.** Representative images of lesions from control and *TcNPR1* transiently transformed leaves two days after *Phytophthora tropicalis* inoculation. **B.** qRT-PCR analysis of *TcNPR1* transcript two days after vacuum infiltration (Control –Ctrl; *TcNPR1* expression – NPR1). Data represent means ± SE of three biological replicates. **C.** Average lesion areas from control and *TcNPR1* overexpressing leaves were measured three days after inoculation using ImageJ. Bar charts represent the means ± SE of measurements from 12 lesion spots from four leaf discs of each genotype. The asterisk denotes a significant difference determined by single factor ANOVA ($p<0.05$). **D.** Pathogen biomass was measured at the lesion sites by qPCR to determine the ratio of pathogen DNA to cacao DNA two days after inoculation. Bar charts represent four biological replicates, each with three technical replicates. The asterisk denotes a significant difference determined by single factor ANOVA analysis ($p<0.05$).

Table 3.1 - Number of genes up- and down-regulated (BH p value < 0.05) by SA treatment for ICS1 and Sca6 genotypes.

Genotype	Up	Down
ICS1	436	601
Sca6	490	447

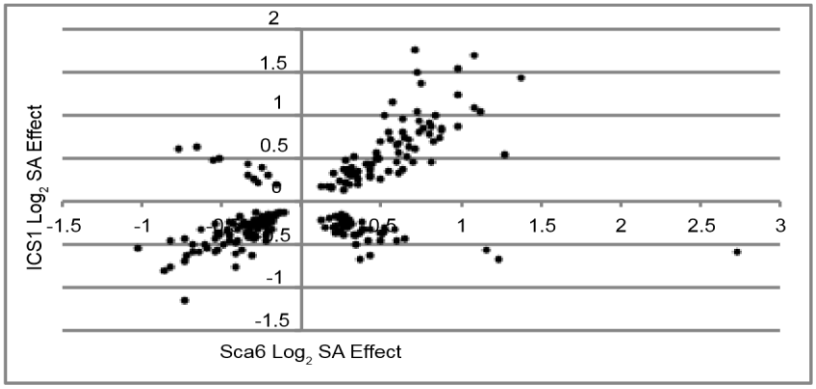


Figure 3.3 - Gene induction differences in Sca6 and ICS1. X-axis represents $\log_2$ expression change in SA-treated versus water-treated Sca6 leaves, after obtaining the general linear model mean of differences across leaf stages. Y-axis represents $\log_2$ expression change in SA-treated versus water-treated ICS1 leaves, after obtaining the general linear model mean of differences across leaf stages. Points represent the 234 genes with statistically significant (BH P<0.05) expression changes in both genotypes.

We also compared transcript changes between Sca6 and ICS1 genotypes in basal (water-treated) and induced (SA-treated) states by using a general linear model to assess differences across leaf developmental stages. More than 2000 genes were differentially expressed between the genotypes in both basal and induced states (BH-adjusted P<0.05). At the basal state 1124 genes had higher expression in Sca6 than ICS1 and 1121 genes had higher expression ICS1 than Sca6. Similarly, in the induced state we detected 1051 genes with higher expression in Sca6 than ICS1 and 1016 genes with higher expression ICS1 than Sca6.

Genotypes differ in induction of PR genes

A hallmark response to SA treatment is the induction of the PR genes, which are families of genes encoding proteins with direct effects on pathogens (Fu and Dong, 2013). Of the 354 PR genes we identified in the *T. cacao* Criollo genome, 136 were represented on the array and 55 passed background normalization (Supplementary Table S4). Unexpectedly, more PR genes were up-regulated by SA in ICS1, the susceptible genotype, than in Sca6 (Fig. 4; cutoffs of BH-adjusted $P<0.05$ and $P<0.1$). At either significance level, only one gene, Tc04_g016440 (a PR-14, putative non-specific lipid transfer protein), was up-regulated in both genotypes, 2.1-fold in Sca6 (BH $P=0.074$) and 4.11-fold in ICS1 (BH $P=0.005$). Of the 16 PR genes induced with BH-adjusted $P<0.1$ in ICS1, seven were class III peroxidase family members (PR-9 family). Conversely, Sca6 had no statistically significantly up-regulated class III peroxidases.

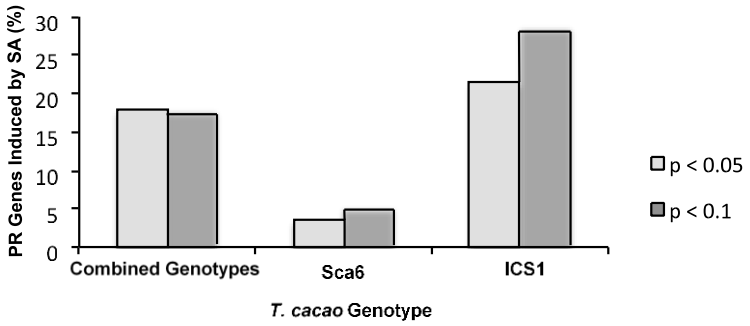


Figure 3.4 - Number of PR genes induced by SA treatment for all leaf stages in Sca6 alone, in ICS1 alone, and across both genotypes. Y-axis shows percentage of PR genes included on the microarray that were induced by SA. Light grey bars represent genes induced with BH $P<0.05$, dark grey bars represent genes induced with BH $P<0.10$.

Functions of most highly up-regulated genes suggest genotype-specificity in response to SA

To further investigate the nature of the response to SA and to identify the functional trends in transcript level changes, we examined the annotations of the genes most highly up-regulated in the two genotypes. The most highly up-regulated genes in Sca6 with available locus IDs for Arabidopsis homologues were encoded in the mitochondrial and chloroplast genomes with functions in the electron transport chains (Supplementary Table S1). Of the 45 most highly

up-regulated genes in Sca6, nine are chloroplastic, including subunits of photosystem I and II and RUBISCO, and nine are mitochondrial, several being NADH dehydrogenase subunits. In the ICS1 genotype, none of the 45 most highly up-regulated genes were annotated as being encoded in the chloroplast or mitochondrial genomes (Supplementary Table S2). However, of the top 21 most highly up-regulated genes in ICS1, six encode proteins with predicted functions as PR proteins, including a glucosidase, three different endochitinases, a class III peroxidase, and a non-specific lipid transfer protein, none of which were found to be in the most highly up-regulated genes in Sca6. This suggested that the mechanisms of SA-induced defense response in the two genotypes might differ significantly.

Using locus IDs for Arabidopsis homologues most related to the cacao genes on our microarray, parametric analysis of gene set enrichment (PAGE) was performed using the gene ontology (GO) (Ashburner *et al.*, 2000) annotations of the genes (Kim and Volsky, 2005). This analysis compares the relative proportions of genes represented on the microarray within specific GO annotation classes, with the proportions of the same classes found in sets of up-regulated genes in each comparison performed. This results in the calculation of a Z score that indicates the difference between two experimental groups for a specific annotation class and the statistical significance of that difference. We performed PAGE on three sets of genes: (i) genes significantly up- or down-regulated by SA in the Sca6 genotype, pooling developmental stages (BH-adjusted $P<0.05$) (Supplementary Table S5), (ii) genes significantly up- or down-regulated by SA in the ICS1 genotype, pooling developmental stages (Supplementary Table S6), and (iii) all genes with statistically significant differential regulation by SA after pooling samples from both genotypes and all three developmental stages (Supplementary Table S7). Overall, 37, 34, and 46 GO terms were enriched (Benjamini-Yekutieli adjusted $P\leq0.05$) (Benjamini and Yekutieli, 2001) in sets 1, 2 and 3 respectively (Supplementary Tables S5−7). Notably, the enriched categories in sets 1 and 2 differ, highlighting the differences in gene expression profiles between the two genotypes. The majority of categories contributing to the differences are associated with photosynthesis or chloroplastic structures, with up-regulation of the genes in these categories in Sca6 (set 1) and down-regulation in ICS1 (set 2). The most enriched terms in each of the three ontologies are included in Fig. 5.

PAGE analysis was also applied to compare transcript differences between genotypes in their basal and SA induced states (Supplementary Tables S8, 9). Using a general linear model,

significant genes were identified (BH-adjusted $P<0.05$) across all leaf stages, for each genotype comparing water-treated Sca6 to water-treated ICS1 (basal state) or SA-treated Sca6 to SA-treated ICS1 (induced state). In the comparison of the basal state of the two genotypes, 39 GO terms were enriched (Supplementary Table S8). Positive Z scores indicate higher expression in Sca6 than in ICS1, and negative Z scores indicate higher expression in ICS1 than in Sca6. Interestingly, several defense-related GO terms, including 'immune response' (GO:0006955), 'response to fungus' (GO:0009620) and 'response to chitin' (GO:0010200), along with 'cellular respiration' (GO:0045333) and 'transcription factor activity' (GO:0003700), were enriched and had negative Z scores, suggesting that ICS1, despite being more susceptible to disease, expressed more active basal defenses. Comparing the induced states, 52 GO terms were enriched (Supplementary Table S9). 'Response to fungus' again exhibited a negative Z score, along with 'response to wounding' (GO:0009611). The majority of the GO terms with positive Z scores were cellular component terms related to chloroplast or biological process terms related to photosynthesis. As processes of energy generation, particularly the light reactions of photosynthesis and electron transport in the mitochondria are known to produce ROS that could be associated with increases in pathogen resistance (Torres, 2010), this enrichment of chloroplast-related terms may explain the tolerance of Sca6 to a variety of pathogens.

qRT-PCR validates genotypic differences in gene induction

qRT-PCR was performed for six selected PR genes to validate the microarray results and obtain more quantitative transcript change measurements. The genes included two class III peroxidases, up-regulated by SA treatment in ICS1, and three chloroplast genes and two mitochondrial genes up-regulated in Sca6 (Fig. 6; Supplementary Table S10). The trends in regulation were consistent between the methods, thus validating the qualitative values from the microarray (Fig. 6A, B). In ICS1, PR genes tended to have higher transcript abundance after SA treatment, while transcript levels for the same genes in Sca6 were more consistent between treatments. Conversely, the three chloroplastic genes showed up-regulation in Sca6 after SA

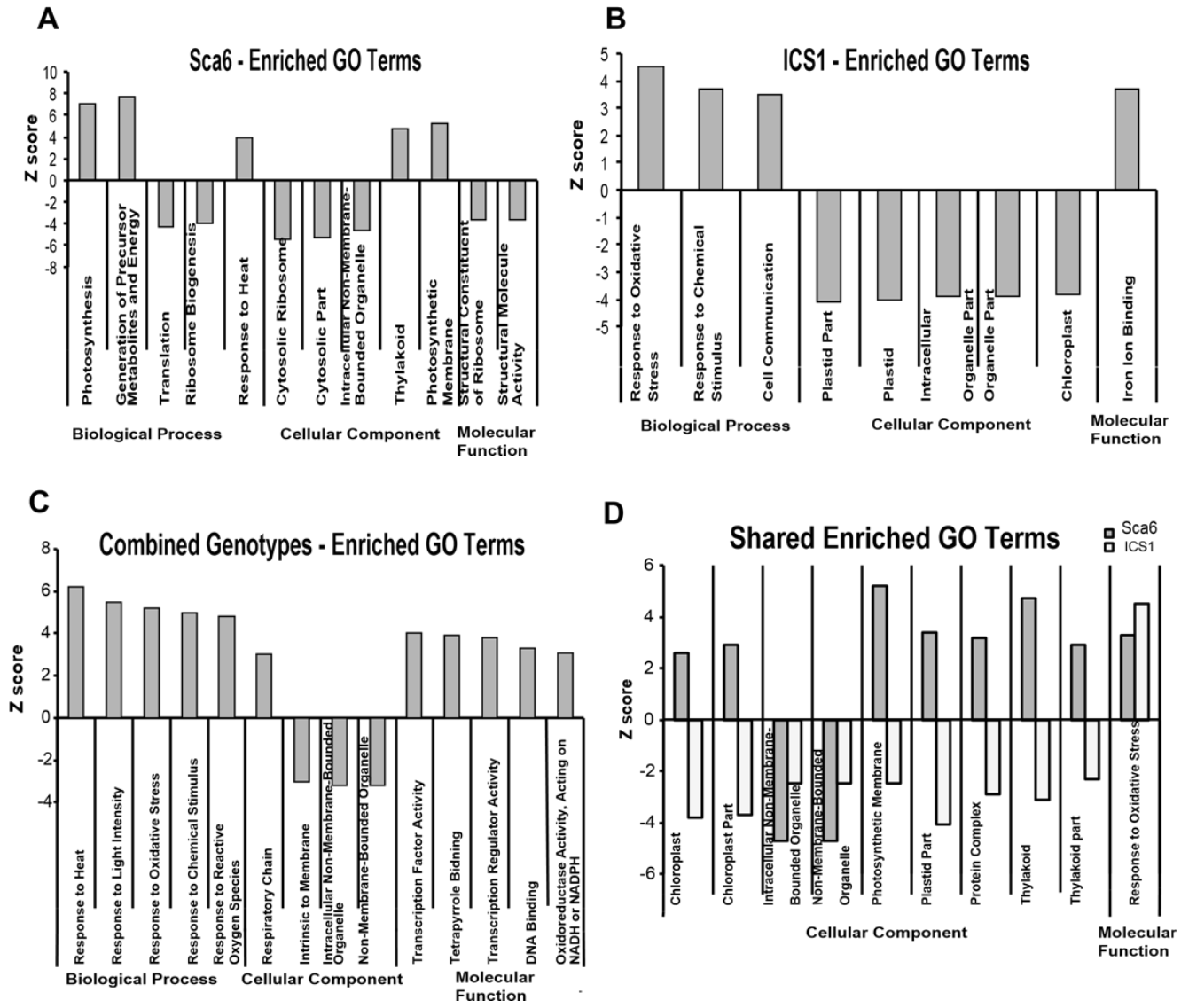


Figure 3.5 - Graphical representation of GO Enrichment by Parametric Analysis of Gene Set Enrichment (PAGE). **A.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes in Sca6. **B.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes in ICS1. **C.** Z scores of select GO terms calculated using PAGE on statistically significantly differentially regulated genes averaging both genotypes across all stages. **D.** Z scores for select GO terms with statistically significant enrichment in both genotypes. See Supplemental Tables 5-9 for detailed data.

treatment and some down-regulation in ICS1. This is consistent with PAGE analysis resulting in positive Z scores for chloroplast-related GO terms in Sca6 and negative Z scores for the same terms in ICS1. The higher variation observed for the chloroplast genes was due to differences among leaf developmental stages (data not shown). Similarly to the chloroplast genes, the two mitochondrial genes were up-regulated in Sca6, while minor differences were detected in ICS1.

Sca6 and ICS1 leaf tissues differ in basal and SA induced ROS levels

Some chloroplast and mitochondrial proteins, such as photosystem components and NADH dehydrogenase, function in electron transfer, contributing to the production of ROS (Moller, 2001; Gechev *et al.*, 2006; Noctor *et al.*, 2007; Sharma *et al.*, 2012). As chloroplast and mitochondrial genes were significantly induced in SA-treated Sca6 and several class III peroxidases were induced in SA-treated ICS1, we hypothesized that the production of ROS in the two genotypes could be driven by different mechanisms. This may contribute to Sca6's greater resistance to certain pathogens and perhaps the faster rate of senescence in ICS1. To test whether accumulation of ROS differed between the two genotypes, we performed nitroblue tetrazolium (NBT) and 3,3'-diaminobenzidine (DAB) staining to quantify the accumulation of superoxide and peroxide, respectively (Fig. 7). NBT reacts with superoxide to form a blue precipitate and DAB reacts with hydrogen peroxide, forming a brown precipitate. Plants of both genotypes were spray treated with 2 mM SA or water, then 24 h after treatment, stage C leaves were harvested from the trees and leaf discs were stained with DAB or NBT (Fig. 7A–H). ROS staining was quantified in replicated samples as described in 'Materials and methods'. Superoxide accumulation was significantly higher in SA-treated Sca6 than in any other genotype-treatment pair (Fisher's PLSD P=0.019 for Sca6 H2O vs. Sca6 SA) (Fig. 7I). Hydrogen peroxide accumulation increased in both genotypes with SA treatment (Fig. 7J) (Fisher's PLSD P=0.002 for ICS1 H2O versus ICS1 SA and P=0.001 for Sca6 H2O versus Sca6 SA). A significant difference in peroxide accumulation was also detected between water-treated ICS1 and Sca6 (Fisher's PLSD P=0.002). This difference in ROS accumulation in water-treated leaf tissue could be an indication of a true difference in basal ROS levels in the leaves or it could be attributed to a faster or stronger wound response in ICS1 that is associated with an ROS burst after excising the leaf discs.

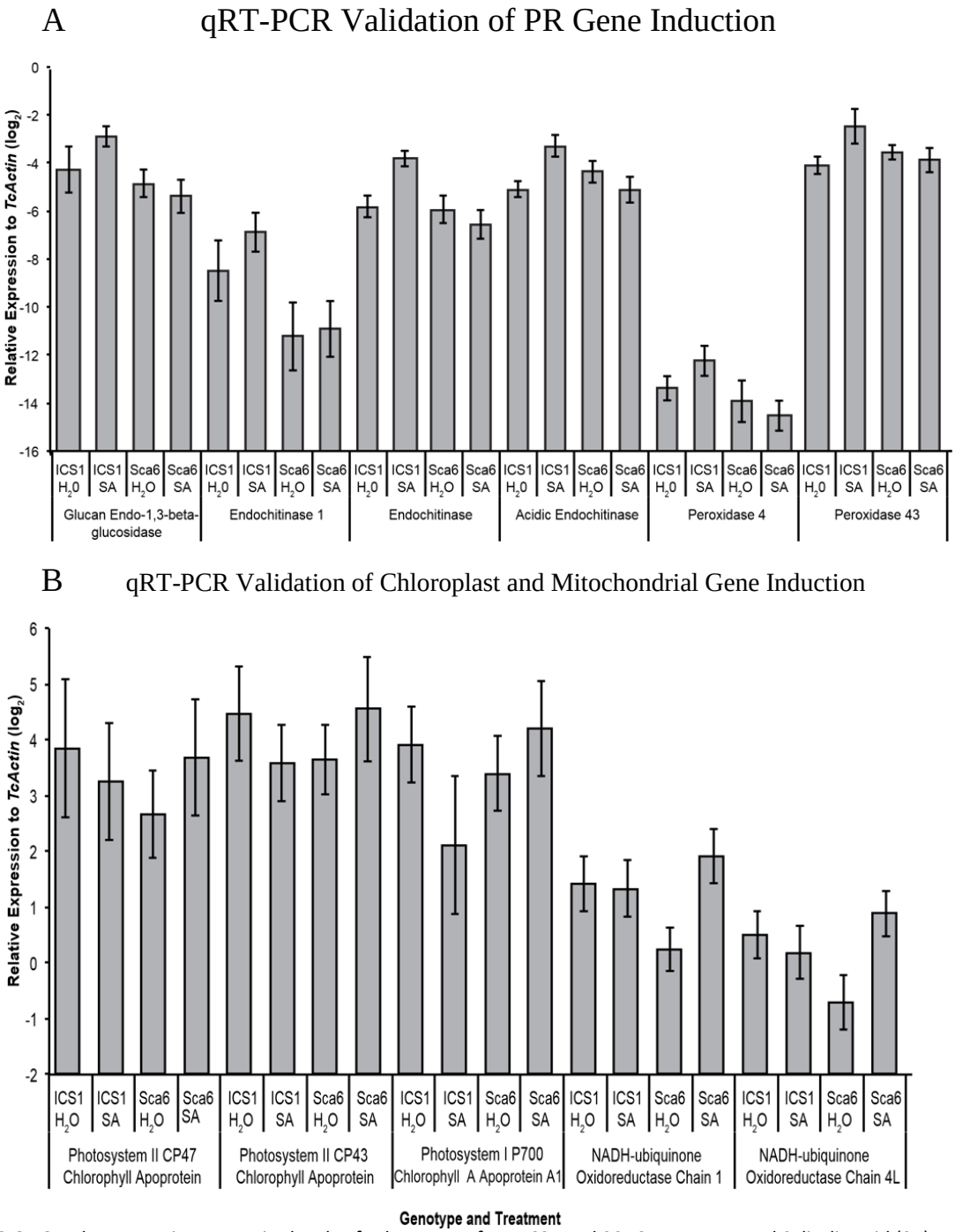
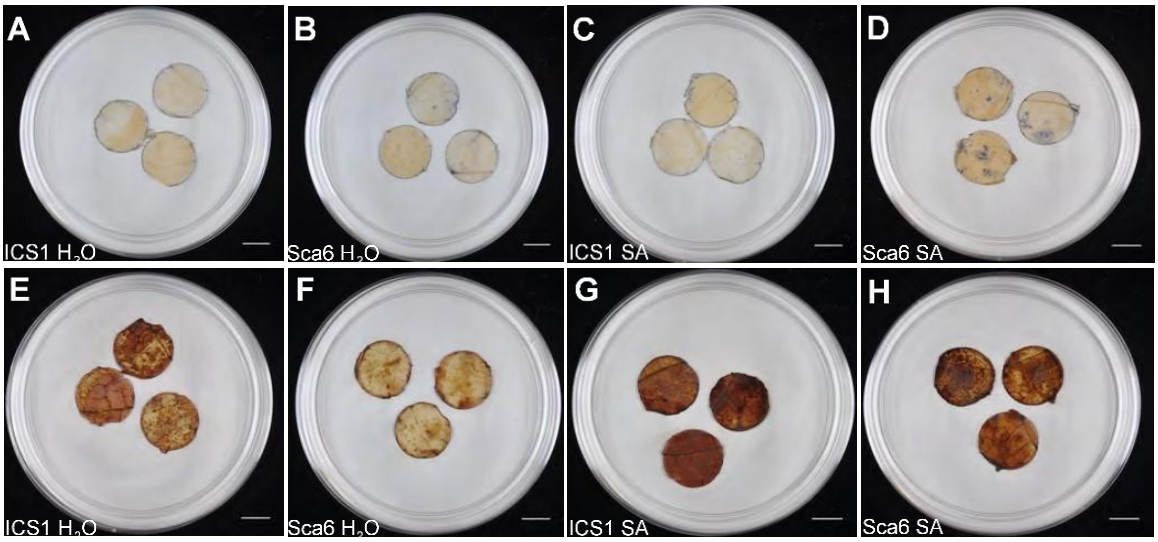
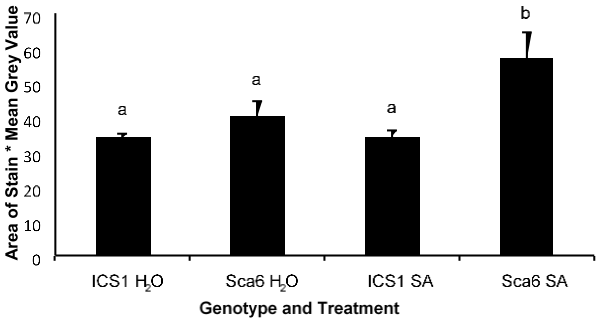


Figure 3.6 - Graphs comparing transcript levels of select genes from ICS1 and SCA6 genotypes and Salicylic Acid (SA) and water control treatments as detected by qRT-PCR. Each bar represents the mean of nine samples ± SE (three replicated from each developmental stage). Values are calculated relative to *TcActin*. **A.** Graph displaying effect of SA treatment on cacao PR genes. **B.** Graph displaying effect of SA treatment on cacao genes with Arabidopsis best hits encoded in chloroplasts and mitochondria. Primer sequences are listed in Supplemental Table 10.



I



J

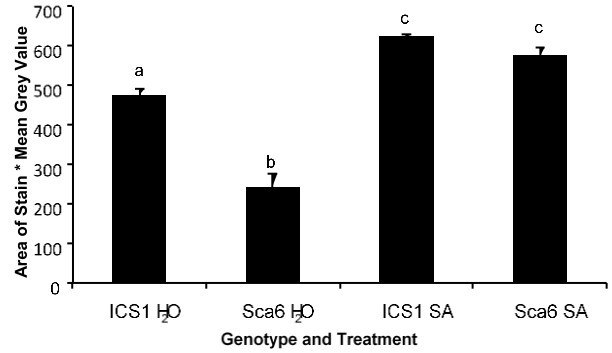


Figure 3.7 - Representative images showing NBT and DAB staining of cacao leaf discs 24 hours after Salicylic Acid (SA) or water (H_2O) treatment of leaves attached to trees: (**A**) NBT-stained ICS1 treated with water, (**B**) NBT-stained Sca6 treated with water (**C**) NBT-stained ICS1 treated with SA, (**D**) NBT-stained Sca6 treated with SA, (**E**) DAB-stained ICS1 treated with water, (**F**) DAB-stained Sca6 treated with water, (**G**) DAB-stained ICS1 treated with SA, (**H**) DAB-stained Sca6 treated with SA. Scale bars represent 1 cm. **I.** Graph displaying mean product area of leaf disc stained by NBT and mean grey value of stained area for each genotype and treatment. **J.** Graph displaying the mean product of area stained by DAB and mean grey value of stained area for each genotype and treatment. In both graphs, differences between bars marked with the same letter are not statistically significant (Fisher's PLSD analysis $p > 0.05$). Standard errors in both graphs were calculated from five biological replicates. Each biological replicate is a plate containing three leaf discs.

Discussion

While the role of SA in model plants has been well studied (Vlot *et al.*, 2009; Fu and Dong, 2013), we sought to verify its role in cacao defense and to explore the molecular basis of

the dramatic differences in resistance to pathogens between two cacao genotypes. We focused our analysis on the SA response, as it is the phytohormone thought to be most important for regulation of defense against biotrophs and hemibiotrophs (Fu and Dong, 2013) and on the difference in response between Sca6 and ICS 1, two genotypes strongly contrasting in their resistance phenotypes (Risterucci *et al.*, 2003; Faleiro *et al.*, 2006). Our data showed that treatment of leaves with 1 mM SA reduced lesion size and pathogen biomass in both ICS1 (more susceptible) and Sca6 (more tolerant) after inoculation of detached leaves with *P. tropicalis,* and their relative susceptibilities were consistent with the known disease resistance phenotypes of these two varieties. We also demonstrated that the cacao homologue to Arabidopsis NPR1 positively contributes to defense in cacao, reinforcing the importance of the SA response in cacao and supporting our previous results. (Shi *et al.*, 2010). Our microarray analysis revealed statistically significant differential transcript abundance for ~1000 genes in each genotype in response to SA, close to the number of genes induced by SA in Arabidopsis (Wang *et al.*, 2006). Interestingly, our study also suggested that there is a significant genotype specificity in the cacao defense response pathway.

A hallmark of induction of systemic acquired resistance is the increased expression of PR genes. Our data revealed that in response to SA treatment, more PR gene transcripts levels were elevated in the susceptible genotype ICS1 than in the tolerant genotype Sca6. This suggested that surprisingly, PR transcript levels were not correlated with the higher resistance of Sca6. Our results are in agreement with to those of Teixeira *et al.* (2014), who showed that transcriptional changes in the susceptible cacao cultivar 'Comum' in response to WBD infection revealed increased transcript levels of PR genes. The measurements of lesion growth and pathogen biomass with and without SA treatment combined with our microarray data suggest that SA-induced gene expression does partially contribute to resistance in ICS1, but it is unclear if the PR genes contribute directly to the observed resistance.

A second plausible hypothesis to explain these observations is that perhaps the higher basal transcript levels of defense genes mediates the resistance of Sca6 that is consequently enhanced by the SA treatment. However, our analysis of basal gene transcript levels indicated that gene transcript levels within several defense-related terms, including 'Immune Response' (GO:0006955), 'Defense Response' (GO:0006952), 'Response to Fungus' (GO:0009620), and 'Response to Chitin' (GO:0010200), were higher in basal ICS1 (Supplementary Table S8).

Detection of PR transcripts in the water-treated samples from the genotypes by the microarray and qRT-PCR also provide evidence against this hypothesis.

Interestingly, in our study nearly half of the PR genes induced in ICS1 by SA were class III peroxidase family members, proteins that are secreted into the cell wall or the apoplast and contribute to ROS generation (O'Brien *et al.*, 2012; Baxter *et al.*, 2013). Moreover, the SA treatment of Sca6 leaves induced the transcript levels of genes located in the chloroplast and mitochondrial genomes also known to contribute to ROS production; this increase could have resulted either from an increase in the transcription rate of the genes, or an increase in the number of organelles or content of organellar DNA. Among this set of genes were mitochondrial NADPH dehydrogenases and chloroplastic photosystem I and II components, both known to be involved in the oxidative burst, a major plant defense mechanism (Rao *et al.*, 1997; Torres and Dangl, 2005; Pogany *et al.*, 2009; Dubreuil-Maurizi *et al.*, 2010). Thus we hypothesized that the mechanism of resistance to pathogens in Sca6 involves expression of genes involved in ROS accumulation in the chloroplasts and mitochondria.

The results from GO enrichment analysis generally supported the hypothesis that ROS production strategy differed between the genotypes. While PAGE analysis (Kim and Volsky, 2005) revealed that both genotypes had a general up-regulation of genes annotated with the 'Response to Oxidative Stress' term, a major difference between the genotypes was elevation of transcripts of genes annotated with plastid-related GO terms in Sca6 and reduced transcripts of the same classes of genes in ICS1. Additionally, comparing the induced (SA-treated) states of the two genotypes, transcripts of genes annotated with 'Response to Fungus' remained higher in ICS1, but gene transcripts annotated with a variety of chloroplast-related terms were again more highly elevated in Sca6. Our data provide strong evidence for the importance of ROS generated in chloroplasts and mitochondria in the Sca6 genotype.

Twenty-four hours after SA treatment, more superoxide accumulated in Sca6, while there was no difference in hydrogen peroxide accumulation between the genotypes. This supports the interpretation of the transcriptome analysis: that higher ROS accumulation in the chloroplast may be unique to Sca6. While SA-treated samples from the genotypes did not differ in hydrogen peroxide accumulation, water-treated Sca6 had less hydrogen peroxide than water-treated ICS1. Thus there was a greater burst of both ROS types resulting from SA treatment in Sca6. The role of ROS in plant defense has been an active area of research (Apel and Hirt, 2004).

Production of superoxide has previously been linked to development of hypersensitive response in potato tubers inoculated with *P. infestans* (Doke, 1983), which is consistent with our finding that the resistant cacao genotype had greater superoxide accumulation in response to another *Phytophthora* species, *P. tropicalis*. Further, it is possible that differences in localization of ROS production differentially affect signalling pathways mediating resistance. It has been demonstrated that the elevation of endogenous SA can induce the production of ROS, which will in turn facilitate cell death at the site of infection (Apel and Hirt, 2004; Torres *et al.*, 2006; Mur *et al.*, 2008). As generation of ROS within the chloroplast (Liu *et al.*, 2007) and mitochondria (Cvetkovska and Vanlerberghe, 2012) have been linked to the hypersensitive response, further investigation of these processes in cacao and the differences between induction of these pathways in susceptible and resistant genotypes is needed. Alternatively, ROS production in response to WBD infection has been proposed to accelerate necrosis, as elevated peroxide concentration could lead to cell death, greater nutrient availability, and a more rapid progression to the pathogen's necrotrophic stage (de Oliveira Ceita *et al.*, 2007). As we detected higher peroxide concentrations in water-treated ICS1 leaves than in water-treated Sca6, it is possible that higher basal ROS levels in the susceptible genotype accelerate its infection by *Phytophthora* as has been proposed with WBD.

Our results reveal several important defense-related physiological differences between the two cacao genotypes. Further research is needed to explore more comprehensively the role of PR gene expression and ROS production in the immune response of cacao. Ultimately, this knowledge can be used to benefit cacao farmers and breeders by providing molecular strategies and markers for accelerated and efficient plant breeding.

References

Alvarez ME. 2000. Salicylic acid in the machinery of hypersensitive cell death and disease resistance. Plant Molecular Biology **44,** 429–442.

Apel K, Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annual Review of Plant Biology **55,** 373–399.

Argout X, Fouet O, Wincker P, *et al.* 2008. Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao L.* generated from various tissues and under various conditions. BMC Genomics **9,** 512.

Ashburner M, Ball CA, Blake JA, *et al.* 2000. Gene ontology: tool for the unification of biology. Nature Genetics **25,** 25–29.

Attaran E, Zeier TE, Griebel T, Zeier J. 2009. Methyl salicylate production and jasmonate signalling are not essential for systemic acquired resistance in *Arabidopsis*. Plant Cell **21,** 954–971.

Bailey BA, Bae H, Strem MD, Antunez de Mayolo G, Guiltinan MJ, Verica JA, Maximova SN, Bowers JH. 2005*a*. Developmental expression of stress response genes in *Theobroma cacao* leaves and their response to Nep1 treatment and a compatible infection by *Phytophthora megakarya*. Plant Physiology and Biochemistry **43,** 611–622.

Bailey BA, Strem MD, Bae H, de Mayolo GA, Guiltinan MJ. 2005*b*. Gene expression in leaves of *Theobroma cacao* in response to mechanical wounding, ethylene, and/or methyl jasmonate. Plant Science **168,** 1247–1258.

Baxter A, Mittler R, Suzuki N. 2013. ROS as key players in plant stress signalling. Journal of Experimental Botany doi: 10.1093/jxb/ert375.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57,** 289–300.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics **29,** 1165–1188.

Borrone JW, Kuhn DN, Schnell RJ. 2004. Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. Theoretical and Applied Genetics **109,** 495–507.

Bretz F, Hothorn T, Westfall P. 2010. Multiple Comparisons Using R. Boca Raton: CRC Press.

Brown JS, Schnell RJ, Motamayor JC, Lopes U, Kuhn DN, Borrone JW. 2005. Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an f2 population using SSR markers and candidate genes. Journal of the American Society for Horticultural Science **130,** 366–373.

Campos MA, Rosa DD, Teixeira JÉC, Targon MLPN, Souza AA, Paiva LV, Stach-Machado DR, Machado MA. 2007. PR gene families of citrus: their organ specific-biotic and abiotic inducible expression profiles based on ESTs approach. Genetics and Molecular Biology **30,** 917–930.

Cao H, Bowling SA, Gordon AS, Dong X. 1994. Characterization of an *Arabidopsis* mutant that is nonresponsive to inducers of systemic acquired resistance. Plant Cell **6,** 1583–1592.

Cvetkovska M, Vanlerberghe GC. 2012. Coordination of a mitochondrial superoxide burst during the hypersensitive response to bacterial pathogen in *Nicotiana tabacum*. Plant, Cell & Environment **35,** 1121–1136.

Daudi A, Cheng Z, O'Brien JA, Mammarella N, Khan S, Ausubel FM, Bolwell GP. 2012. The apoplastic oxidative burst peroxidase in *Arabidopsis* is a major component of pattern-triggered immunity. The Plant Cell Online **24,** 275–287.

De Block M, Herrera-Estrella L, Van Montagu M, Schell J, Zambryski P. 1984. Expression of foreign genes in regenerated plants and in their progeny. The EMBO Journal **3,** 1681–1689.

de Oliveira Ceita G, Macêdo JNA, Santos TB, *et al.* 2007. Involvement of calcium oxalate degradation during programmed cell death in *Theobroma cacao* tissues triggered by the hemibiotrophic fungus *Moniliophthora perniciosa*. Plant Science **173,** 106–117.

Doke N. 1983. Involvement of superoxide anion generation in the hypersensitive response of potato tuber tissues to infection with an incompatible race of *Phytophthora infestans* and to the hyphal wall components. Physiological Plant Pathology **23,** 345–357.

Dong X. 2004. NPR1, all things considered. Current Opinion in Plant Biology **7,** 547–552.

Drummond A, Ashton B, Buxton S, *et al.* 2012. Geneious v5.6, Available from http://www.geneious.com.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research **38,** 64–70.

Dubreuil-Maurizi C, Trouvelot S, Frettinger P, Pugin A, Wendehenne D, Poinssot B. 2010. Beta-aminobutyric acid primes an NADPH oxidase-dependent reactive oxygen species production during grapevine-triggered immunity. Molecular Plant-Microbe Interactions **23,** 1012–1021.

Durrant WE, Dong X. 2004. Systemic acquired resistance. Annual Review of Phytopathology **42,** 185–209.

Faleiro F, Queiroz V, Lopes U, *et al.* 2006. Mapping QTLs for Witches' Broom (*Crinipellis perniciosa*) Resistance in Cacao (*Theobroma cacao* L.). Euphytica **149,** 227–235.

Fu ZQ, Dong X. 2013. Systemic acquired resistance: turning local infection into global defense. Annual Review of Plant Biology **64,** 839–863.

Fu ZQ, Yan S, Saleh A, Wang W, Ruble J, Oka N, Mohan R, Spoel SH, Tada Y, Zheng N, Dong X. 2012. NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. Nature **486,** 228–232.

Gechev TS, Van Breusegem F, Stone JM, Denev I, Laloi C. 2006. Reactive oxygen species as signals that modulate plant stress responses and programmed cell death. Bioessays **28,** 1091–1101.

Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho KP, Schuster I, Macedo JN, Pereira GA, Cascardo JC. 2007. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa.* Annals of Botany **100,** 129–140.

Grant M, Lamb C. 2006. Systemic immunity. Current Opinion in Plant Biology **9,** 414–420.

Hebbar PK. 2007. Cacao diseases: a global perspective from an industry point of view. Phytopathology **97,** 1658–1663.

Kim S-Y, Volsky D. 2005. PAGE: Parametric analysis of gene set enrichment. BMC Bioinformatics **6,** 144.

Lee J, Nam J, Park HC, *et al.* 2007. Salicylic acid-mediated innate immunity in *Arabidopsis* is regulated by SIZ1 SUMO E3 ligase. Plant Journal **49,** 79–90.

Lee S, Hong JC, Jeon WB, Chung YS, Sung S, Choi D, Joung YH, Oh BJ. 2009. The salicylic acid-induced protection of non-climacteric unripe pepper fruit against *Colletotrichum gloeosporioides* is similar to the resistance of ripe fruit. Plant Cell Reports **28**, 1573–1580.

Lichtenstein C, Fuller S. 1987. Vectors for Plant Genetic Engineering. Academic Press.

Liu Y, Ren D, Pike S, Pallardy S, Gassmann W, Zhang S. 2007. Chloroplast-generated reactive oxygen species are involved in hypersensitive response-like cell death mediated by a mitogen-activated protein kinase cascade. The Plant Journal **51,** 941–954.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative pcr and the $2-\Delta\Delta C_t$ method. Methods **25,** 402–408.

Loake G, Grant M. 2007. Salicylic acid in plant defense—the players and protagonists. Current Opinion in Plant Biology **10,** 466–472.

Maximova SN, Dandekar AM, Guiltinan MJ. 1998. Investigation of *Agrobacterium*-mediated transformation of apple using green fluorescent protein: high transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting. Plant Molecular Biology **37,** 549–559.

Maximova SN, Marelli JP, Young A, Pishak S, Verica JA, Guiltinan MJ. 2006. Over-expression of a cacao class I chitinase gene in *Theobroma cacao* L. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224,** 740–749.

Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S. 2012. Expression of designed antimicrobial peptides in *Theobroma cacao* L. trees reduces leaf necrosis caused by *Phytophthora* spp. Small Wonders: Peptides for Disease Control **1905,** 379–395.

Mejía LC, Herre EA, Sparks JP, *et al.* 2014. Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5,** 479.

Mitsuhara I, Ugaki M, Hirochika H, *et al.* 1996. Efficient promoter cassettes for enhanced expression of foregin genes in dicotyledonous and monocotyledonous plants. Plant and Cell Physiology **37,** 49–59.

Moller IM. 2001. Plant mitochondria and oxidative stress: electron transport, NADPH turnover, and metabolism of reactive oxygen species. Annual Review of Plant Physiology and Plant Molecular Biology **52,** 561–591.

Mur LA, Kenton P, Lloyd AJ, Ougham H, Prats E. 2008. The hypersensitive response; the centenary is upon us but how much do we know? Journal of Experimental Botany **59,** 501–520.

Noctor G, De Paepe R, Foyer CH. 2007. Mitochondrial redox biology and homeostasis in plants. Trends in Plant Science **12,** 125–134.

O'Brien J, Daudi A, Butt V, Paul Bolwell G. 2012. Reactive oxygen species and their role in plant defense and cell wall metabolism. Planta **236,** 765–779.

Overmyer K, Brosché M, Kangasjärvi J. 2003. Reactive oxygen species and hormonal control of cell death. Trends in Plant Science **8,** 335–342.

Pfaffl MW, Horgan GW, Dempfle L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research **30,** e36.

Phillips-Mora W, Wilkinson MJ. 2007. Frosty pod of cacao: a disease with a limited geographic range but unlimited potential for damage. Phytopathology **97,** 1644–1647.

Pogany M, von Rad U, Grun S, Dongo A, Pintye A, Simoneau P, Bahnweg G, Kiss L, Barna B, Durner J. 2009. Dual roles of reactive oxygen species and NADPH oxidase RBOHD in an *Arabidopsis-Alternaria* pathosystem. Plant Physiology **151,** 1459–1475.

Purdy LH, Schmidt RA. 1996. Status of cacao witches' broom: biology, epidemiology, and management. Annual Review of Phytopathology **34,** 573–594.

Rao MV, Paliyath G, Ormrod DP, Murr DP, Watkins CB. 1997. Influence of salicylic acid on H_2O_2 production, oxidative stress, and H_2O_2-metabolizing enzymes. Salicylic acid-mediated oxidative damage requires H_2O_2. Plant Physiology **115,** 137–149.

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C. 2003. Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theoretical and Applied Genetics **108,** 168–174.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9,** 671–675.

Shah J. 2003. The salicylic acid loop in plant defense. Current Opinion in Plant Biology **6,** 365–371.

Sharma P, Jha AB, Dubey RS, Pessarakli M. 2012. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. Journal of Botany doi: 10.1155/2012/217037.

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan M. 2010. Functional analysis of the *Theobroma cacao* NPR1 gene in *Arabidopsis*. BMC Plant Biology **10,** 248.

Shi Z, Zhang Y, Maximova S, Guiltinan M. 2013. TcNPR3 from *Theobroma cacao* functions as a repressor of the pathogen defense response. BMC Plant Biology **13,** 204.

Shimono M, Sugano S, Nakayama A, Jiang CJ, Ono K, Toki S, Takatsuji H. 2007. Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. Plant Cell **19,** 2064–2076.

Swanson JD, Carlson JE, Guiltinan MJ. 2008. Comparative flower development in *Theobroma cacao* based on temporal morphological indicators. International Journal of Plant Sciences **169,** 1187–1199.

Teixeira PJPL, Thomazella DPdT, Reis O, *et al.* 2014. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. The Plant Cell **26,** 4245–4269.

Torres MA. 2010. ROS in biotic interactions. Physiologia Plantarum **138,** 414–429.

Torres MA, Dangl JL. 2005. Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. Current Opinion in Plant Biology **8,** 397–403.

Torres MA, Jones JDG, Dangl JL. 2006. Reactive oxygen species signalling in response to pathogens. Plant Physiology **141,** 373–378.

Uquillas C, Letelier I, Blanco F, Jordana X, Holuigue L. 2004. NPR1-independent activation of immediate early salicylic acid-responsive genes in *Arabidopsis*. Molecular Plant-Microbe Interactions **17,** 34–42.

Verica JA, Maximova SN, Strem MD, Carlson JE, Bailey BA, Guiltinan MJ. 2004. Isolation of ESTs from cacao (*Theobroma cacao L.*) leaves treated with inducers of the defense response. Plant Cell Reports **23,** 404–413.

Vlot AC, Dempsey DMA, Klessig DF. 2009. Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47,** 177–206.

Vlot AC, Klessig DF, Park SW. 2008. Systemic acquired resistance: the elusive signal(s). Current Opinion in Plant Biology **11,** 436–442.

Wang D, Amornsiripanitch N, Dong X. 2006. A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants. PLoS Pathogens **2,** e123.

Wood GAR, Lass R. 2008. *Cocoa*, 4th edition. Wiley Blackwell.

Wu Y, Zhang D, Chu Jee Y, Boyle P, Wang Y, Brindle Ian D, De Luca V, Després C. 2012. The *Arabidopsis* NPR1 protein is a receptor for the plant defense hormone salicylic acid. Cell Reports **1,** 639–647.

Yamada MM, Lopes UV. 1999. Paternity analysis of cacao trees selected for resistance to witches' broom in plantations of Bahia, Brazil. Agrotrópica **11,** 83–88

Chapter 4: Genetic Diversity Assessment of Three Cacao Populations using SNP Identification in Putative Defense Genes

Abstract

Assessment of genetic diversity within and between populations of crop plants, particularly those with long generation times, is critical for enabling breeding programs to effectively utilize available germplasm. Here we used a PCR product sequencing strategy to assess genetic diversity in three defense genes using three populations of cacao plants including more than 150 genotypes. The analysis identified 16 SNPS within the ~2200 bp surveyed, corresponding to nine non-synonymous and seven synonymous mutations. Eight of the non-synonymous mutations sit within predicted functional domains of the encoded polypeptides, which may be more likely to affect protein function. Nucleotide diversity assessment showed that the surveyed French Guianan population harbored less variability than the Peruvian or Ecuadorian populations, consistent with earlier work showing less diversity within French Guianan cacao. Populations contained up to 9 gene haplotypes, and encode up to seven unique protein variants. Two genes, a polygalacturonase inhibitor and cysteine-rich repeat secretory peptide 38, show signatures of diversifying selection in one and two surveyed populations, respectively. While the study focuses on haplotypic analysis of only three genes, it identifies differences between the test populations, and may provide a powerful strategy for detecting useful genetic variation to incorporate into breeding programs.

Background

Recent advances in crop plant genomics and genotyping strategies have revolutionized strategies for identifying markers and accelerating breeding (Huang and Han, 2014). *Theobroma cacao L.*, a tree crop and the source of cocoa powder and cocoa butter, recently had two genome sequences published (Argout et al., 2011; Motamayor et al., 2013), which motivate and

greatly expedite molecular and population genetic analysis. Studying genetic diversity in cacao populations is crucial as ~40% of the crop is lost to a variety of pathogens annually (Guiltinan et al., 2008; Wood and Lass, 2008). A better understanding of the functional diversity within and between cacao populations, particularly relating to defense genes, is integral to developing superior, pathogen-tolerant varieties.

Early work characterizing the genetic diversity of cacao described the region between Colombia, Ecuador, and Peru as cacao's center of diversity (Cheeseman, 1944). At the time, populations were divided into three groups, the widely produced Forastero, the fine-flavor Criollo, and the Trinitario, which were hybrids of the former two. More recently, analysis of accessions from a wider geographic area revealed 10+ clusters of genetic diversity in cacao, which likely diverged due to ancient ridges that obstructed dispersal (Motamayor et al., 2008). The designations given to these clusters are: Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañon, Nacional, Nanay, and Purús (Motamayor et al., 2008).

Work assessing genetic markers within populations has begun to characterize genetic variation. Researchers have begun to characterize diversity within the Brazilian Amazon (Sereno et al., 2006), Peruvian Amazon (Zhang et al., 2006), and a large germplasm collection maintained in Costa Rica (Zhang et al., 2009). Analysis of populations from different countries revealed varying degrees of genetic diversity, with the French Guianan population showing a markedly low number of alleles and observed heterozygosity and a high inbreeding coefficient in the analyzed set (Zhang et al., 2009). However so far most of the diversity analyses in cacao germplasm were based on neutral markers (e.g. SSR), which cannot reflect the full impact of genetic diversity, especially for functional genetic variation affecting adaptation, agronomic traits, and quality attributes of this species. Further, the ability to infer evolutionary forces acting on gene sequences based on sequence variability is a widely established strategy (Hughes and Nei, 1998), and has been applied to studying plant defense evolution (Michelmore and Myers, 1998, Karasov et al., 2014). Surveying allelic variability of defense genes in different populations and germplasm groups would be informative for understanding the genetic background relating to cacao's history of adaptation, establish the link between genotype and phenotype in different environmental conditions, and ultimately improve the efficiency in conservation and use of cacao germplasm for crop improvement.

In this study, we analyzed sequence polymorphism of putative defense genes in cacao germplasm originating from French Guiana, Northern Peru, and the Pacific coast of Ecuador. Our objectives were to compare the allelic diversity of defense genes in different geographic regions, as well as between wild and cultivated populations. We also performed genotyping and nucleotide diversity analyses to assess the cause of the current diversity distribution. While further functional assessment is necessary to discover the phenotypic variation created by the genetic variation described here, this study serves as a proof of principle for a strategy of discovering potentially important genetic variation in defense genes and potentially use them to identify candidate gene markers for germplasm evaluation.

Materials and Methods

Plant material

The analysis here focused on cacao germplasm from three geographical regions - Northern Peru, French Guianan, and the coastal region of Ecuador. Among them, Northern Peru is located in the center of diversity of cacao and the Pound collection, a set of widely used germplasm from this region, includes members of the Nanay (NA), Iquitos (IMC), Marañon (PA), and Contamana (SCA) genetic clusters (Motamayor et al., 2008; Pound, 1943; Zhang et al., 2009). French Guiana is at the border region of cacao's primary gene pool in the Amazon, which represents a different population with lower genetic diversity, as revealed by microsatellite markers (Lachenaud and Zhang, 2008). The sample set from the coastal region of Ecuador is a cultivated population called 'Refractario', which is a mixed hybrid group derived from "Nacional" - a traditional Ecuadorian variety and other cultivated varieties from Venezuela (Zhang at al., 2007).

qRT-PCR measurement of target genes

qRT-PCR analyses were performed on the RNA samples produced for the microarray experiments measuring gene induction after pathogen treatment. One microgram of RNA from each sample was used for reverse transcription to cDNA by M MuLV Reverse Transcriptase (New

England Biolabs, Ipswich, MA) using oligo-$(dT)_{15}$ as a primer. qRT-PCR was performed in 10 µL reactions, consisting of 5 µL of SYBR Green PCR Master Mix (TAKARA, Tokyo, Japan), 0.2 µL of Rox, and 0.4 µL of each primer, diluted to 5 µM, and 4 µL of cDNA. Each reaction was performed in technical duplicate using the Applied Biosystem Step One Plus Realtime PCR System (Roche Applied Science, Foster City, CA) with the following program: 15 min at 94 °C, 40 cycles of 15 s at 94 °C, 20 s at 60 °C and 40 s at 72 °C. Cacao Tubulin1 and ACP1 genes were used as reference genes, and expression of target genes was compared to these using REST (Pfaffl et al., 2002).

PCR amplification of target genes

Standard PCR was used to amplify the three target genes from genomic DNA samples. PCR was performed with Phusion (Thermo-Fisher Scientific, Waltham, MA) high-fidelity polymerase, using the recommended protocols, in 50 µL reactions. The thermocycling protocol was: 2 min at 98^{o}C for initial denaturation, 30 s at 98^{o}C denaturation, 30 sec annealing, 1 min extension at 72^{o}C, 30 cycles of steps 2-4, followed by a 7 min final extension at 72^{o}C. The annealing temperatures were 57^{o}C for PR1, 51^{o}C for PGIP, and 53^{o}C for CRSP38. 5µL of PCR products were run on a 2% agarose gel to ensure that a single band was generated.

PCR cleanup

PCR cleanup was performed using Denville SpinSmart (Denville Scientific, Denville, NJ) columns according to the supplied protocol. The recommended additional wash step was used in order to increase sample purity. Samples were eluted from columns using 20 µL of nuclease free water. Concentration of purified PCR product was measured using a Nanodrop 2000 (Thermo-Fisher Scientific, Waltham, MA) spectrophotometer. Concentrations were consistently >30ng/µL, with 260/280 ratios of ~1.8 and 260/230 ratios of ~1.

Sanger sequencing

2 µL of PCR product was aliquotted into 96-well plates with 2 µL of the 0.05 µM primer as requested by the Penn State Genomics Core Facility. Sequencing was performed at the facility

using an Applied Biosystems 3730XL sequencing system (Thermo Fisher Scientific, Waltham, MA) and was downloaded from their server when available.

Sequence Analysis

Downloaded sequences were imported into Geneious (Drummond et al., 2012) and aligned with the Criollo genome's reference sequence for each gene using the most stringent alignment parameters. After alignment, sequences were manually curated. In cases where low quality sequence prevented identification of a base at a given position, the sequence was ignored. Typically, the forward and reverse primers each gave sequence for the majority of the gene, so most samples were represented by two sequences. Electropherogram peaks were manually analyzed to perform base-calling. To identify a heterozygote, both strands of a sequence had to show partially overlapping peaks in a high quality region of sequence. If only one sequence was available, that genotype was not included in subsequent analyses unless the base call at the position could be easily discriminated. Ultimately, each sample was represented by a single sequence using IUPAC degeneracy code for heterozygous sites and an N in positions with missing or unreliable data.

After sequences were curated, alignment files were exported from Geneious and analyzed using DnaSP v 5.10 (Librado and Rozas, 2009). Only sequences with a definitive base call (A,C,T,G or degeneracy code for a heterozygous site) at each site where variability was seen during manual curation were included for analysis in DnaSP. In the program, the PHASE algorithm (Stephens et al., 2001) was used for prediction of haplotypes, estimation of nucleotide diversity, and detection of selection. PHASE was run with the default parameters (100 iterations, 1 thinning interval, and 100 burn-in iterations).

Functional Domain Prediction

The Criollo genome (Argout et al., 2011) reference nucleotide sequence for each of the three genes was used as a query in NCBI's CDSearch (v3.14) (Marchler-Bauer et al., 2015). All identified domains have an E-value <0.01. The domains were annotated onto the sequence in using Geneious to display domain positions.

The EMBOSS (Rice et al., 2000) plugin in Geneious was also used to predict secondary structure of amino acid sequences.

Results

Gene Prioritization and Expression Analysis

In this study, we sought to identify variability within the sequence of putatively important defense genes. Two parameters were used to select genes, induction by pathogens and high pN/pS ratios (normalized polymorphism in protein-coding regions in several species), as measured in other tropical trees. An ongoing project evaluating Resistance gene diversity in a set of tropical trees (Marden et al, 2016, in review) identified orthologous genes from six tree species with that showed signatures of diversifying selection (high pN/pS) in several or all of the six species surveyed. Ultimately three genes were chosen, a Pathogenesis-Related 1 family member (PR-1, Tc02_g002410), a Cysteine-rich repeat secretory peptide 38 (CRSP38, Tc06_g009580), and a polygalacturonase inhibitor protein (PGIP, Tc05_g018290). Orthologs of CRSP38 and PGIP in other tropical trees were found to be highly variable, so we chose these to assess the ability to predict a high degree of genetic variability in cacao orthologs. All of these genes are members of gene families with putative roles in pathogen defense.

A previously described microarray experiment (GEO: GSE73804) measuring gene induction in response to treatment with *P. palmivora* and *C. theobromicola* showed that the PR-1 family member was upregulated 125.4-fold and 91.3-fold by the pathogens, respectively. To verify the microarray data qRT-PCR was performed using the RNA collected for the microarray experiment. Validation by qRT-PCR analysis using REST indicated that *P. palmivora* increased expression of the gene 763-fold, where *C. theobromicola* increased expression 55.7-fold. Validation of upregulation, and its low predicted variability from the tropical tree diversity dataset led to its being selected as a candidate gene considered as a low variability control.

Analysis of these microarray data revealed that transcript abundance of TcCRSP38 in leaves was also highly induced by pathogen, up-regulated 97-fold by treatment of plants with *Phytophthora palmivora* and 46-fold by treatment with *Colletotrichum theobromicola*. To verify

the microarray results, cDNA was produced from the RNA used in the original microarray experiment. Subsequent qRT-PCR verified that the gene was induced by both pathogens, and REST (Pfaffl et al., 2002) was used to calculate fold-induction. These data showed a 33.9-fold induction by *P. palmivora* and 11.9 fold by *C. theobromicola.* As the gene was highly induced and was predicted to have high variability in tropical tree orthologs, it was used for subsequent genotyping experiments.

The microarray used did not detect induction of the polygalacturonase inhibitor gene. qRT-PCR was performed, and subsequent qRT-PCR analysis using REST showed that the gene was not induced, and had low transcript abundance across all treatments. Nonetheless, the high pN/pS ratio in the Marden *et al*. analysis and its putative role in inhibiting pathogenic polygalacturonases led to its inclusion in subsequent experiments.

Genotyping analysis of three putative defense genes

The Pathogenesis-Related 1 (PR-1) locus (Tc02_g002410) was not predicted to have high variability, but has been shown to be strongly induced by infection of cacao with two pathogens. In analyzing sequence data from four populations of trees, only one site was found to be variable, but this mutation was only present at a low frequency (0.07%) in the Pound population. Proportions of bases detected at each variable site in PR1 are shown in Table 1. Site 160 showed either an A or G in the first position of a codon, encoding either a valine or isoleucine residue.

Table 4.1 - Genotype counts and frequencies for the variable site detected in PR-1 (Tc02_g002410).

			French Guiana		**Nacional**		**Pound**		**Total**	
	Criollo Reference Base	**Detected Genotypes**	**Counts**	**Freq.**	**Counts**	**Freq.**	**Counts**	**Freq.**	**Counts**	**Freq.**
		A/A	0/68	0	0/48	0	1/78	.013	1/194	.005
Site		A/G	0/68	0	0/48	0	11/78	.141	11/194	.057
160	G	G/G	68/68	1	48/48	1	66/78	.846	182/194	.938

In the CRSP38 locus, four positions were found to be variable. Bases detected and the frequency of the alleles detected are shown in Table 2. Interestingly the same four bases were

detected as variable in the Nacional and Pound populations, and the three variable sites in the French Guianan population were included in those four. In short, no variability was population-specific. Notably, one variable site, base 212, was found to have two alleles in the Nacional and Pound populations, but was fixed in the French Guiana population. Two of these were non-synonymous mutations and the others were synonymous.

Within the polygalacturonase gene (PGIP) (Tc05_g018290), eleven sites were found to be variable, with six resulting in synonymous variability and five resulting in changes to the amino acid sequence. Bases detected in the PGIP gene and their frequencies are shown in Table 3. Again we found that fewer variable sites were detected in the French Guianan, with only four of the eleven sites showing variation. The Pound population had variation at 10 sites, with site

Table 4.2 - Genotype counts and frequencies for variables sites detected in CRSP38 (Tc06_g009580).

			French Guiana		Nacional		Pound		Total	
	Criollo Reference Base	Detected Genotypes	Counts	Freq.	Counts	Freq.	Counts	Freq.	Counts	Freq.
Site 212	A	A/A	68/68	1.0	29/49	.592	32/55	.582	129/172	.750
		A/G	0/68	0	19/49	.388	18/55	.327	37/172	.215
		G/G	0/68	0	1/49	.020	5/55	.091	6/172	.035
Site 378	G	C/C	8/66	.121	1/49	.020	5/54	.093	14/169	.083
		C/G	17/66	.268	30/49	.612	14/54	.259	61/169	.361
		G/G	41/66	.621	18/49	.367	35/54	.648	94/169	.556
Site 404	A	A/A	8/65	.123	10/49	.204	23/56	.411	41/170	.241
		A/G	17/65	.262	36/49	.735	21/56	.375	74/170	.435
		G/G	40/65	.615	3/49	.061	12/56	.214	55/170	.324
Site 482	G	A/A	8/68	.118	1/49	.020	7/55	.127	16/172	.093
		A/G	17/68	.250	30/49	.612	16/55	.291	63/172	.366
		G/G	43/68	.632	18/49	.367	32/55	.582	93/172	.541

843 fixed. The Nacional population was the only population to show variation at site 843; the other sites were shared by Nacional and Pound.

Overall, this survey identified 16 variable bases. The majority of alleles detected for the CRSP38 and PGIP loci exist at intermediate frequencies, often an indicator of balancing selection. Eleven of these sites were transition mutations, eight being A/G and three being C/T. Of the five transversions, three were C/G, one was A/C, and one was G/T. Notably, more variation was detected in genes predicted using the high pN/pS metric than in the PR1 gene. In PGIP we found 11 variable sites within its 993 bases, giving 11.08 variable sites per kilobase. For CRSP, we found 4 variable sites in 729 bases of its sequence, giving 5.48 variable sites per kilobase. In the PR1 gene, we found 1 variable site in 489 bases, giving 2.04 variable sites per kilobase. While additional experimentation would give further support, this is a good first indication that selecting genes based on high pN/pS in orthologs is a valuable strategy for variation that could be used to develop molecular markers to be used in breeding programs.

Table 4.3 - Genotype counts and frequencies for variable sites detected in PGIP (Tc05_g018290).

			French Guiana		Nacional		Pound		Total	
	Criollo Ref. Base	Detected Genotypes	Counts	Freq.	Counts	Freq.	Counts	Freq.	Counts	Freq.
		C/C	0/53	0	1/53	.019	1/50	.020	2/156	.013
Site		C/G	0/53	0	14/53	.265	3/50	.060	17/156	.109
270	C	G/G	53/53	1	38/53	.717	46/50	.920	137/156	.878
		C/C	49/49	1	38/53	.717	46/50	.920	133/152	.875
Site		C/T	0/49	0	14/53	.265	3/50	.060	17/152	.112
423	T	T/T	0/49	0	1/53	.019	1/50	.020	2/152	.013
		C/C	36/50	.720	34/53	.642	9/50	.180	79/153	.516
Site		C/T	4/50	.200	17/53	.321	18/50	.360	39/153	.255
552	C	T/T	10/50	.080	2/53	.038	23/50	.460	35/153	.229
		A/A	48/48	1	38/53	.717	46/50	.920	132/151	.874
Site		A/G	0/48	0	14/53	.264	3/50	.060	17/151	.113
562	G	G/G	0/48	0	1/53	.019	1/50	.020	2/151	.013
		C/C	27/48	.563	22/53	.415	40/50	.800	89/151	.589
Site		C/G	3/48	.063	24/53	.453	8/50	.160	35/151	.232
604	C	G/G	18/48	.375	7/53	.132	2/50	.040	27/151	.179
		G/G	9/49	.184	0/53	0	4/49	.082	13/151	.086
Site		G/T	3/49	.061	1/53	.019	9/49	.184	13/151	.086
628	T	T/T	37/49	.755	52/53	.981	36/49	.735	125/151	.828

		A/A	0/49	0	1/53	.019	1/49	.020	46/151	.305
Site		A/C	0/49	0	14/53	.264	3/49	.061	17/151	.113
650	A	C/C	49/49	1	38/53	.717	45/49	.918	88/151	.583
		A/A	0/49	0	1/53	.019	1/49	.020	2/151	.013
Site		A/G	0/49	0	14/53	.264	3/49	.061	17/151	.113
669	A	G/G	49/49	1	38/53	.717	45/49	.918	132/151	.874
		A/A	0/49	0	1/53	.019	1/48	.021	2/150	.013
Site		A/G	0/49	0	14/53	.264	5/48	.104	19/150	.127
714	A	G/G	49/49	1	38/53	.717	42/48	.875	129/150	.860
		A/A	0/49	0	8/53	.151	0/44	0	8/146	.055
Site		A/G	0/49	0	27/53	.509	0/44	0	27/146	.185
843	G	G/G	49/49	1	18/53	.340	44/44	1	111/146	.760
		C/C	13/47	.277	6/53	.113	1/38	.026	20/138	.145
Site		C/T	8/47	.170	27/53	.509	8/38	.211	43/138	.312
851	T	T/T	26/47	.553	20/53	.377	29/38	.763	75/138	.543

Nucleotide diversity and haplotype analysis

We next used DNAsp (Librado and Rozas, 2009) to evaluate nucleotide diversity and predict the haplotypes of the surveyed sequences for each gene within each population and combining sequences across the three populations. In DNAsp, the PHASE (Stephens et al., 2001) algorithm was applied to estimate the number of haplotypes (H) in each population. DNAsp was used to calculate θ, Watterson's estimator of mutation rate, and π, the number of average pairwise differences per base, and to perform Tajima's test (Tajima, 1989) and Fu and Li's F and D tests (Fu and Li, 1993; Fu, 1995, 1997). Results are displayed in Table 4.

Table 4.4 - Nucleotide diversity and haplotype analysis for three genes and three populations.

Gene	Population	Sequences Included	SNP No.	θ (Per Site)	π	H	Tajima's D-test	Fu and Li's D*	Fu and Li's F*	GC (%)
PR1	French Guiana	136	0	0	0	1	0	0	0	54
	Nacional	96	0	0	0	1	0	0	0	54
	Pound	144	1	.00037	0.0003	2	-0.0889	0.471	0.351	54
	Combined	376	1	.00031	0.0001	2	-0.5197	0.427	0.151	54
CRSP38	French Guiana	122	3	.00077	0.0016	2	1.9	0.81	1.36	46
	Nacional	98	4	.00106	0.0023	7	2.327 †	0.942	1.63	46

	Pound	106	4	.00105	0.0022	7	2.135 †	0.935	1.557	46
	Combined	326	4	.00086	0.0023	8	2.829 ‡	0.849	1.81 †	46
PGIP	French Guiana	82	4	.00081	0.0017	6	2.289 †	0.958	1.621	44
	Nacional	100	11	.00195	0.0033	8	1.438	0.759	1.187	44
	Pound	66	10	.00212	0.0022	9	0.0919	1.39	1.13	44
	Combined	248	11	.00182	0.0027	12	1.133	1.369	1.538	44
† $p < 0.05$, ‡ $p < 0.01$										

The PR-1 gene, having only one variable site detected, also showed the lowest values for θ and π, CRSP38 had intermediate values, and PGIP had the highest detected. showed statistically significant values for Tajima's D in the Nacional and Pound populations and the combined data and a marginal value for French Guiana, and had a statistically significant Fu and Li's F* for the combined data. In PGIP, a statistically significant Tajima's D was calculated for the French Guiana population, but the other two populations, which had more variable sites than haplotypes, had lower, insignificant values.

Having identified the haplotypes for each gene within the populations, we generated maximum-likelihood trees for CRSP38 (Fig. 1A) and PGIP (Fig. 2A). With the sequences being so closely related, the CRSP38 tree has low bootstrap support at many nodes, although there is some support for the separation of haplotypes 1 and 2 from the rest of the tree. The PGIP is somewhat better supported, with low support for differentiating the clades containing haplotypes 7, 10, 11, and 12 and haplotypes 1, 2, 3, 4, and 5. We graphed the frequency of each haplotype within each population for CRSP38 and PGIP (Figs. 1B and 2B). The most common haplotype within each population differs noticeably for PGIP, while for CRSP38 the distribution for the Pound and Nacional populations are similar.

We next analyzed the polypeptides encoded by each haplotype to determine how many protein sequences exist within each population (Tables 5 and 6). We detect 6 and 7 polypeptide variants for CRSP and PGIP, respectively. CRSP38 haplotypes 1 and 2 do encode polypeptides differing at two amino acids. Interestingly for PGIP, coding variant 1 is most often encoded by a different haplotype in each population. However, haplotypes 2 and 4, common in certain populations, do encode different polypeptides.

Functional domain analysis

The NCBI CDSearch algorithm was used to identify functional domains within the coding sequences of the three genes described above. Figure 3A shows predicted domain structure of the PR1 gene. The majority of the gene's length is part of a predicted SCP/PR1-like domain. The only variable site detected sits close to center of this domain. The detected variation leads to two isoforms that differ in the presence of an isoleucine (I) or valine (V) residue, both of which are nonpolar. As this site was the only detected variation within the gene after surveying 213 genotypes, it appears this gene is highly conserved. The function of the PR1 family is not well understood, but the lack of variation suggests that it may not interact with a rapidly evolving pathogen molecule. Isoleucine and valine have structurally similar sidechains, so the difference we do detect may not have an effect on its biochemical activity.

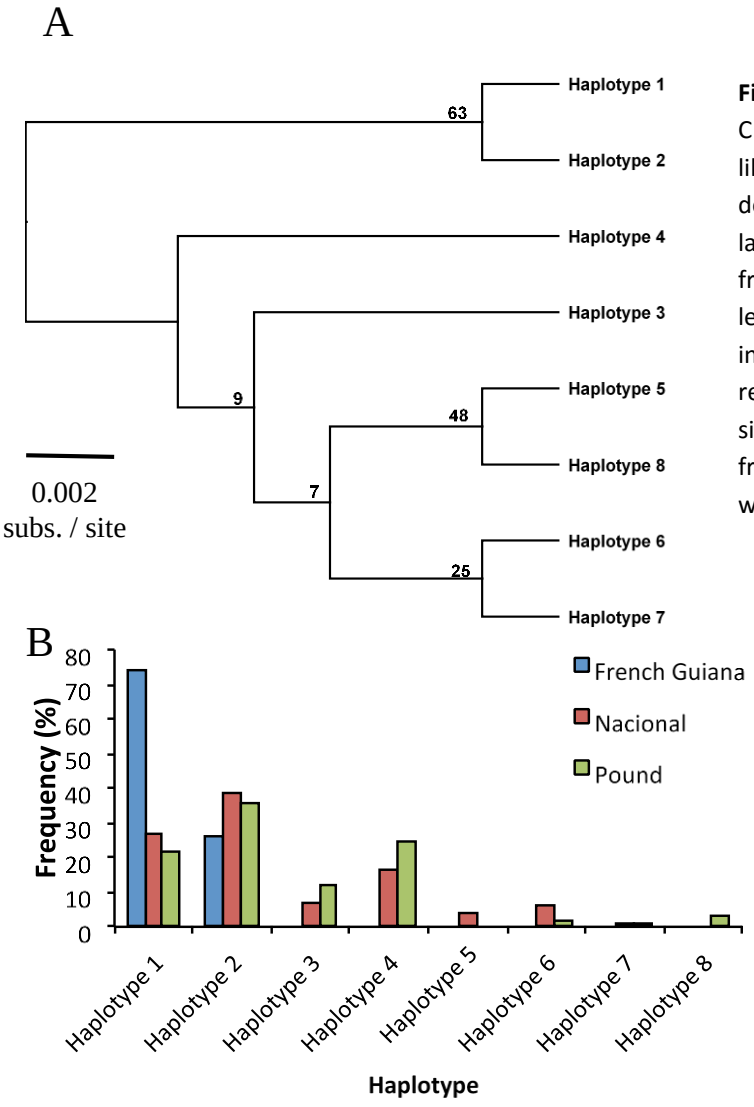


Figure 4.1 - Haplotype analysis of CRSP38 gene. A) Maximum-likelihood phylogenetic tree of detected CRSP38 haplotypes. Node labels represent bootstrap support from 100 replicates. Branch lengths represent genetic distance in substitutions per site. Scale bar represents 0.002 substitutions per site. B) Bar graph displaying frequency of each haplotype within each population.

Table 4.5 - Haplotype coding variation for CRSP38.

	Amino acids encoded at variable sites				
Haplotype ID	Site 212	Site 404	Site 482	Variable AA Seq	Coding Variant ID
Criollo Ref	Q	Q	S	QQS	3
1	Q	Q	N	QQN	1
2	Q	R	S	QRS	2
3	Q	Q	S	QQS	3
6	Q	Q	S	QQS	3
4	R	Q	S	RQS	4
5	R	R	S	RRS	5
7	R	Q	N	RQN	6
8	R	Q	N	RQN	6

Figure 3B shows the domain prediction for CRSP38. Two of three variable sites resulting in coding changes were located within a copy of a Stress/Antifungal domain (pfam 01657), while the third fell in the gap between the two domains. The synonymous mutation was also positioned toward the end of the first of these two domains. As glutamine (Q) has a polar sidechain and arginine (R) has a basic sidechain, these coding variants may be more likely to contribute to different biochemical activity of the isoforms. Asparagine (N) and serine (S) are both polar, so this difference is less likely to contribute to differences in biochemical activity. Cysteine residues within the structure of proteins with these antifungal domains have been shown to contribute to their structure and antifungal properties, but we did not detect any variation affecting cysteine residues.

Figure 3C shows domain prediction for PGIP. CDSearch detected 6 types of LRR domains within the gene sequence, some of which partially overlap. All five of the coding variants detected in the PGIP sequence sit within the longest of these predicted LRR domains, LRR_RI (cd00116). As LRR domains are known to have a role in facilitating protein-protein interactions, these variants may confer differential ability to bind to and inhibit pathogenic polygalacturonases. Three of the coding changes result in amino acids within the same class

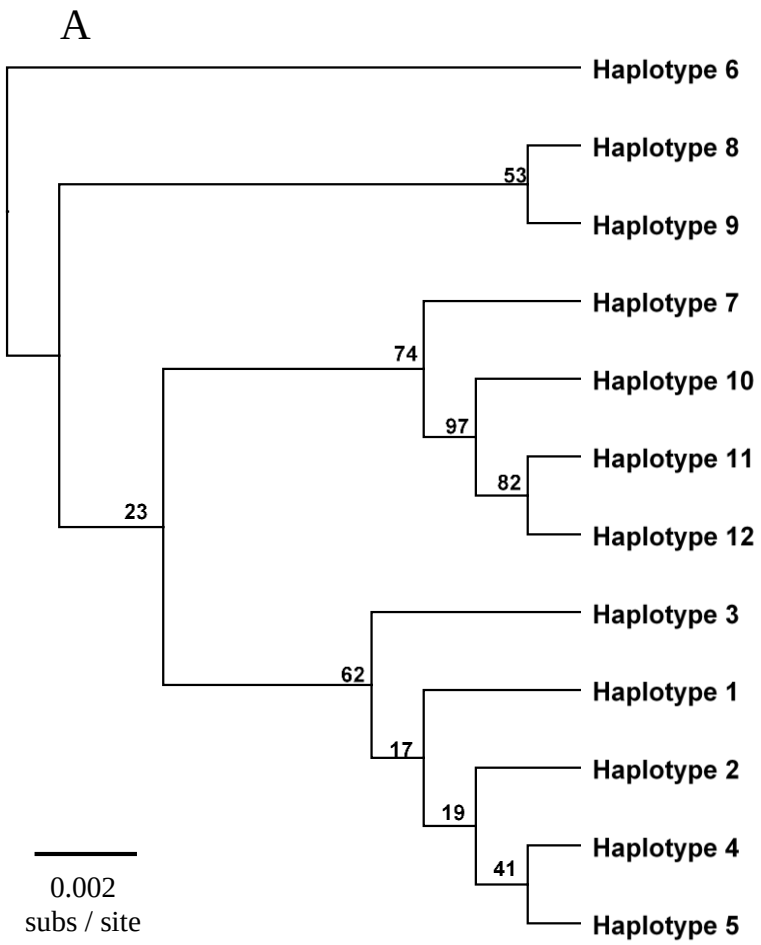


Figure 4.2 - Haplotype analysis of PGIP gene. A) Maximum-likelihood phylogenetic tree of detected PGIP haplotypes. Node labels represent bootstrap support from 100 replicates. Branch lengths represent genetic distance in substitutions per site. Scale bar represents 0.002 substitutions per site. B) Bar graph displaying frequency of each haplotype within each population.

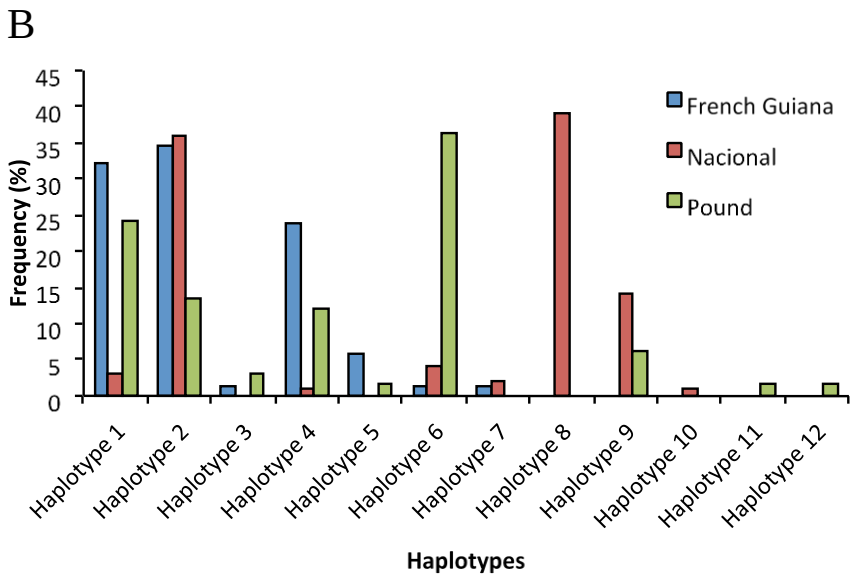


Table 4.6 - Haplotype coding variation for PGIP.

	Amino acids encoded at variable sites						
Haplotype ID	Site 562	Site 604	Site 628	Site 650	Site 851	Variable AA Seq	Coding Variant ID
Criollo Ref	V	R	S	N	V	VRSNV	7
1	M	R	S	T	V	MRSTV	1
6	M	R	S	T	V	MRSTV	1
8	M	R	S	T	V	MRSTV	1
12	M	R	S	T	V	MRSTV	1
2	M	G	S	T	A	MGSTA	2
3	M	G	A	T	V	MGATV	3
4	M	R	A	T	V	MRATV	4
5	M	G	S	T	V	MGSTV	5
7	M	R	S	T	A	MRSTA	6
9	V	R	S	N	V	VRSNV	7
10	V	R	S	N	V	VRSNV	7
11	V	R	S	N	V	VRSNV	7

(methionine (M) and valine (V) are both nonpolar, asparagine (N) and threonine (T) are both polar, alanine (A) and valine (V) are both nonpolar. However, arginine (R) has a basic sidechain and glycine (G) has a nonpolar sidechain, so the mutation at site 604 may have more impact on the protein's function. Also, the serine (S) to alanine (A) switch results in the change of a polar to nonpolar amino acid, and could be more likely to affect structure. All but one of the noncoding variant sites also sit within predicted LRR domains.

Discussion

Our analysis, while small in scale, was successful at identifying cacao loci with highly variable nucleotide and polypeptide sequences. Screening cacao orthologs of genes identified by

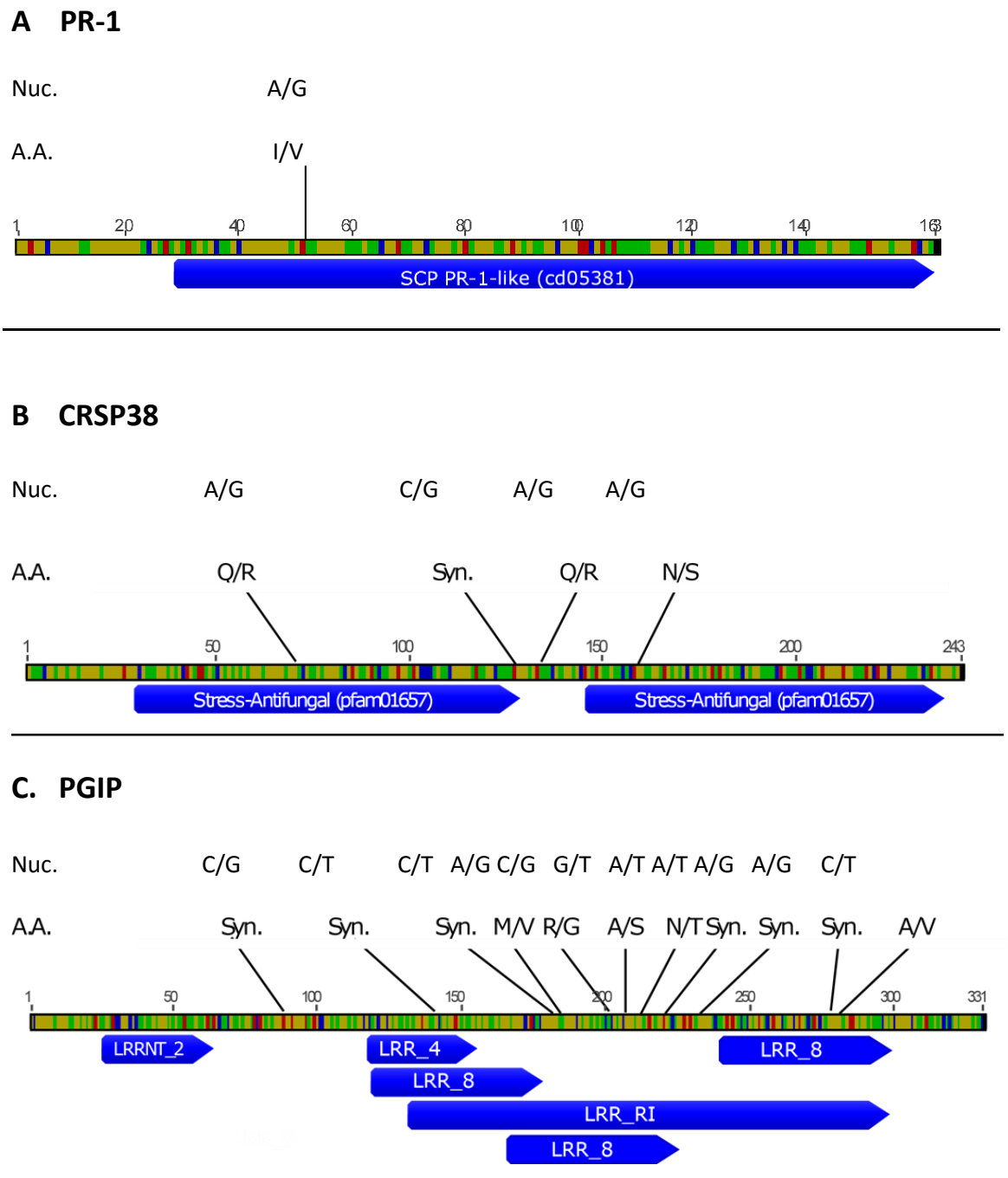


Figure 4.3 - Positioning of variable sites within predicted protein domains for A) PR-1, B) CRSP38, and C) PGIP. Sequence depicted is the amino acid sequence, with color coding representing polarity. Domain IDs for PR-1 and CRSP-38 are shown in annotation arrows. Domain database IDs for PGIP are as follows: LRRNT 2 (pfam 08263), LRR 4 (pfam 12799), LRR RI (cd00116), LRR 8 (pfam 13855).

Marden et al.'s tropical tree R gene genetic diversity assessment appears to be a viable strategy for discovering variation in cacao; however, a larger set of genes must be analyzed to assert with statistical power that variable proteins in other tropical trees also tend to be more variable in cacao.

Our assessment of genetic diversity in the three populations largely supports previously noted trends. Earlier analysis of the French Guianan population showed it to be the least diverse of a group of surveyed populations (Zhang et al., 2009). Here, we show that it contained variation at only 7 of the 16 sites found to be variable across our three populations. Interestingly, the two more variable populations shared almost all variable loci. Only one site, a synonymous variant in Nacional and not in Pound, differentiated the populations. This suggests that the majority of the detected variation predates germplasm dispersal, as it appears in populations believed to have long been geographically isolated (Motamayor et al., 2008). However, recombination has shuffled the alleles, creating varied haplotypes and altering frequencies of these haplotypes between populations. In PGIP in particular, differences in haplotype frequencies are striking. PGIP haplotypes 1, 6, 8, and 12 all encode the same polypeptide sequence, and the phylogenetic tree of haplotypes suggests that these emerged separately through convergent evolution. In total, this suggests the encoded protein could be functionally important in responding to some pressure that has become ubiquitous in cacao populations after the populations became reproductively isolated.

Tajima's D is a commonly used statistical test for identifying non-neutral evolution of a sequence. However, signatures of balancing or purifying selection can be muddied by demographic history of populations such as migration, changes in population size, and bottlenecks (Maruyama and Fuerst, 1985; Schmidt and Pool, 2002; Ramírez-Soriano et al., 2008). The statistically significant Tajima's D that we see for CRSP38 and PGIP could therefore be true signatures of balancing selection or artifacts of cacao domestication. For both genes, the existence of haplotypes with frequencies >10% and which encode different polypeptide sequences could suggest these variants are being selectively maintained for recognition of pathogen proteins. The more stringent Fu and Li's F*(Fu and Li, 1993) gave a significant value for all combined CRSP38 sequences, and was marginal for individual populations for CRSP38 and PGIP.

Pathogenesis-related genes, and PR-1 in particular, have long been studied as indictors of activation of the plant biotic stress response (van Loon and van Strien, 1999; van Loon et al., 2006). They belong to the SCP/TAPS family which play roles in host-pathogen interactions across eukaryotes, as well as the cysteine-rich secretory protein superfamily, but their function in plants is still not known (Cantacessi et al., 2009). Overexpression of PR-1 proteins in plants has provided some evidence of enhanced resistance to oomycete pathogens through an unknown mechanism (Sarowar et al., 2005). Conservation of PR-1-like sequences across eukaryotes indicates that they may serve an evolutionarily conserved role in the biology of higher organisms. Our analysis here focused on one of 14 PR-1 family members in the Criollo genome, revealing a single variable site which existed in only one of the three surveyed populations and a low frequency.

CRSP38 is the cacao ortholog of a gene identified and described in *Ginkgo biloba* (Sawano et al., 2007*)*. It contains cysteine-rich motifs which were shown to form disulfide bridges needed to form its active structure, and a group of arginine residues in an alpha helix were presumed to play a role in binding fungal cell wall components (Miyakawa et al., 2009). It has since been shown to interact with mannan in fungal walls, inhibiting growth of pathogenic *Fusarium* (Miyakawa et al., 2014). Interestingly two coding variants we detected in the CRSP38 protein caused a change from an arginine to a glutamine reside, but neither sat in alpha helixes predicted by EMBOSS.

Polygalacturonase inhibitor proteins bind to and inhibit the cleavage ability of pathogenic polygalacturonases, preventing degradation of the pectin network in the plant cell wall (Howell and Davis, 2005). Genes in the family exhibit varied expression dynamics, some undergoing induction by wounding, infection, and development of fruit (Yao et al., 1999; Ferrari et al., 2003; Howell and Davis, 2005). Overexpression of a pear PGIP in tomato reduced lesion development after inoculation with *B. cinerea* reduced fungal growth, demonstrating direct activity of PGIPs in limiting infection severity (Powell et al., 2000). Our investigation of genetic diversity in a cacao PGIP revealed 11 SNPs in its 993 bp sequence. Five non-synonymous mutations cluster within a single predicted leucine-rich repeat (LRR) domain. LRRs frequently participate in protein-protein interactions (Bella et al., 2008), making the detected coding variation more interesting. The variation we detect may be a sign of an arms race in plant

evolution, driving evolution of polypeptide sequences capable of recognizing rapidly evolving pathogenic polygalacturonases.

Overall, the study offers preliminary support to the approach of searching for variation in orthologs of genes measured to have high proportions of non-synonymous to synonymous variation. Pursuing this strategy. we were able to identify 15 SNPs in ~1700 bp (the CRSP38 and PGIP sequence lengths), identifying polypeptide variants that should be functionally validated to confirm differential roles in interacting with pathogen proteins. Further, using sequence variability to detect directional and balancing selection is a useful tool for identifying loci encoding proteins which are involved in host-pathogen interaction (Karasov et al., 2014). Given that interactions between cacao and its pathogens are poorly understood (Gutierrez et al., 2016), analysis of cacao gene sequences could provide a shortcut for identifying key components of interactions. After validation, this approach would be a powerful strategy to adopt into breeding programs, selecting for haplotypes required for successful protection against key pathogens.

References

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JSS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of *Theobroma cacao*. Nature Genetics **43:** 101-108

Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. Cellular and Molecular Life Sciences **65:** 2307-2333

Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, Matthews JB, Loukas A, Hofmann A, Otranto D, Sternberg PW, Gasser RB (2009) A portrait of the “SCP/TAPS” proteins of eukaryotes — Developing a framework for fundamental research and biotechnological outcomes. Biotechnology Advances **27:** 376-388

Cheeseman EE (1944) Notes on the nomenclature, classification, and possible relationships of cocoa populations. Tropical Agriculture **21:** 144-159

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2012) Geneious v5.6, Available from http://www.geneious.com.

Ferrari S, Vairo D, Ausubel FM, Cervone F, De Lorenzo G (2003) Tandemly Duplicated Arabidopsis Genes That Encode Polygalacturonase-Inhibiting Proteins Are Regulated Coordinately by Different Signal Transduction Pathways in Response to Fungal Infection. The Plant Cell **15:** 93-106

Fu Y-X (1995) Statistical properties of segregating sites. Theoretical Population Biology **48:** 172-197

Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. Genetics **147:** 915-925

Fu YX, Li WH (1993) Statistical Tests of Neutrality of Mutations. Genetics **133:** 693-709

Guiltinan M, Verica J, Zhang D, Figueira A (2008) Genomics of Theobroma cacao, "The foods of the gods". *In* PM Moore, R Ming, eds, Genomics of tropical crop plants. Springer, New York

Gutiérrez O A, Campbell A S, and Phillips-Mora W (2016). Breeding for disease resistance in cacao. In *Cacao Diseases* (pp. 567-609). Springer International Publishing.

Howell JT, Davis M (2005) Plant defense mechanisms against fungal pathogens: polygalacturonase inhibitor proteins. Canadian journal of plant pathology **27:** 5-15

Huang X, Han B (2014) Natural Variations and Genome-Wide Association Studies in Crop Plants. Annual Review of Plant Biology **65:** 531-551

Hughes A L, and Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**(6186), 167-170

Karasov T L, Horton M W and Bergelson, J (2014). Genomic variability as a driver of plant–pathogen coevolution?. *Current opinion in plant biology*, **18**, 24-30.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451-1452

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. Nucleic Acids Research **43:** D222-D226

Maruyama T, Fuerst PA (1985) POPULATION BOTTLENECKS AND NONEQUILIBRIUM MODELS IN POPULATION GENETICS. II. NUMBER OF ALLELES IN A SMALL POPULATION THAT WAS FORMED BY A RECENT BOTTLENECK. Genetics **111:** 675-689

Michelmore R W, and Meyers B C (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome research*, **8**(11), 1113-1130

Miyakawa T, Hatano K-i, Miyauchi Y, Suwa Y-i, Sawano Y, Tanokura M (2014) A Secreted Protein with Plant-Specific Cysteine-Rich Motif Functions as a Mannose-Binding Lectin That Exhibits Antifungal Activity. Plant Physiology **166:** 766-778

Miyakawa T, Miyazono K-i, Sawano Y, Hatano K-i, Tanokura M (2009) Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. Proteins: Structure, Function, and Bioinformatics **77:** 247-251

Motamayor J, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley S, Zheng P, Utro F, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler B, Stack J, Feltus F, Mustiga G, Amores F, Phillips W, Marelli JP, May G, Shapiro H, Ma J, Bustamante C, Schnell R, Main D, Gilbert D, Parida L, Kuhn D (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology **14:** R53

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao L*). PLoS ONE **3:** e3311

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research **30:** e36

Pound FJ (1943) Cacao and witches' broom disease (Marasmius perniciosa). Report on a recent visit to the Amazon territory of Peru, September, 1942 - February, 1943. Government Print, Port of Spain

Powell ALT, van Kan J, ten Have A, Visser J, Greve LC, Bennett AB, Labavitch JM (2000) Transgenic Expression of Pear PGIP in Tomato Limits Fungal Colonization. Molecular Plant-Microbe Interactions **13:** 942-950

Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical Power Analysis of Neutrality Tests Under Demographic Expansions, Contractions and Bottlenecks With Recombination. Genetics **179:** 555-567

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics **16:** 276-277

Sarowar S, Kim YJ, Kim EN, Kim KD, Hwang BK, Islam R, Shin JS (2005) Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. Plant Cell Reports **24:** 216-224

Sawano Y, Miyakawa T, Yamazaki H, Tanokura M, Hatano K-i (2007) Purification, characterization, and molecular gene cloning of an antifungal protein from Ginkgo biloba seeds. Biological Chemistry **388:** 273

Schmidt D, Pool J (2002) The effect of population history on the distribution of the Tajima's D statistic. *In*. New York: Cornell University Press

Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A (2006) Genetic Diversity and Natural Population Structure of Cacao (Theobroma cacao L.) from the Brazilian Amazon Evaluated by Microsatellite Markers. Conservation Genetics **7:** 13-24

Stephens M, Smith NJ, Donnelly P (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. The American Journal of Human Genetics **68:** 978-989

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology **44:** 135-162

van Loon LC, van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiological and Molecular Plant Pathology **55:** 85-97

Wood GAR, Lass R (2008) Cocoa. Technical, New York

Yao C, Conway WS, Ren R, Smith D, Ross GS, Sams CE (1999) Gene encoding polygalacturonase inhibitor in apple fruit is developmentally regulated and activated by wounding and fungal infection. Plant Molecular Biology **39:** 1231-1241

Zhang D, Arevalo-Gardini E, Mischke S, Zuniga-Cernades L, Barreto-Chavez A, Aguila JAD (2006) Genetic Diversity and Structure of Managed and Semi-natural Populations of Cocoa (Theobroma cacao) in the Huallaga and Ucayali Valleys of Peru. Annals of Botany **98:** 647-655

Zhang D, Mischke S, Johnson ES, Phillips-Mora W, Meinhardt L (2009) Molecular characterization of an international cacao collection using microsatellite markers. Tree Genetics & Genomes **5:** 1-10

Chapter 5: Protocol: Transient expression system for functional genomics in the tropical tree *Theobroma cacao* L.

Published as:

Fister, A. S., Shi, Z., Zhang, Y., Helliwell, E. E., Maximova, S. N., & Guiltinan, M. J. (2016). Protocol: transient expression system for functional genomics in the tropical tree Theobroma cacao L. *Plant methods*, *12*(1), 1.

Abstract

Theobroma cacao L., the source of cocoa, is a crop of significant economic value around the world. To facilitate the study of gene function in cacao we have developed a rapid *Agrobacterium*-mediated transient genetic transformation protocol. *Agrobacterium tumefaciens* cultures are induced then vacuum-infiltrated into cacao leaves. Transformation success can be gauged 48 hours after infiltration by observation of green fluorescent protein (GFP) and by qRT-PCR. Leaves expressing transgenes of interest can be used in subsequent functional genetic assays such as a pathogen bioassay, metabolic analysis, gene expression analysis etc. This transformation protocol can be carried out in one day, and the transgene expressing leaf tissue can be maintained in petri dishes for 5-7 days, allowing sufficient time for performance of additional downstream gene functional analysis. Here we also present a pathogen infection bioassay used to assess gene function after transient transformation of leaves.

Background

Theobroma cacao L., the source of cocoa, is a tree crop of great international economic importance and the center of the multi-billion-dollar chocolate industry. While the tree is native to the Amazon basin (Motamayor et al., 2008), approximately 70% of cocoa is now produced in

West Africa, with the remainder coming from South America and Southeast Asia (Wood and Lass, 2008; Lopes et al., 2011). Each year the crop suffers significant losses to a variety of fungal, oomycete, and viral diseases (Guiltinan et al., 2008), resulting in significant financial loss for cacao farmers and nations exporting cocoa. Cacao research has benefited from the recent publication of the genome sequences of two genotypes (Argout et al., 2011; Motamayor et al., 2013). Availability of these data increases the speed with which putatively important cacao genes can be functionally characterized, which could lead to crop improvement through application of novel breeding strategies or biotechnological approaches (Guiltinan and Maximova, 2015), although progress with long-generation crops is inherently slow. Accordingly, development of strategies enabling gene characterization is important to expedite the process of genetic improvement of cacao.

Agrobacterium-mediated transient and stable plant transformation techniques were developed to enable the introduction of recombinant DNA into plant cells in plants (Schell, 1987; Janssen and Gardner, 1990). Whereas transient expression is largely the result of transcription and translation of non-integrated T-DNA, stable transformation by definition implies the integration of T-DNA into the host genome (Lacroix and Citovsky, 2013). Transiently transfected plants typically show a peak in expression 2-4 days after infection with *Agrobacterium* that subsequently declines (Lacroix and Citovsky, 2013), while stable transformation is typically achieved through selection and culturing of transformed tissue, and leads to persistent expression of transgenes (Křenek et al., 2015). If germ line cells are transformed, integration of T-DNA is heritable (Bent, 2006). While stable transformation is essential for applications in crop improvement, transient transformation enables rapid testing of gene function, and is therefore an invaluable tool for plant genetics research. Both transformation strategies have been applied to a number of tree crops including cacao (Maximova et al., 2003; Maximova et al., 2006; Mejia et al., 2012; Shi et al., 2013; Mejía et al., 2014; Zhang et al., 2014; Fister et al., 2015; Helliwell et al., 2015), and have been applied to enhancement the of disease resistance, abiotic stress response, improvement of quality traits, and the general study of functional genetics (Gambino and Gribaudo, 2012).

Traditional breeding strategies for tree crops are laborious and expensive. For cacao, generation of new varieties through breeding programs can take 15-20 years (Lopes et al., 2011). A strategy for generation of stable transgenic cacao trees was previously published

(Maximova et al., 2003), however even this process takes several years to produce a mature tree that could be used to assay experimentally the effect of a transgene's overexpression or knockdown. The transient transformation protocol and subsequent functional analysis described here can be performed in a week, and has been used to demonstrate effect of overexpression (Mejía et al., 2014; Fister et al., 2015) and knockdown (Shi et al., 2013) of cacao genes with roles in defense, expression of non-native phosphatidylinositol 3-phosphate binding proteins in cacao (Helliwell et al., 2015), and the function of a transcription factor controlling embryogenesis (Zhang et al., 2014).

Here we present the protocol for *Agrobacterium*-mediated transient transformation of detached leaf tissue of *Theobroma cacao.* Growth conditions described here were extensively tested to optimize transformation efficiency. The strategy enables functional gene characterization to be performed in a matter of weeks, rather than the years that would be required to generate a stably transgenic cacao tree.

Experimental Design

The protocol described here has been used to rapidly screen vectors to measure the effect of gene overexpression or knockdown in cacao leaf tissue (Shi et al., 2013; Mejía et al., 2014; Zhang et al., 2014; Fister et al., 2015; Helliwell et al., 2015). Prior to transformation, binary vector constructs were transferred into competent *Agrobacterium* strain AGL1 as previously described (Maximova et al., 2003). Typically, the experiment is performed using two vectors: an experimental construct and a control construct (typically pGH00.0126, GenBank: KF018696). Leaves are divided into two sections, one closer to the tip and one closer to the base, such that each leaf can be transformed with both constructs. Preliminary experiments have showed that transformation success usually does not differ significantly between the two sections of a given leaf (data not shown). The two sections of a leaf are simultaneously infiltrated by submerging leaf discs in cultures of *Agrobacterium* and applying a vacuum. Transformation success is evaluated 48 hours after infiltration by observing EGFP fluorescence. A leaf is only used for subsequent functional characterization of EGFP is uniformly present across >80% of the surface area of the control and experimental sections of a given leaf. A workflow diagram of the

transient transformation process is depicted in Fig. 1. It is important to note that efficiency of transformation varies significantly between leaves, and proper appraisal of leaf stage is critical for a successful experiment. At least 3 replicates per transgene are typically used for statistical power. In order to ensure that 3-5 leaf sections per construct are successfully transformed, we

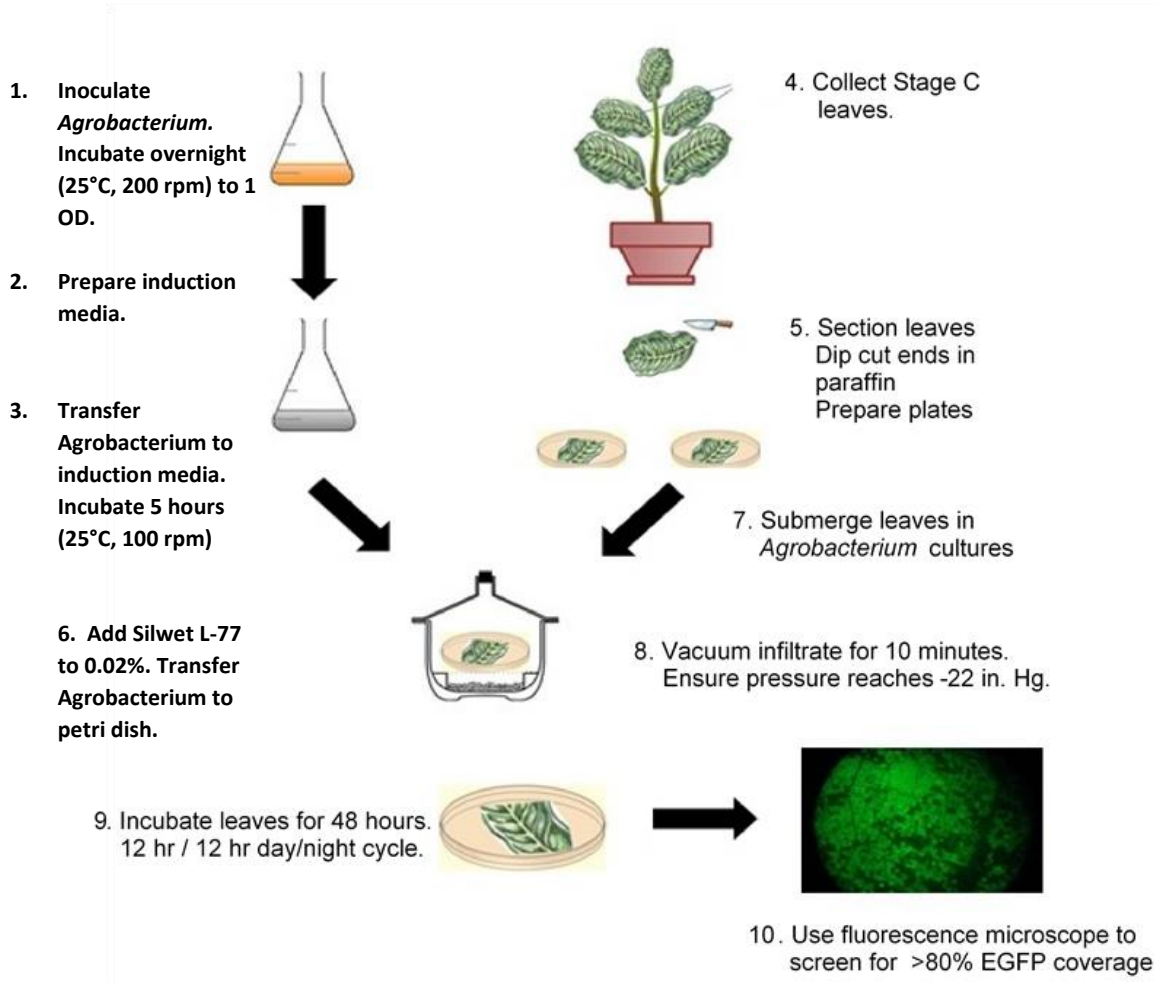


Figure 5.1 - Workflow diagram for transient transformation of cacao leaf tissue.

recommend infiltrating 8-10, anticipating several leaves will not pass the EGFP coverage threshold.

Cacao leaf stages were previously described (Mejia et al., 2012); however, as accurate determination of leaf stage is integral to successful transient transformation, we sought to more quantitatively describe the stages to enhance reproducibility of the protocol. In developing the protocol, we found that leaf age affected transformation efficiency, with both earlier and later developmental stages showing lower transformation success as measured by EGFP fluorescence. This resulted in our using Stage C leaves (Fig. 2A), which are expanded but still supple, for our transient transformation experiments. To demonstrate this observation, we transformed leaves of each stage, and 48 hours after infiltration, photographed EGFP fluorescence (Fig. 2B-F). To measure leaf toughness, we used a force gauge and performed a punch test on leaves of stages A through E. Fig. 2G shows the mean force to puncture, averaged across five leaves, for each leaf stage. Our protocols for collection and transformation and photographing of the five leaf stages, as well as the protocol for the force to puncture test, can be found in the supplemental files. The data indicates that early in their development (through stage C), leaves do not significantly increase in rigidity. Stage D and E leaves, however, are measurably more rigid. Therefore, it is essential to take into account both leaf color (stage C leaves are bronze to light green) and rigidity to select leaves most likely to be successfully transformed.

In order to evaluate the rate at which cacao leaves infiltrated with *Agrobacterium* become transformed, we monitored expression of an EGFP transgene over a time course after infiltration. Leaves were imaged using a fluorescence stereo-microscope. Images were acquired immediately after transformation and every three hours after bacterial infiltration (ABI) for the first 48 hours, and at hours 60, 84, 108, 132, and 156. No EGFP fluorescence was detected until 18 hours ABI. Fluorescence intensity increased until its peak at 45 hours ABI, remained high until 60 hours, and then steadily declined. EGFP fluorescence was quantified using ImageJ and is graphed as a percentage of the level detected at 45 hours ABI (Fig. 3). Because the intensity peaks approximately two days ABI, this time point was selected to evaluate transformation success before proceeding into subsequent experiments. Further, our earliest detection of transient expression at hour 18 was consistent with findings in tobacco

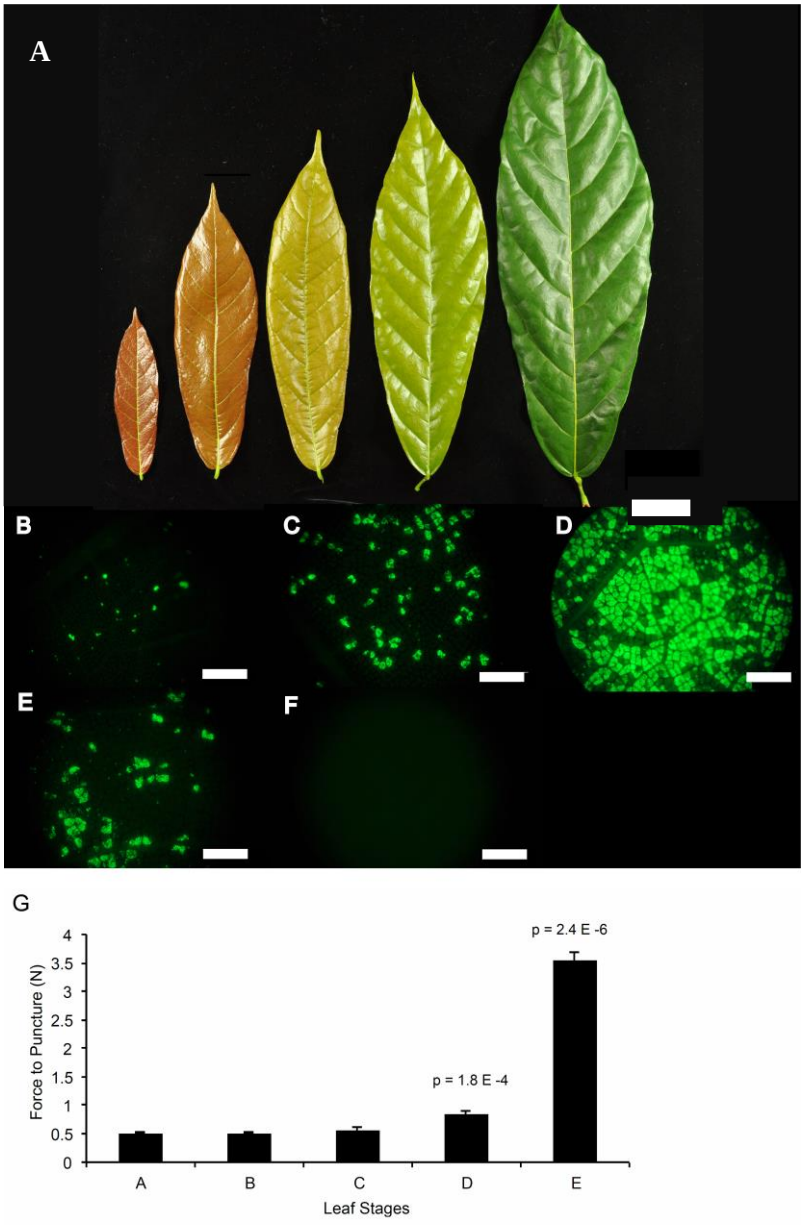


Figure 5.2 – Leaf stages and force to puncture measurements. **A)** Photograph displaying representative leaves of stages A (leftmost) to E (rightmost) collected from genotype Scavina 6. Scale bar represents 5 cm. **B-F)** Representative photographs of EGFP fluorescence taken 48 hours after infiltration of leaves (stages A-E) with *Agrobacterium.* Scale bars represent 1 mm. **G)** Measurement of force to puncture for each leaf stage. Bars represent mean of five measurements, each representing one leaf from that stage. Bars represent standard deviation across five replicates. T-test p values are shown above bars for Stage D and Stage E, which are comparisons of measurements of Stage C leaves with those of the older stages. Differences between Stage A and C and B and C were not significant.

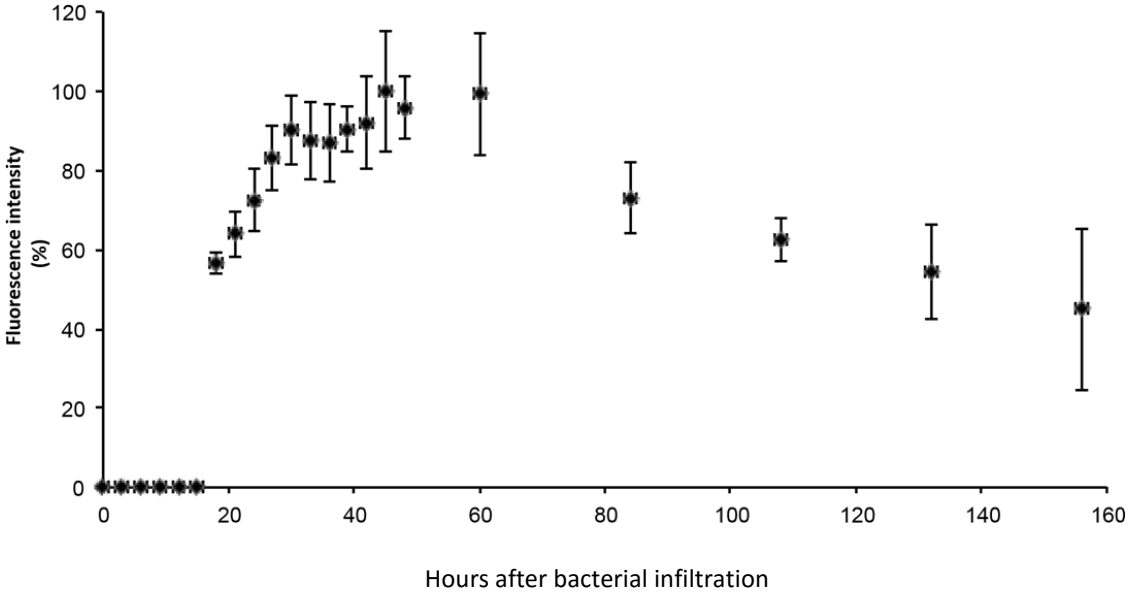


Figure 5.3 - Time course of EGFP fluorescence intensity after infiltration of leaf tissue with *Agrobacterium.* Fluorescence is expressed as a percentage of the intensity measured at hour 45, the peak time point. Error bars represent standard deviation calculated from three biological replicates.

(Narasimhulu et al., 1996), and peak expression in our time course is consistent with results from transient transformation of *Arabidopsis* (Nam et al., 1999).

While the protocol was optimized for transformation of Stage C leaves (Mejia et al., 2012) from genotype Scavina 6, it can be applied to other genotypes. Figure 4 includes photographs of stage C leaves from eight genotypes (Fig. 4A), as well as representative photographs showing transformation efficiency of these genotypes (Fig. 4 B-I). In Figure 4J, the transformation efficiency of each genotype was calculated and graphed relative to that measured in the Scavina 6 genotype. Our protocol for this genotype transformation optimization test, including calculation of transformation efficiency with ImageJ (Schneider et al., 2012), can be found in the supplemental files. While Scavina 6 exhibited the highest transformation efficiency, three other genotypes (CCN51, ICS1, TSH1188) had mean transformation efficiencies greater than 80%, suggesting that our protocol could likely be easily applied to these varieties. Physiological differences between leaves of different genotypes may contribute to decreased efficiency, and some alterations to the protocol may be necessary to

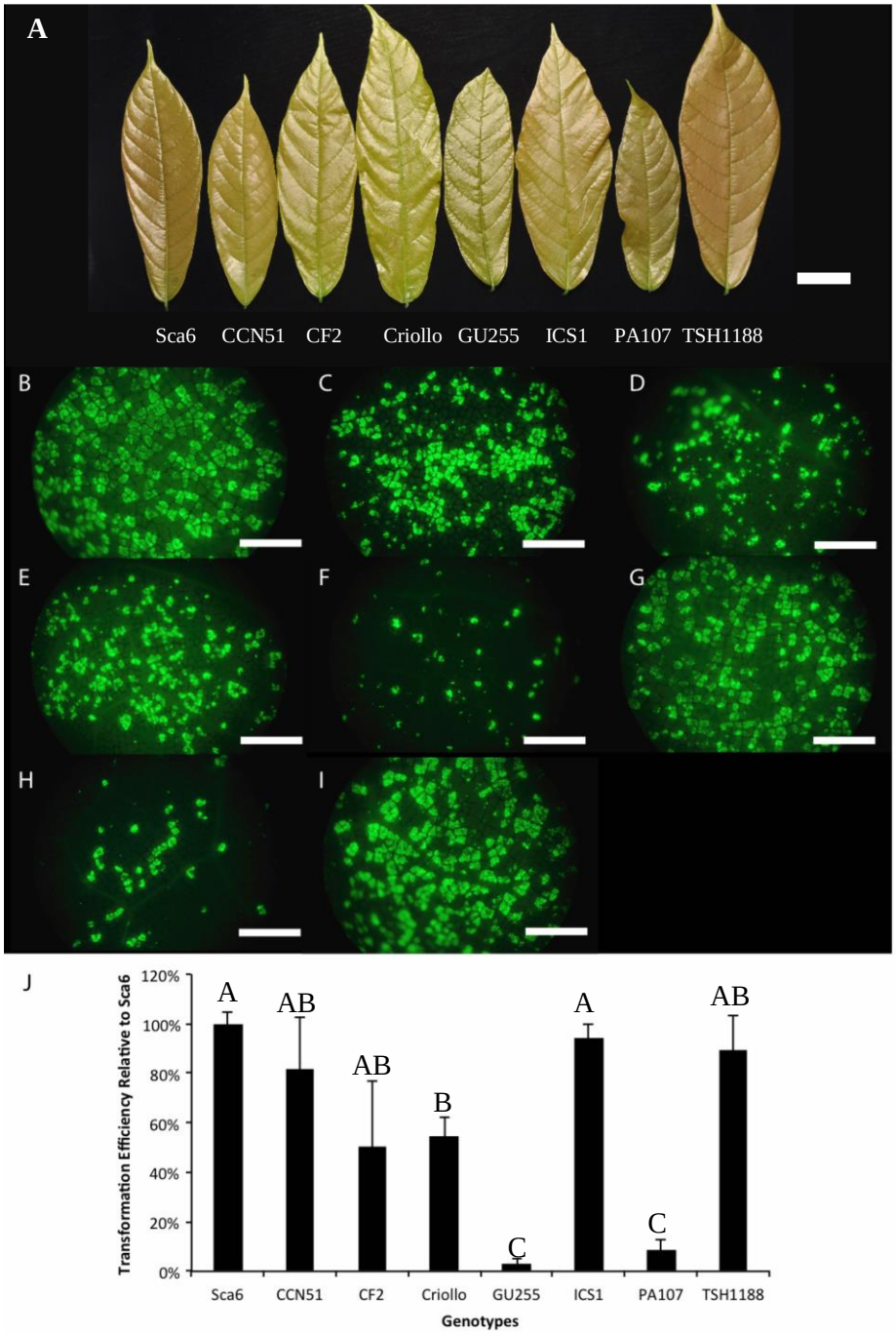


Figure 5.4 – Transformation of eight cacao genotypes. **A)** Photograph showing stage C leaves selected from eight cacao genotypes. Some genotype identifiers are abbreviated: Sca6 = Scavina 6, Criollo = B97-61/B2, ICS1 = Imperial College Selection 1. Scale bar represents 5 mm. **B-H)** Representative images of EGFP coverage 48 hours after agrobacterium infiltration using the eight genotypes shown in panel A. Scale bars represent 1 mm. **B)** Sca6; **C)** CCN51; **D)** CF2; **E)** Criollo; **F)** ICS1; **G)** GU255**; H)** PA107**; I)** TSH1188. **J)** Bar graph depicting transformation efficiency expressed as a percentage of that calculated for Scavina 6 samples. Error bars represent standard deviation calculated from three biological replicates. Bars labelled with the same letter are not statistically significant ($p > 0.05$).

overcome low efficiencies of the transformation-recalcitrant varieties. We have also previously noted that Scavina 6 leaves appear to remain green and survive longer in petri dishes than other genotypes (Fister et al., 2015), so it may be generally more suitable to long-duration experiments.

After identifying successfully transformed leaves, subsequent experiments including RNA extractions, pathogen inoculations, and lipid extractions can be performed, as have been described (Shi et al., 2013; Zhang et al., 2014; Fister et al., 2015; Helliwell et al., 2015). Leaves will show significant desiccation 5-7 days after being detached from plants; therefore, experiments should not require more 3-5 days after transformation success is confirmed. Other than this limitation, the transformation strategy can be widely applied to gene characterization studies. In addition to the transformation protocol, we also provide here a detailed methodology for infection of leaves with pathogen after transformation.

In Figure 5, we have included additional data demonstrating the effect of transient overexpression of a previously described cacao chitinase gene (Maximova et al., 2006). Our protocol for these experiments is available in the supplemental file. Two constructs were used for the transient transformation, pGH00.0126 (GenBank: KF018696), in which EGFP is driven by the CaMV 35S promoter, and another (pGAM00.0511, described in [15]) which has an additional cassette containing the cacao chitinase gene (Tc02_g003890) under the CaMV 35S promoter. Chitinase overexpression using this system resulted in decreased lesion size after infection with *Phytophthora tropicalis* (Fig. 5A-B), a decrease in the ratio of pathogen to cacao DNA detected in the tissue (Fig. 5C), and an approximately six-fold increase in chitinase transcript abundance as assessed by qRT-PCR (Fig. 5D).

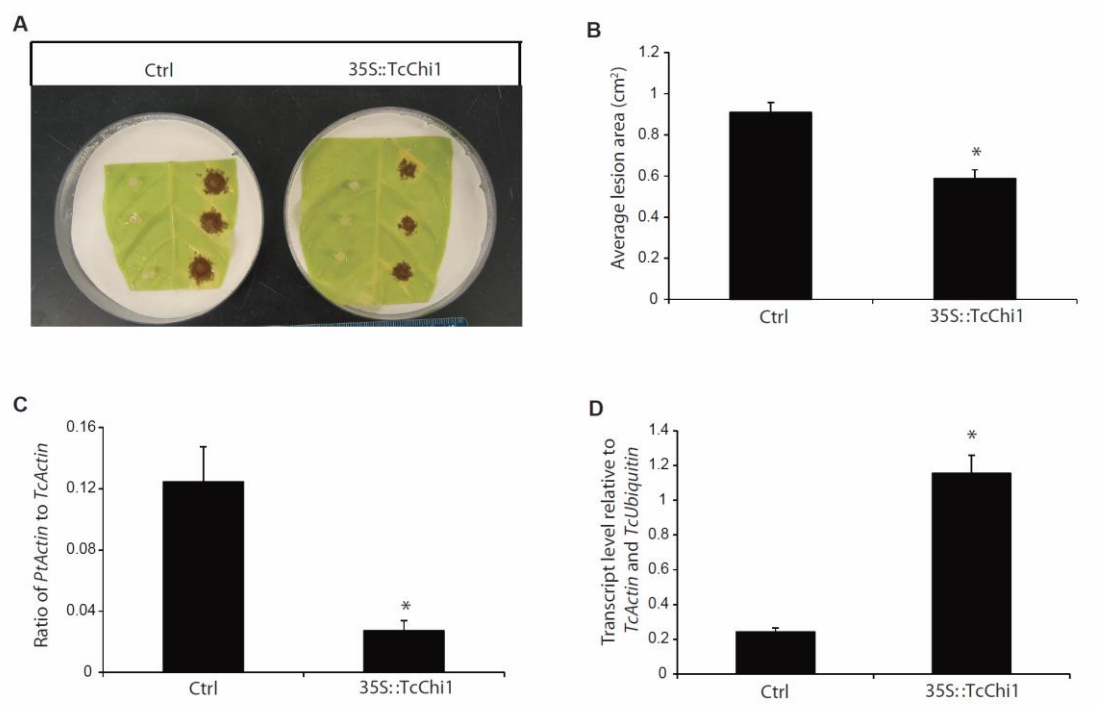


Figure 5.5 – Functional analysis of *TcChi1*. **A.** Representative images of lesions from control (Ctrl, transformed with pGH00.0126) and leaves transiently transformed to overexpress *TcChi1* two days after *Phytophthora tropicalis* inoculation. **B.** Average lesion areas from control and *TcChi1* overexpressing leaves were measured three days after inoculation using ImageJ. Bar charts represent the means ± SE of measurements from 12 lesion spots from four leaf discs of each genotype. **C.** Pathogen biomass was measured at the lesion sites by qPCR to determine the ratio of pathogen DNA to cacao DNA two days after inoculation. Bar charts represent four biological replicates, each with three technical replicates. **D.** qRT-PCR analysis of *TcChi1* transcript two days after vacuum infiltration. Data represent means ± SE of three biological replicates. The asterisk denotes a significant difference determined by single factor ANOVA ($p<0.05$).

Reagents and Equipment

For transformation:

- *Agrobacterium* is cultured in 523 media, and induced as previously described (Li et al., 1998). Recipes for these media can be found in Table 1.
- A Fast PES Filter unit (Thermo Scientific, Cat. No. 124-0045) is used to sterilize induction media.
- Before infiltration of leaves, Silwet L-77 (Lehle Seeds, Cat. No. VIS-01) is added to *Agrobacterium* cultures to act as a surfactant.

- After leaves are infiltrated with *Agrobacterium*, they are maintained in a controlled environment at 25°C with 50% relative humidity and a 12 hr / 12 hr light dark cycle. Light levels are maintained at 55 µmol m^{-2} s^{-1}, using fluorescent bulbs 4100K Kelvin ratings. Higher light levels did not affect transgene expression, but did lead to faster desiccation of leaves.
- Gast G582DX Vacuum Pump

Table 5.1 – Media recipes for *Agrobacterium* growth and induction and pathogen growth

523 Medium (1 Liter)	
Reagent	**Amount per Liter**
Sucrose	10 g
Casein enzymatic hydrolysate	8 g
Yeast extract	4 g
K_2HPO_4 anhydrous	2 g
$MgSO_4$ anhydrous	0.15 g
Notes: Add distilled water to 1 L. Adjust pH to 7.1 and autoclave.	
Induction Medium (recipe per 30 mL volume)	
Liquid ED (recipe described in (Maximova et al., 2005))	30 mL
0.1 M acetosyringone (Sigma Cat. # D134406)	30 µL
L-proline	0.00465 g
Notes: Adjust pH to 5.25 – 5.3 using 0.1 M KOH. Discard if pH exceeds 5.32. Do not adjust pH using HCl. Prepare induction medium on morning of leaf infiltration experiment. Use liquid ED less than 30 days old.	

20% V8 Media (1 Liter)	
Reagent	**Amount per Liter**
Bacto Agar	15 g
Calcium Carbonate ($CaCO_3$)	3 g
Campbell's V8 Vegetable Juice	200 mL
Notes: Add distilled water to 1 L. Adjust pH to 7.1 and autoclave. Shake frequently while pouring into petri dishes to maintain homogeneity of media. Pour about 20 mL of media into each plate to ensure that agar plugs and do not fall during leaf assay.	

- Science-Ware vacuum desiccator (Cat# 420270000)
- Whatman grade 5 qualitative filter paper, 90mm diameter discs (Cat# 1005-090)
- Sterile 100 mm x 20 mm petri dishes (Fisher Brand Cat# FB0875711Z)
- Paraplast Plus tissue embedding medium (McCormick Scientific Cat# 39503002)
- Orbital shaker
- General lab supplies: pipettors, pipette tips, Parafilm, paper towels

For pathogen bioassay:

- Pathogen subcultures (age depends on pathogen)
- Appropriate media for pathogen growth (recipe for 20% V8 media is listed in Table 1)
- Laminar flow hood
- Atomizer of sterile water
- 3 mm diameter cork borer, 6 mm diameter cork borer, 1.5 cm diameter cork borer
- General lab supplies: forceps, probe, petri dishes

Protocol

A. Preparation of *Agrobacterium* working stocks for transformation

Timing: Approximately 1 hour, plus overnight incubation

1. Prepare 523 media (see Reagents and Equipment).

2. *Agrobacterium* for transformations are cultured using working stocks at a known OD to ensure that cultures grow at consistent rates. To create working stocks, inoculate freezer stocks of AGL1 colony containing desired plasmid in 2 mL 523 medium with appropriate antibiotic and shake overnight at 200 rpm, 25^{o}C.

3. Measure OD at 600 nm. Let the culture grow until OD600 is 1, or dilute to 1 with 523 media if above. Take 750 µL culture and transfer into a sterile 1.5 mL tube. Add 250 µL of 60% glycerol. Mix well. Aliquot 100 µL of the mixture into cyrovial tubes. Store at -80 ^{o}C.

B. Day 1 – Inoculation and incubation of *Agrobacterium* culture

Timing: Approximately 10 minutes, plus overnight incubation

1. Thaw a 100 µL working stock of AGL1 for each desired plasmid.

2. Inoculate 90 µl of AGL1 stock into 30 ml of 523 media with Kanamycin (50 mg/ml) in a sterile 125 ml Erlenmeyer flask covered with aluminum foil.
3. Shake in the dark overnight at 200 rpm at 25°C (approximately 16 hours).

C. Day 2, Part I – Virulence Induction of *Agrobacterium culture*

Timing: Approximately 1 hour of active time, plus 5 hour incubation

1. For every 30 mL culture of Agrobacterium, prepare 30 mL of induction media (see Reagents and Equipment). Vacuum sterilize the induction media using Fast PES Filter unit (Thermo Scientific, Cat. # 124-0045).
2. Measure OD of overnight cultures at 600 nm. Use 523 media as a blank. Wait for all cultures to reach OD of 1. Remove those that have passed this point from the shaker to prevent overgrowth. If OD has passed 1.3, discard cultures. If OD is between 1 and 1.3, dilute to OD 1.0 with 523 media.
3. Transfer culture to a 50 mL centrifuge tube. Centrifuge the *Agrobacterium* at 1500 x g, 25^{o}C, for 17 minutes to pellet the bacteria.
4. Discard supernatant, gently pipette and vortex to re-suspend cultures using 30 mL of induction media and transfer to new 250 mL flasks. Ensure that pelleted bacteria are thoroughly suspended in the solution.
5. Shake at 100 rpm at 25^{o}C for 5 hours. During this step, collect leaves and prepare plates.

D. Day 2, Part II – Plate Preparation and Leaf Selection

Timing: Approximately 1 hour

Note: Plate preparation and leaf collection will take approximately an hour, so perform these steps about 4 hours after beginning *Agrobacterium* induction, typically early in the afternoon.

1. Place ten Paraplast Plus chips onto a glass petri dish and apply low heat (~56°C) until they melt.
2. For each plate, fold a paper towel into a square, and cut off the corners to fit it into a 100x20mm petri dish. Place Whatman #5 filter paper on top of the paper towel and gently press down to create a flat surface. Add 10ml of sterile water to the plate to maintain humidity.
3. Collect Stage C leaves from greenhouse grown plants. It is essential to the success of the experiment that leaves are soft and supple, and Stage C leaves are bronze to light green in color.

Cut the petiole to remove the leaf from the plant without damaging the leaf's surface area. Place the leaves in a sealable plastic bag containing wet paper towels to maintain humidity.

4. Cut leaves with a scalpel to produce leaf two sections. First, the tip and base of the leaf are removed (Fig. 6A). Next the leaf is divided into two sections of equal size. Ensure that each section is large enough to accommodate subsequent experiments (i.e. inoculation with pathogens). As leaves are cut, seal the cut edges by dipping into melted paraffin. This will limit desiccation from exposed veins. Place the leaf discs onto plates for temporary storage, abaxial side up (Fig. 6B) and close the plates. Let sit on the lab bench until induction of *Agrobacterium* is complete.

D. Day 2, Part III – Vacuum Infiltration

Timing: Approximately 1-2 hours, depending on replicate number

1. After 5 hours in the incubator, add pure Silwet L-77 to the *Agrobacterium* culture to a final concentration of 0.02% (for a 30 ml culture, this is 6µL of Silwet). Silwet L-77 is necessary for successful transformation. Our preliminary results indicated that higher Silwet L-77 concentrations do not increase transformation success rates.

2. Pour induced *Agrobacterium* suspension onto 100 mm x 20 mm petri dishes labeled with the construct name on the bottom of the plate as lids are removed during infiltration.

3. With lids removed, place the petri dishes of induced *Agrobacterium* into the desiccator.

4. Select a leaf section to be placed into each dish of *Agrobacterium,* abaxial side down*.* The other section of the same leaf should receive the other treatment. *Agrobacterium* containing control and experimental vectors are typically infiltrated into their respective leaf sections concurrently. Place the lid on the desiccator.

5. Vacuum-infiltrate the leaves.

    a. Turn the stopcock valve to open airflow between vacuum pump and desiccator. Start a timer as pressure begins to reduce.
    b. Ensure that the pressure reaches -22 in. Hg on pressure gauge. Wait 10 minutes. As leaves sit in vacuum, small air bubbles should appear at the edges of the leaf.
    c. Turn the stopcock valve to release vacuum inside desiccator.

6. Using separate tweezers and paper towels for each construct, gently remove the leaf disc from the desiccator, blot dry in paper towels, and hold up to light to look for flooding of cells to assess infiltration success.

a. Spots of translucence on the underside of the leaf indicate successful infiltration, which correlates with high EGFP and transgene expression on day 4.
b. If none of the leaves have noticeable flooding, the transformation experiment will likely be unsuccessful.

7. Place the leaf abaxial side up onto its petri dish from step 2. Ensure complete contact between the leaf and the filter paper by placing one corner down first and slowly lowering the leaf so that it adheres to the filter paper. Place lid on the petri dish, and seal it with parafilm.
8. Repeat steps 2-6 for all remaining leaf discs.
9. Incubate leaves in a growth chamber at 25^{o}C with 12 hr: 12 hr light dark cycle for two days with a light intensity of 55 μmol m^{-2} s^{-1}. Higher light levels were found to lead to faster deterioration of leaf tissue.

E. Day 4 – Evaluating EGFP expression

Timing: Approximately 1 hour

1. Forty-eight hours after infiltration, gather leaf tissue.
2. Using a fluorescence microscope, scan the surface area of each leaf for EGFP as previously described (Fister et al., 2015). In order for leaves to be useful for subsequent experiments, at least 80% of the surface area of the leaf should fluoresce, and there should be no large patches of tissue not expressing EGFP. Representative image of EGFP fluorescence over a small area of leaf tissue is included in Fig. 6C. Coverage across the entire surface of the leaf should match this level of expression.
3. Leaf-to-leaf physiological variability may contribute to some variability in downstream experiments. Consequently, only pairs of leaf sections that both pass the EGFP threshold should be retained. Any pairs of leaves where either has less than 80% of its surface showing EGFP can be discarded.

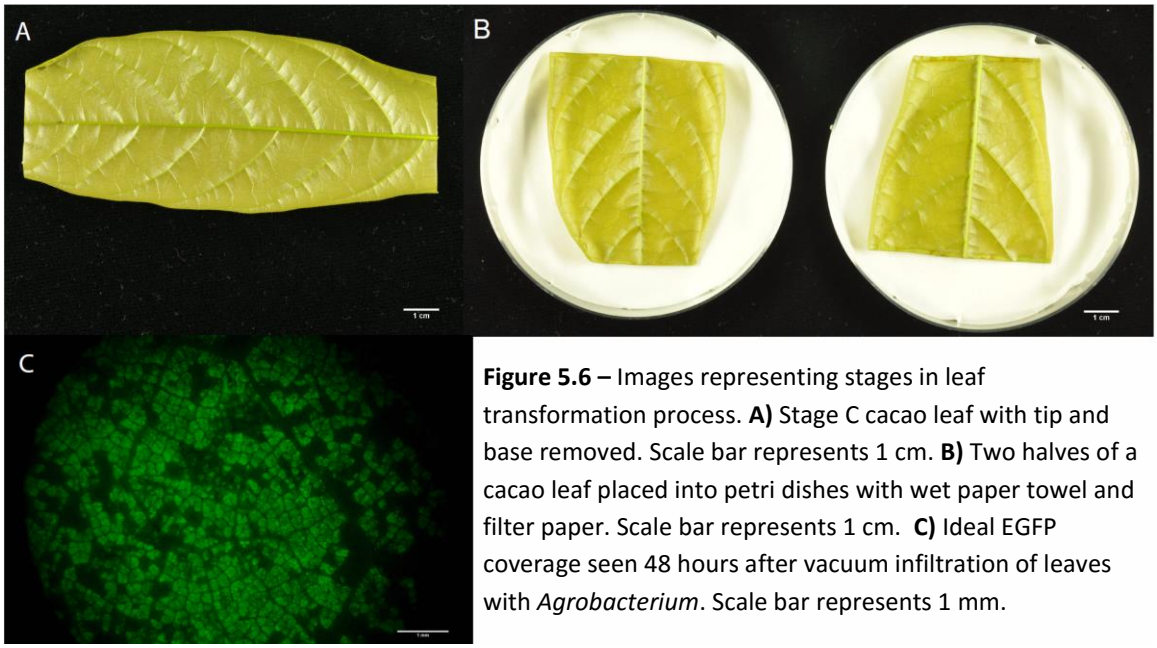


Figure 5.6 – Images representing stages in leaf transformation process. **A)** Stage C cacao leaf with tip and base removed. Scale bar represents 1 cm. **B)** Two halves of a cacao leaf placed into petri dishes with wet paper towel and filter paper. Scale bar represents 1 cm. **C)** Ideal EGFP coverage seen 48 hours after vacuum infiltration of leaves with *Agrobacterium*. Scale bar represents 1 mm.

F. Demonstration of the utilization of transient transformation protocol for functional studies of candidate genes for disease resistance

***Phytophthora* bioassay**

Part I – Subculturing pathogen

Timing: Approximately 15 minutes

1. Sterilize a laminar flow hood with UV light for 2 minutes, and wipe the area with 70% ethanol.
2. Sterilize the 6 mm diameter cork borer and forceps using 70% ethanol and flame. Let cool briefly.
3. Use the cork borer to create agar plugs in a mature plate of pathogen (Fig. 7A-B).
4. Transfer agar plugs, pathogen side down, to a new plate of V8 media (Fig. 7C).
5. Incubate the leaves at 28^{o}C, 12:12 light/dark cycle for 48 hours.

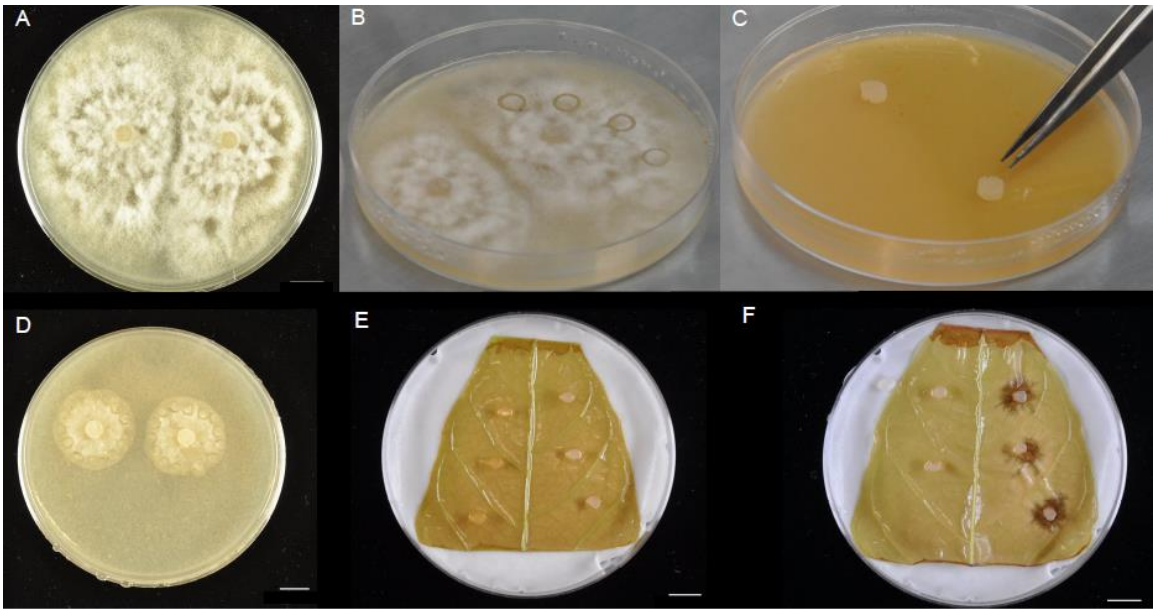


Figure 5.7 – Images representing leaf infection process. **A)** Mature (approximately 1 week since inoculation) plate of the cacao pathogen *Phytophthora palmivora*. **B)** Plate of *P. palmivora* with four agar plugs bored into V8 media. **C)** Inoculation of a new plate of 20% V8 media by transferring agar plugs, pathogen side down, onto the media. **D)** Typical size of pathogen growth 48 hours after inoculation of new plate. Agar plugs are bored around the edges of the cultures to be used for leaf inoculation. **E)** Inoculation of a Stage C cacao leaf with pathogen. Control (media only) plugs are placed on the left side, plugs containing pathogen are placed on the right. **F)** Lesion development 48 hours after inoculation. All scale bars represent 1 cm.

Part II – Inoculation of leaf tissue

Timing: Approximately 1 hour

1. Sterilize the laminar flow hood with UV light for 2 minutes, and wipe the area with 70% ethanol.
2. Sterilize the 3 mm cork borer with ethanol and flame. Let it cool, and then use it to bore holes into a plate of V8 media with no pathogen. These agar plugs will be used to demonstrate that placing the media on the leaves does not result in formation of a lesion.
3. Re-sterilize the cork borer and let cool briefly. Use it to bore agar plugs around the outside edges of the pathogen culture, as shown in Fig 5.7D. Creating plugs from the edges of the

culture ensures that the pathogen is actively growing, and that all agar plugs used will be equally virulent.

4. Place agar plugs on the leaf as shown in Fig. 7E. First, sterilize forceps and a probe, and let them cool. Use it to place three V8 agar plugs without pathogen along the left side of the leaf in a line parallel to the midvein. Place agar plugs with containing pathogen mycelia along the right side of the leaf. Ensure that the pathogen's side of the agar plug is in contact with the leaf. Avoid placing an agar plug near the primary or secondary veins as they affect the shape of lesion growth, or too close to another plug so that lesions do not coalesce. For reference, when harvesting lesions, a 1.5 cm diameter disc will be cut for each lesion.

5. Repeat the inoculation for all remaining leaf sections.

6. Use the atomizer to spray each leaf with sterilized water. Ensure that the leaf was uniformly misted.

7. Seal plates with parafilm, handling carefully so as to not disturb agar plugs.

8. Incubate leaves in a growth chamber at 25^{o}C with 12 hr: 12 hr light dark cycle for two days with a light intensity of 55 μmol m^{-2} s^{-1}.

Part III – Leaf photography and tissue collection

Timing: Approximately 2 hours

1. If inoculation was successful lesions will have developed after 48 hours, (Fig. 7F). Photograph the leaves, including a ruler as reference for measurement. Use ImageJ (Schneider et al., 2012) to trace the area of the lesions, and average the three lesions on a leaf to serve as a biological replicate.

2. Remove agar plugs using forceps. Follow appropriate guidelines for disposal of the pathogen.

3. Cut lesions of each leaf using a 1.5 cm diameter cork borer with location of agar as center. Using a sharpened cork borer will prevent leaf tearing. For each leaf, place the three leaf discs into a 2 mL microfuge tube. Flash freeze the tissue with liquid nitrogen, and store at -80^{o}C. This tissue will be used for DNA extractions and subsequent qPCR to compare relative abundance of pathogen to host DNA within the infected tissue.

4. Use a scalpel to excise the "donut" of tissue around where the lesions developed. Again, place this tissue in a 2 mL microfuge tube, flash freeze, and store at -80^{o}C. This tissue can be used for

RNA extraction as previously described (Zhang et al., 2014) to verify overexpression of the transgene, and to compare expression level of other genes of interest between the transgene-overexpressing samples and those treated with vector control.

Conclusions

The transient transformation procedure described here offers a rapid means of performing functional genetic characterization studies on cacao, a long generation tree crop of significant economic importance. The strategy has already been applied to several studies (Shi et al., 2013; Mejía et al., 2014; Zhang et al., 2014; Fister et al., 2015; Helliwell et al., 2015), which were studies investigating single gene overexpression and knockdown. The cacao transient transformation protocol was first described by Shi *et al* (Shi et al., 2013). In this study, the transcription factor NPR3 was shown to be a negative regulator of the cacao defense response by using transient microRNA-mediated knockdown of the TcNPR3 transcript in cacao leaves followed by *Phytophthora* inoculation assays. The protocol was also applied to the study of cacao defense response by Mejia *et al*., who demonstrated that overexpression of a cacao gene induced by presence of the endophyte *Colletotrichum tropicale,* decreased susceptibility to *Phytophthora* infection (Mejía et al., 2014). This result suggested that the presence of endophytes in cacao leaves confers a mutualistic enhanced defense response to attack by pathogens (Mejía et al., 2014). The transient transformation was also applied by Fister *et al.* in a study demonstrating the positive role of NPR1, the master regulator of systemic acquired resistance, in cacao's response to infection by *Phytophthora* (Fister et al., 2015). Helliwell *et al.* applied cacao leaf transient transformation to show that expression of phosphatidylinositol-3-phosphate binding proteins can decrease susceptibility to infection by competitively inhibiting pathogens' effector proteins' abilities to bind to host cell membranes (Helliwell et al., 2015). Finally, Zhang *et al.* used the transient transformation strategy to characterize the role of the transcription factor TcLEC2, transiently overexpressing it in leaves to demonstrate its role in regulating genes related to embryo development (Zhang et al., 2014). The development and application of this leaf transformation study enables these types of gene characterization studies to be performed rapidly and at lower cost than through the creation of stably transgenic

plants. Without this strategy for rapid gene testing, similar analyses require several years and extensive resources in order to generate stably transgenic cacao trees. The transient transformation strategy is also in the process of being adopted for altering expression of multiple genes by including additional cassettes, and will also be used to develop CRISPR/CAS9-mediated genome editing in cacao leaves.

References

Argout X, Salse J, Aury J, Guiltinan M, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova S, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto J, Sabot F, Kudrna D, Ammiraju J, Schuster S, Carlson J, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A. 2011. The genome of *Theobroma cacao*. Nature Genetics **43**, 101 - 108.

Bent A. 2006. Arabidopsis thaliana Floral Dip Transformation Method. In: Wang K, ed. *Agrobacterium Protocols*, Vol. 343: Humana Press, 87-104.

Fister AS, O'Neil ST, Shi Z, Zhang Y, Tyler BM, Guiltinan MJ, Maximova SN. 2015. Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. Journal of Experimental Botany **66**, 6245-6258.

Gambino G, Gribaudo I. 2012. Genetic transformation of fruit trees: current status and remaining challenges. Transgenic Research **21**, 1163-1181.

Guiltinan M, Verica J, Zhang D, Figueira A. 2008. Genomics of *Theobroma cacao*, "The foods of the gods". In: P.M. Moore, Ming R, eds. *Genomics of Tropical Crop Plants*: New York: Springer.

Guiltinan MJ, Maximova SN. 2015. Chapter 4 Applications of Genomics to the Improvement of Cacao. *Chocolate and Health: Chemistry, Nutrition and Therapy*: The Royal Society of Chemistry, 67-81.

Helliwell EE, Vega-Arreguín J, Shi Z, Bailey B, Xiao S, Maximova SN, Tyler BM, Guiltinan MJ. 2015. Enhanced resistance in Theobroma cacao against oomycete and fungal pathogens by secretion of phosphatidylinositol-3-phosphate-binding proteins. Plant Biotechnology Journal, n/a-n/a.

Janssen B-J, Gardner RC. 1990. Localized transient expression of GUS in leaf discs following cocultivation with Agrobacterium. Plant Molecular Biology **14**, 61-72.

Křenek P, Šamajová O, Luptovčiak I, Doskočilová A, Komis G, Šamaj J. 2015. Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications. Biotechnology Advances.

Lacroix B, Citovsky V. 2013. The roles of bacterial and host plant factors in Agrobacterium-mediated genetic transformation. Int. J. Dev. Biol **57**, 467-481.

Li Z, Traore A, Maximova S, Guiltinan M. 1998. Somatic embryogenesis and plant regeneration from floral explants of cacao (Theobroma cacao L.) using thidiazuron. In Vitro Cell and Developmental Biology - Plant **34**, 293 - 299.

Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP. 2011. Cacao breeding in Bahia, Brazil: strategies and results. Crop Breeding and Applied Biotechnology **11**, 73-81.

Maximova S, Marelli J-P, Young A, Pishak S, Verica J, Guiltinan M. 2006. Over-expression of a cacao class I chitinase gene in *Theobroma cacao L*. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224**, 740-749.

Maximova S, Miller C, Antúnez de Mayolo G, Pishak S, Young A, Guiltinan MJ. 2003a. Stable transformation of Theobroma cacao L. and influence of matrix attachment regions on GFP expression. Plant Cell Reports **21**, 872-883.

Maximova S, Miller C, De Mayolo GA, Pishak S, Young A, Guiltinan MJ. 2003b. Stable transformation of Theobroma cacao L. and influence of matrix attachment regions on GFP expression. Plant Cell Reports **21**, 872-883.

Maximova S, Young A, Pishak S, Miller C, Traore A, Guiltinan M. 2005. Integrated system for propagation of Theobroma cacao L. Protocol for Somatic Embryogenesis in Woody Plants, 209 - 229.

Mejia L, Guiltinan M, Shi Z, Landherr L, Maximova S. 2012. Expression of Designed Antimicrobial Peptides in *Theobroma cacao L.* Trees Reduces Leaf Necrosis Caused by *Phytophthora* spp. Small Wonders: Peptides for Disease Control **1905**, 379-395.

Mejía LC, Herre EA, Sparks JP, Winter K, García MN, Van Bael SA, Stitt J, Shi Z, Zhang Y, Guiltinan MJ, Maximova SN. 2014. Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5**.

Motamayor J, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley S, Zheng P, Utro F, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler B, Stack J, Feltus F, Mustiga G, Amores F, Phillips W, Marelli JP, May G, Shapiro H, Ma J, Bustamante C, Schnell R, Main D, Gilbert D, Parida L, Kuhn D. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology **14**, R53.

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao L*). PLoS ONE **3**, e3311.

Nam J, Mysore K, Zheng C, Knue M, Matthysse A, Gelvin S. 1999. Identification of T-DNA tagged Arabidopsis mutants that are resistant to transformation by Agrobacterium. Molecular and General Genetics MGG **261**, 429-438.

Narasimhulu SB, Deng XB, Sarria R, Gelvin SB. 1996. Early transcription of Agrobacterium T-DNA genes in tobacco and maize. The Plant Cell **8**, 873-886.

Schell JS. 1987. Transgenic plants as tools to study the molecular organization of plant genes. Science **237**, 1176-1183.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Meth **9**, 671-675.

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan MJ. 2013a. The Salicylic Acid Receptor NPR3 Is a Negative Regulator of the Transcriptional Defense Response during Early Flower Development in Arabidopsis. Molecular Plant **6**, 802-816.

Shi Z, Zhang Y, Maximova S, Guiltinan M. 2013b. TcNPR3 from *Theobroma cacao* functions as a repressor of the pathogen defense response. BMC Plant Biology **13**, 204.

Swanson JD, Carlson JE, Guiltinan MJ. 2008. Comparative Flower Development in *Theobroma cacao* Based on Temporal Morphological Indicators. International Journal of Plant Sciences **169**, 1187-1199.

Wood GAR, Lass R. 2008. *Cocoa*: Longman Scientific & Technical, copublished by John Wiley & Sons, New York.

Zhang Y, Clemens A, Maximova S, Guiltinan M. 2014. The Theobroma cacao B3 domain transcription factor TcLEC2 plays a duel role in control of embryo development and maturation. BMC Plant Biology **14**, 106

Chapter 6: Retrospective

Outro

In the Dissertation Overview section of Chapter 1, a question was put forth that this work sought to answer: what genes are most important in cacao's defense response? Each of the subsequent chapters offered a partial answer to the question. In Chapter 2, the PR genes of cacao were defined, and a whole genome microarray was used to identify which genes were induced by treatment with two pathogens. This revealed a set of PR genes which are collectively induced in leaf tissue to combat infection. In Chapter 3, we compare the responses of two genotypes to salicylic acid treatment. This experiment begins to reveal the complexity of defense response induction, as the susceptible genotype showed a more canonical response to SA, including PR gene induction, while the more disease tolerant genotype exhibited a marked induction of chloroplastic and mitochondrial machinery linked to ROS generation. In Chapter 4, a genotyping survey is presented where three defense genes were sequenced from trees representing three South American cacao population. Here we garnered support for the ability to predict more variable cacao genes from variable orthologs in other tropical trees. It is likely that the various isoforms of the different proteins have differential capacity for interaction with and recognition of pathogen strains in the trees' area. In Chapter 5, a protocol for transient transformation of cacao leaf tissue is presented, and its usefulness is demonstrated through overexpression of a chitinase gene that confers increased resistance to *Phytophthora* infection. Finally, in Appendix A, we describe a strategy whereby layers of information are used to prioritize candidate genes for functional study. While the transformation protocol was successful at overexpressing the genes, this overexpression did not confer improved resistance to infection. While there are a few interpretations for this result, it serves as a reminder that coordination of hundreds to thousands of genes is critical for defense, and causing overexpression of a single gene may not significantly modulate a phenotype as complex as infection severity. The results of our analyses are reviewed below, with a focus on strategies for

improvement of our analyses and experiments to be performed which would continue to elucidate the key players in cacao immunity.

Defining cacao's induced defenses

We were able to apply protein sequence information from model organisms to classify genes into the 15 PR families present in cacao. We identified 359 PR genes in these families, with sizes ranging from 5 (PR-17) to 81 (PR-9). Comparing family size to other species, we found that cacao is fairly typical in terms of PR gene family size. Rough family sizes appear to be conserved across species, indicating that evolutionary processes have tolerated expansions of families across monocots and dicots.

Further, we found that ~50% of cacao PR genes are organized into tandem arrays, as has been described for a variety of other defense gene families (Ferrari et al., 2003; Rizzon et al., 2006; Li et al., 2009; Lebel et al., 2010), and we found this organization to be present in three additional dicots and two monocots. It has been argued that because a duplicated defense gene can dramatically impact fitness without altering the plant's regulatory stoichiometry, these types of duplications persist in the genome (Sterck et al., 2007). This creates a positive feedback loop, where increasingly repetitive stretches of the genome become more prone to polymerase slippage and non-homologous recombination, which leads to generation of more gene copies (Freeling, 2009).

We noted that expression of tandem array members was highly divergent, both basally and with regard to induction by pathogen. For example, in a six-member array of adjacent PR-1s, only one gene was induced by *Phytophthora palmivora*. This trend is consistent with evolutionary theory, where genome-wide analyses showed that genes duplicated in local events tend to diverge in expression pattern more rapidly than those duplicated in polyploidization or large segmental duplication events (Blanc and Wolfe, 2004). In such cases, it is unsurprising that genes within a family, despite close physical linkage, are not induced by a pathogen. However, the PR-3 and PR-4 families both had tandem arrays containing multiple genes induced by both

tested pathogens. This may indicate that these result from relatively recent duplications which are still diverging.

Compilation of transcriptomic resources is essential for characterizing cacao's defense response. By performing this analysis, we have hopefully identified the PR genes which play critical roles in defense in foliar tissue. To make the data more impactful to breeding, a next step would be to characterize natural variation in expression of these genes at a basal level and after pathogen treatment. Higher basal or induced expression may confer greater tolerance, and therefore breeding to select for high PR expression would be a strategy for improving the crop. As more cacao genome sequences become available, *in silico* analysis of defense gene promoters may become a powerful tool for predicting genotypes with more responsive regulatory machinery. Other groups have used *in silico* promoter analyses to validate the role of putative transcription factor binding sites and to predict novel domains integral to stress responses (Maruyama et al., 2012). Leveraging dozens to hundreds of cacao genome sequences and expression data linked to pathogen response would likely enable identification of genotypes with ideal motif concentration or motif positioning within promoter sequences that optimizes defense gene inducibility.

Scavina 6 and ICS1: Unique strategies for (partial) tolerance

The foray investigating genotype specificity of defense was successful, but the results may have revealed the complicated nature of breeding to improve the defense response. Further, targeting the Sca6 and ICS1 genotypes was useful because they have been employed in breeding programs and QTL mapping projects (Lanaud et al., 2009). Ultimately we showed that while salicylic acid treatment reduced infection symptoms in both genotypes, the underlying mechanism of SA-induced gene regulation differed.

The study first validated the expected differences in defense response expected for Sca6 and ICS1; Sca6 was more resistant to infection by *Phytophthora tropicalis* than ICS1 at a basal level, and salicylic acid treatment enhanced defense in both genotypes, with SA-treated Sca6 showing the strongest immune response. The master regulatory of systemic acquired resistance,

the defense branch governed by salicylic acid, is NPR1 (Vlot et al., 2009). To garner further support for the role of the systemic acquired resistance pathway to defense in cacao, we next sought to functionally characterize cacao NPR1. Previous work from our lab had already shown that expression of cacao NPR1 could partially complement Arabidopsis *npr1* mutants (Shi et al., 2010), so we next demonstrated that NPR1 overexpression in Sca6 enhanced the defense response using our transient transformation protocol. Transient overexpression increased NPR1 transcript abundance by roughly two-fold, and it decreased lesion size and pathogen biomass by roughly half. At face value these results corroborate the effect of salicylic acid in defense by showing that modulation further along the pathway also promotes defense. However, there is evidence of an NPR1 independent mechanism through which SA can act (Prithiviraj et al., 2005; Herrera-Vásquez et al., 2015). While NPR1 and its family members were recently shown to be part of the SA receptor complex (Fu et al., 2012; Wu et al., 2012), npr1 mutant Arabidopsis plants are still capable of SA-dependent gene induction, particularly of ROS generating pathways (Yan and Dong, 2014). Therefore, the effect of salicylic acid treatment that we measured in cacao may not be exclusively filtered through the NPR1 protein.

Analysis of the SA response using transcriptome data revealed mechanistic similarities and differences in the responses of Sca6 and ICS1. Gene Ontology analysis showed that while both genotypes showed a general upregulation of genes involved in oxidative stress response, the cellular localization of this response differed. About 30 chloroplastic and mitochondrial genes in Sca6 were upregulated, while they were downregulated in ICS1. ROS staining confirmed greater accumulation of superoxide in Sca6, potentially linking transcriptional differences between genotypes to differential ROS production. Further, we examined response of PR family members in both genotypes, as induction of these genes is a hallmark of systemic acquired resistance induction. Counterintuitively, we found that ICS1 had more PR genes induced. Staining of reactive oxygen species verified that Sca6 produced more superoxide after salicylic acid treatment, providing a potential link between transcriptomic modulation and increased defense in that genotype.

Notably, this analysis was performed at in greenhouses and growth chambers, and field evaluation of the genotypes is critical. It is widely reported that Sca6 shows superior resistance and that ICS1 is widely susceptible, but molecular analyses of these genotypes in the field have

yet to be performed. This experiment linked exogenous salicylic acid treatment to enhanced resistance; however, we also have yet to measure differences in naturally occurring salicylic acid concentrations between cacao genotypes. Genetic variability affecting its synthesis or that of hormones with which salicylic acid has crosstalk could dramatically influence signal transduction and defense response.

Assessing germplasm for variation in defense genes

Our investigation of genetic diversity of defense genes had two primary goals. First, the Pound, French Guianan, and Ecuadorian Nacional populations are of great interest to cacao breeders because of their desirable quality and disease resistance traits, so assessing variation is integral to successfully utilizing the germplasm. Secondly, we sought to probe whether genes shown to be more polymorphic in other tropical trees had cacao orthologs that also showed signatures of diversifying selection. While our sample size of three genes is too low to make any definitive claims about differences between the populations or the broad applicability of the approach, it did provide interesting data regarding both of the goals.

By sequencing only three genes, differences between the populations already became apparent. Foremost, the French Guianan population had much less variability than the other two, only showing SNPs at 7 of 15 variable sites. The Pound and French Nacional populations were more similar in that they had more variable loci, with 15 and 16 SNPs respectively. The similarity of the variation detected in these populations was striking: if they were recently diverging, one would expect to find more population-specific SNPs. The fact that so many alleles are shared likely indicates that the mutations generating the variation predate the dispersal of the germplasm to geographically isolated regions. Nonetheless, for the CRSP38 and PGIP loci, multiple haplotypes are retained at intermediate frequencies. In combination with significant or marginal positive values of Tajima's D, this suggests these genes are under diversifying selection. One explanation would be that pathogen strains in the environment of these populations encode interacting proteins that also have varied structure, so preservation of multiple forms of the defense genes in the cacao populations is favored. Importantly, however, no gene-for-gene

interaction has been demonstrated between cacao and any of its pathogens, albeit cacao's interactions with its pathogens are far less fully described than those between model species and their pathogens. *In vitro* demonstration of interaction of a cacao and pathogen protein, an R gene/effector pair or a PGIP/PG, perhaps with different cacao protein isoforms showing differential binding capacity for pathogenic isoforms, would be useful for confirming our hypotheses about functional genetic diversity.

Transient transformation: a strategy for rapid gene assessment

With the help of collaborators and other Guiltinan lab members, an assay was developed whereby Agrobacterium containing overexpression and knockdown constructs can efficiently be infiltrated into leaf tissue, allowing detection of phenotypes associated with modulated gene expression. We found that transient transformation in cacao followed a similar timeline to that seen in other species: transgene expression appears to peak approximately 48 hours after infiltration of leaves with Agrobacterium, and expression deteriorates over the course of 3-5 additional days. While our protocol is optimized for use with Stage C leaves, and our use has been limited to the Sca6 genotype, we found that it appears to be effective for transformation of other genotypes. Further optimization may be required to adapt it to poor performing varieties.

Availability of this method is a major asset to screening gene function in a rapid manner. While generating stably transgenic cacao trees can take months to years, this experiment can be carried out in a week, plus the time required to design the T-DNA vector. Further optimization of the protocol which would increase the proportion of successfully transformed leaves from a given batch would be useful, as it would allow for larger sample sizes to be collected from a given experiment, eliminating potential variation introduced between repetitions.

In this dissertation, the transient transformation protocol was successfully applied to show the positive effect on resistance from NPR1 and Chi1 overexpression. It has also been applied to studying blocking of pathogen entry expressing PI3P binding proteins (Helliwell et al., 2015), knockdown of a negative regulator of defense (Shi et al., 2013), and the role of a master

transcriptional regulator in embryogenesis (Zhang et al., 2014). While the genes selected for assessment in Appendix A did not provide the expected results, the assay remains a useful tool for probing gene function, and prioritization of genes to study through creation of stable transgenic trees.

Gene prioritization strategy

Previous work in our lab identified two candidate genes with major roles in resistance, and their function was validated by generating a stably transgenic tree overexpressing the gene (chitinase) (Maximova et al., 2006), or by performing functional complementation in an Arabidopsis mutant (NPR1) (Shi et al., 2010). When these were tested using out transient transformation system, their overexpression again improved resistance to pathogen inoculation. Overexpression of the genes we tested using the prioritization strategy described in Appendix A, however, did not confer enhanced resistance. While this could be an indictment of our approach, it may also simply be a reflection of the fact that hundreds of genes play important roles in defense, and overexpressing single genes will not always confer a phenotype.

The data collected in Chapters 2-4 shed light on reasons why this may be the case. Transcriptomic assessment of infection symptoms clearly shows differences in response to cacao's various pathogens. While overexpressing a single gene may be effective against a pathogen strain, or perhaps even a species or family, it is unlikely to be a universal silver bullet. It is also critical to consider the role of genetic variation in resistance. A given locus, for example PGIP, encodes seven variant polypeptide sequences. It is possible, albeit unlikely, that each of these evolved to recognize and inhibit a specific pathogenic polygalacturonase. If we cloned and overexpressed each of these, we may see no phenotype unless we work with the specific pathogen strain encoding the correct interacting partner. Even if the pair is correctly tested in tandem, it remains a possibility that no phenotype would be detected because the effect of inhibition of this polygalacturonase may not be especially detrimental to a particular pathogen's entry into the cell.

A major hurdle for this approach is the lack of an observed gene-for-gene defense response in cacao. A new approach may be derived from the availability of genome sequences for cacao's pathogens, such as witches' broom and the variety of pathogenic *Phytophthora spp*. As pathogenic effectors are defined, it may be possible to model which cacao R genes are most likely to interact with these proteins by comparing to known interacting pairs identified in model species. A major hurdle for our approach is the lack of an observed gene-for-gene defense response in cacao. Application of this bioinformatics prediction of protein interactions may be useful in identifying germplasm that could provide an example of R gene-mediated resistance in cacao.

In conclusion…

Technological advancements of the genomics era have made previously unimaginable quantities of data available at low cost and with relative ease, and the resources developed offer a variety of new approaches in biology. As shown in this dissertation, genomics can dramatically benefit the study of tree crops, which can, in many contexts, be difficult to study. Here, I have identified and partially characterized 15 families of cacao's defense genes, compared responsiveness of two agriculturally important genotypes to treatment with salicylic acid, a key defense regulator, examined polymorphism within and between three populations of cacao trees using a preliminary set of three defense genes of interest, demonstrated a protocol for gene functional characterization using transient transformation of cacao leaves, and have begun to screen candidate defense genes to attempt to define their role in resistance. But this is only the beginning of a much longer process of genotype evaluation and gene functional characterization. Genomic and transcriptomic analysis of dozens to hundreds of additional cacao genotypes is underway, enabling more comprehensive SNP identification, more powerful QTL prediction, and more precise prediction of key candidate genes. I am excited to remain involved in the world of cacao functional genomics, to polish projects I have already begun, to begin new experiments in new areas, and to continue asking and answering questions about life.

References

Blanc G, Wolfe KH (2004) Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. The Plant Cell **16:** 1679-1691

Ferrari S, Vairo D, Ausubel FM, Cervone F, De Lorenzo G (2003) Tandemly Duplicated Arabidopsis Genes That Encode Polygalacturonase-Inhibiting Proteins Are Regulated Coordinately by Different Signal Transduction Pathways in Response to Fungal Infection. The Plant Cell **15:** 93-106

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology **60:** 433-453

Fu ZQ, Yan S, Saleh A, Wang W, Ruble J, Oka N, Mohan R, Spoel SH, Tada Y, Zheng N, Dong X (2012) NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. Nature **486:** 228-232

Helliwell EE, Vega-Arreguín J, Shi Z, Bailey B, Xiao S, Maximova SN, Tyler BM, Guiltinan MJ (2015) Enhanced resistance in Theobroma cacao against oomycete and fungal pathogens by secretion of phosphatidylinositol-3-phosphate-binding proteins. Plant Biotechnology Journal**:** n/a-n/a

Herrera-Vásquez A, Carvallo L, Blanco F, Tobar M, Villarroel-Candia E, Vicente-Carbajosa J, Salinas P, Holuigue L (2015) Transcriptional control of glutaredoxin GRXC9 expression by a salicylic acid-dependent and NPR1-independent pathway in Arabidopsis. Plant Molecular Biology Reporter **33:** 624-637

Lanaud C, Fouet O, Clément D, Boccara M, Risterucci AM, Surujdeo-Maharaj S, Legavre T, Argout X (2009) A meta–QTL analysis of disease resistance traits of Theobroma cacao L. Molecular Breeding **24:** 361-374

Lebel S, Schellenbaum P, Walter B, Maillot P (2010) Characterisation of the *Vitis vinifera* PR10 multigene family. BMC Plant Biology **10:** 184

Li Z, Zhang H, Ge S, Gu X, Gao G, Luo J (2009) Expression pattern divergence of duplicated genes in rice. BMC Bioinformatics **10:** S8

Maruyama K, Todaka D, Mizoi J, Yoshida T, Kidokoro S, Matsukura S, Takasaki H, Sakurai T, Yamamoto YY, Yoshiwara K, Kojima M, Sakakibara H, Shinozaki K, Yamaguchi-Shinozaki K (2012) Identification of Cis-Acting Promoter Elements in Cold- and Dehydration-Induced Transcriptional Pathways in Arabidopsis, Rice, and Soybean. DNA Research **19:** 37-49

Maximova SN, Marelli JP, Young A, Pishak S, Verica JA, Guiltinan MJ (2006) Over-expression of a cacao class I chitinase gene in *Theobroma cacao L.* enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224:** 740-749

Prithiviraj B, Bais HP, Jha AK, Vivanco JM (2005) Staphylococcus aureus pathogenicity on Arabidopsis thaliana is mediated either by a direct effect of salicylic acid on the pathogen or by SA-dependent, NPR1-independent host responses. The Plant Journal **42:** 417-432

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol **2:** e115

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan M (2010) Functional analysis of the *Theobroma cacao* NPR1 gene in arabidopsis. BMC Plant Biology **10:** 248

Shi Z, Zhang Y, Maximova S, Guiltinan M (2013) TcNPR3 from *Theobroma cacao* functions as a repressor of the pathogen defense response. BMC Plant Biology **13:** 204

Sterck L, Rombauts S, Vandepoele K, Rouzé P, Van de Peer Y (2007) How many genes are there in plants (… and why are they there)? Current Opinion in Plant Biology **10:** 199-203

Vlot AC, Dempsey DMA, Klessig DF (2009) Salicylic acid, a multifaceted hormone to combat disease. Annual Review of Phytopathology **47:** 177-206

Wu Y, Zhang D, Chu Jee Y, Boyle P, Wang Y, Brindle Ian D, De Luca V, Després C (2012) The Arabidopsis NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. Cell Reports **1:** 639-647

Yan S, Dong X (2014) Perception of the plant immune signal salicylic acid. Current Opinion in Plant Biology **20:** 64-68

Zhang Y, Clemens A, Maximova S, Guiltinan M (2014) The Theobroma cacao B3 domain transcription factor TcLEC2 plays a duel role in control of embryo development and maturation. BMC Plant Biology **14:** 106

Appendix A:
A strategy for functional characterization of defense genes

Introduction

Gene prioritization strategy

Availability of the cacao genome sequence (Argout et al., 2011; Motamayor et al., 2013) has enabled new methods for gene functional analysis. Using these data, specific primers can easily be created for gene cloning, design of knockdown or knockout constructs, or qRT-PCR to further characterize gene expression and function. However, the scale and complexity of defense systems makes identifying the most important genes a challenge.

If we consider the stages of the defense response, the top tier would be genes involved recognition of pathogens, RLKs and NLRs. Both of these classes are large superfamilies across plant species (Sanseverino et al., 2010), and initial annotation of genes in cacao predicts that they are composed of 200-300 genes. The second tier of defense, signal transduction, and it involves dozens of MAPKs and associated proteins (Meng and Zhang, 2013) and hundreds of transcription factors (van Verk et al., 2009; van Verk et al., 2011). The final tier, the induced genes, includes hundreds of PR genes (van Loon et al., 2006) and ROS generating machinery (O'Brien et al., 2012), and crosstalk between defense and developmental processes links defense to growth, maturation, and general health (Naseem et al., 2015). While genomic tools make the process much easier, functional screening of genes is still a time and resource intensive process, which requires focused analysis of only top candidates predicted to be the most important for defense.

Research performed on model organisms and other crop plants provides an invaluable first filter for gene prioritization. Thanks to a multitude of studies, many key players in plant-pathogen interactions have already been studied, and cacao homologs of these genes can be selected. Still, thousands of genes with putative roles in defense have been described, and the divergence of species and the idiosyncrasies of cacao's interactions with its pathogens must be taken into account when for choosing candidates. We have applied several filters based on different types of support for the importance of a candidate gene's role in defense in cacao (Fig. A-1). These filters are outlined below.

Defining the optimal strategy for selection of high priority candidate genes within the defense response is an ongoing challenge. The filters and data described here are useful for identifying candidates, but as new transcriptomic and gene functional analyses are performed, they will be incorporated to refine the strategy. In this pilot study, six genes were selected using

our existing criteria, and their role in defense was screened using transient overexpression and subsequent infection of leaf tissue.

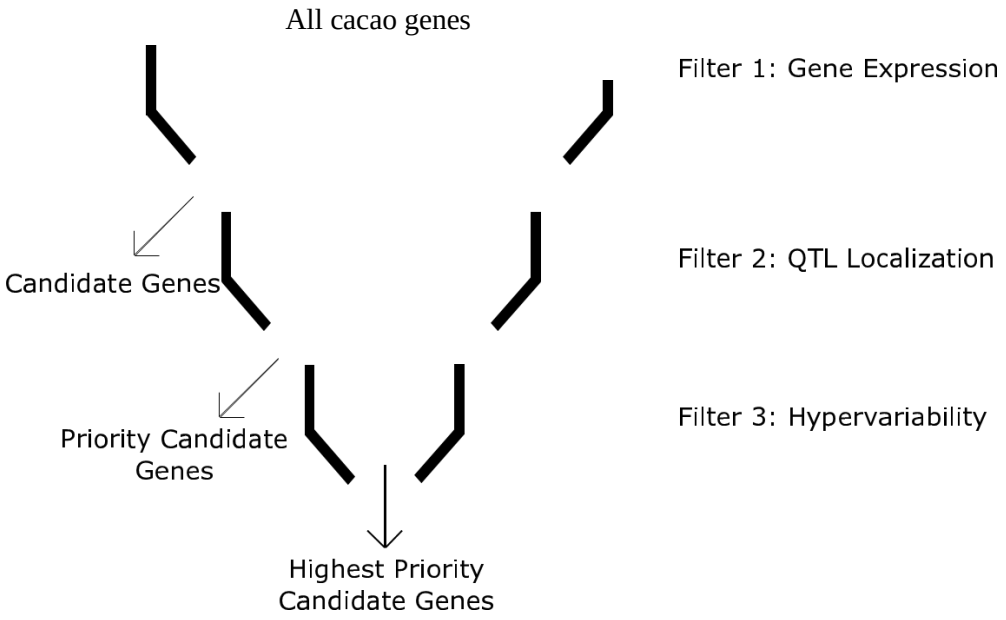


Figure A.1 – Schematic representation of gene prioritization strategy. Genes that pass more filters have more evidence supporting the importance of their role in defense, and are considered higher priority candidates for functional analysis.

Expression dynamics

To explore the expression profiles of the PR family members we utilized the dataset described in Argout *et al*., 2008, which contains the sequences of 56 cDNA libraries created using RNA from a variety of cacao genotypes, tissues, and conditions (e.g. infection, drought, fermentation of seeds) (Argout *et al*., 2008). These data can be viewed using the GenomeThreader track in GBrowse on the Criollo genome browser (cocoagendb.cirad.fr/). While the breadth of tissues and conditions used makes this a useful reference for identifying activity in specific conditions, the small size of the libraries means that some potentially important genes were not identified.

The Guiltinan-Maximova Lab has used microarrays to analyze gene expression in a variety of tissues and after a number of treatments. These include development of somatic and zygotic embryos (Maximova et al., 2014), treatment of rooted-cuttings of two cacao genotypes with salicylic acid (Fister et al., 2015), treatment of seedlings with fungal endophytes (Mejía et al., 2014), and treatment of seedlings with a fungal and oomycete pathogen (GEO: GSE73804). Another group published their analysis of RNA-seq data from an experiment involving treatment

of seedlings with *Moniliophthora perniciosa*, the witches' broom pathogen (Teixeira et al., 2014). Collectively, these data provide another useful set of references for gene prioritization. Genes upregulated by pathogen treatment are considered strong contenders for genes with important roles in the defense response. These often include pathogenesis-related genes, for example chitinases, which can be overexpressed to reduce the growth of a pathogen after infection (Maximova et al., 2006). While the majority of our data are from infected leaf and shoot tissue, non-infection based data, like that from the somatic embryogenesis-related experiments, also provide useful references for tissue-specific regulation of genes in defense families.

Chapters 3 and 4 present microarray data on salicylic acid treatment and two pathogen treatments, respectively, both of which could be examples of the sorts of databases leveraged within this filter.

QTL maps

To identify regions of chromosomes conferring resistance to a variety of pathogens, QTL mapping populations of cacao trees have been grown, screened and analyzed. By crossing genotypes that are resistant and susceptible to one or more of cacao's major diseases, more than 20 of these populations have been created (Lanaud et al., 2009; Gutiérrez et al., 2016). These analyses predict 65 QTLs for black pod rot resistance, five QTLs for frosty pod resistance, six for witches broom resistance, and ten for vascular streak dieback (Gutiérrez et al., 2016). The number of individuals evaluated in these experiments was often small, resulting in QTL that are quite large; as a result, the majority of the cacao genome sits within disease resistance QTL, albeit there are a number of large QTL of minor effect. To resolve this problem, a meta-analysis of QTL studies was performed to identify regions where multiple QTL, particularly for black pod resistance, overlap (Lanaud et al., 2009). Even the strong QTL or these meta-QTL can span more than a megabase, and some of these still contain more than a thousand genes. Clearly finer mapping is required, but targeting versions of genes that come from resistant varieties and are located within QTL identified in populations where the resistant variety served as the parent provides a potentially valuable second filter.

Gene hypervariability

Recent work by collaborators identified sets of genes with elevated evolutionary rates. Marden *et al.* (unpublished data) identified R gene orthologs in a set of 6 tropical tree species which exhibited higher than average polymorphism and elevated pN/pS ratios. A goal here was first to assess whether genes with high pN/pS orthologs in these tropical tree species were also more polymorphic in cacao (addressed in Chapter 5). Marden *et al.* conclude that balancing selection favors existence of diverse haplotypes of genes involved in pathogen interactions at a

population level such that adjacent plants would have reduced co-susceptibility. Therefore, we predict that genes showing signatures of diversifying selection are likely to encode proteins which interact with pathogenic molecules and which may be involved in the defense response. Overexpressing these likely pathogen interactors may increase cacao's defense response by enhancing pathogen detection (e.g. R genes interacting with effectors) or inhibit pathogen encoded proteins that break down plant cell walls (e.g. polygalacturonase inhibitors interacting with polygalacturonases).

Methods

Gene cloning

Using the gene sequence annotated in the Criollo genome browser, primers were designed to amplify target genes from gDNA. The primers were appended with restriction sites to make them compatible with a binary vector used for Agrobacterium-mediated transformation (base vector is pGH00.0126, GenBank ID: KF018690). Genes were amplified using Phusion polymerase (New England Biolabs, Ipswich, MA), and A-tailed using standard Taq polymerase. PCR products were run on a 1% agarose gel to assess that amplicons were the correct size. Bands were cut from the gels and purified using the protocol described in the GeneClean II kit (MP Biomedicals, Santa Ana, CA). Purified PCR products were ligated into pGEM-T vector (Promega, Madison, WI), transformed into competent *E. coli*, and colonies were screened using blue/white selection. Five to ten colonies were sequenced to detect any mutations from PCR errors. A colony with the correct sequence was used to inoculate a liquid culture, which was mini-prepped using Wizard Minicolumns (Promega, Madison, WI). The collected cloning vector was then digested using the enzymes with recognition sites appended to the primers, and the binary vector was digested using complementary enzymes. The dropout product from the cloning vector (the gene of interest) and the backbone of the binary vector were purified from 1% agarose gel and were ligated overnight at 4°C using T4 DNA ligase (New England Biolabs, Ipswich, MA). The ligation product was transformed into *E. coli*, and colony PCR was performed to screen for positive transformants.

After identifying a successfully transformed *E. coli* colony, it was used to inoculate a liquid culture, and was then mini-prepped. The collected DNA was used to transform electro-competent Agrobacterium, strain AGL1. Again colony PCR was used to identify positive transformants. For each gene, two to three positive colonies were digested to verify size and structure of the vector, and the region containing the CaMV 35S promoter driving the gene of interest was sequenced to verify the integrity of the sequence.

Transient transformation and infection

The transient assay and leaf infection protocol described in Chapter 5 were used to test the function of gene overexpression in detached cacao leaves. Briefly, stage C leaves of Sca6 plants were vacuum infiltrated with *Agrobacterium* containing a vector with a cassette overexpressing the gene of interest or a control vector. Forty-eight hours after infiltration, leaves were screened for GFP fluorescence, and those with GFP coverage over 80% of the leaf area were used for subsequent experiments. Unsuccessfully transformed leaves were discarded. Tissue was isolated from successfully transformed and used for RNA extractions. The majority of successfully transformed tissue was used for infection assays. Agar plugs containing mycelia of *Phytophthora palmivora* (or no pathogen as control) were placed on leaves 48 hours after *Agrobacterium* infiltration. Leaves were placed in a growth chamber for 72 hours and then photographed. Lesion sizes were compared and statistically analyzed using linear models in JMP.

Evaluation of leaf detachment effect

Stage C leaves were collected from greenhouse grown Sca6 and ICS1 plants. Leaves were immediately cut into three sections. One section, corresponding to 'fresh' samples, was frozen using liquid nitrogen and was used to measure basal expression level. Another sample had its edges sealed and plated, and was placed in a growth chamber for 48 hours under conditions described in Chapter 5. The last was sealed and plated, then infected using agar plugs containing mycelia of *P. palmivora*. Seventy-two hours after inoculation leaves were frozen with liquid nitrogen for RNA extraction.

qRT-PCR

RNA Extractions were performed using PureLink RNA extraction reagent (Thermo-Fisher Scientific, Waltham, MA) following the manufacturer's protocol, cDNA was synthesized using M-MuLV reverse transcriptase (New England Biolabs, Ipswich, MA), and qRT-PCR was performed using TaKaRa Premix Ex Taq SYBR Green reagents (Clontech, Mountain View, CA). Reactions were performed in technical duplicates and followed the following thermocycling protocol in an ABI StepOne Plus Real Time PCR System (Roche, Nutley, NJ): 15 min 94°C, 40 cycles of 15 s at 94°C, 20s at 60°C, and 40 s at 72°C. Relative expression values were analyzed using JMP, and the $2^{-(\Delta\Delta Ct)}$ method was used to calculate the degree of overexpression compared to control vector (Livak and Schmittgen, 2001).

Results

Gene Selection

By considering gene expression dynamics, localization within QTL, and whether genes had highly variable tropical tree orthologs, six genes were selected for cloning and overexpression using our transient transformation protocol. The selected genes are described in Table 1. Two of these genes sat in a major QTL on chromosome 9, which was thought to be important in conferring black pod resistance to a breeding population at CNRA in Ivory Coast. Notably, they also sat near TcNPR1, the master transcriptional co-regulator of systemic acquired resistance, which was previously functionally characterized in our lab (Shi et al., 2010).

Putative functions were also considered in selection of these genes. Although its mechanism of action is unclear overexpression of PR-1 in tobacco increased resistance to oomycete pathogens (Sarowar et al., 2005; Freeling, 2009), and knockdown of a PR-1 family member decreased tolerance of the *Blumeria graminis* fungus (Chae et al., 2014). The TcCRSP38 is the cacao ortholog of a *Ginkgo biloba* secreted protein shown to have antimicrobial properties in vitro (Sawano et al., 2007) through interaction with cell wall carbohydrates (Miyakawa et al., 2014). Both Myb (Britto et al., 2013; Royaert et al., 2016) and WRKY (Ülker and Somssich, 2004; Pandey and Somssich, 2009; van Verk et al., 2011) family members are known to regulate the defense response downstream of the SA and JA/ET pathways. By targeting such transcription factors, we hoped to modulate expression of many downstream anti-microbial proteins. Polygalacturonase inhibitors have known roles in preventing pathogens from breaking down plant cell walls (Yao et al., 1999; De Lorenzo and Ferrari, 2002; Misas-Villamil and van der Hoorn, 2008). While the above proteins all have predicted positive effects on the plant's ability to defend itself, we also targeted GID1L3, a predicted gibberellin receptor, which is a negative regulator of defense (De Bruyne et al., 2014; Ploetz, 2016). By demonstrating a phenotype from its overexpression, we hoped to garner evidence for its role in defense, which would motivate knockdown our knockout experiments that would promote defense.

The Scavina 6 cacao genotype was used as the resistant parent in the populations used to identify the black pod resistance QTL containing CRSP38 and GID1L3 (Risterucci et al., 2003) and the witches' broom resistance QTL containing WRKY50 and Myb251 (Brown et al., 2005; Faleiro et al., 2006). Scavina 6 is also considered to have broad spectrum resistance to a variety of diseases (Pound, 1943; Gutiérrez et al., 2016). Accordingly, these genes were all cloned from Scavina 6 genomic DNA.

Table A.1 - Genes selected for functional characterization and traits used for their prioritization.

Common Name	Gene ID	Description	Microarray Expression Data	In resistance QTL?	Hypervariable orthologs?
PR-1	Tc02_g002410	Defense gene of unknown function	Up 125x by *P. palmivora* Up 56x by *C. theobromicola*	Frosty pod	No
CRSP38	Tc06_g009580	Homolog of secreted *Ginkgo* defense protein	Up 96x by *P. palmivora* Up 46x by *C. theobromicola*	Black pod	Yes
Myb251	Tc09_g005560	Putative defense transcription factor	Up 17x by *P. palmivora*	Black pod, witches' broom	No
WRKY50	Tc09_g005290	Putative defense transcription factor	Up 8x by *P. palmivora* Up 14x by *C. theobromicola*	Black pod witches' broom	No
PGI	Tc05_g018290	Polygalacturonase inhibitor	No differential regulation detected	No	Yes
GID1L3	Tc10_g013340	Putative gibberellin receptor	Up 5x by *P. palmivora* Up 6x by *C. theobromicola*	Black pod	Yes

Overexpression of target genes

Relative expression of target genes was calculated relative to the geometric mean of two housekeeping genes, ACP1 and Tubulin (Fig. A2). Expression of the target genes in vector control-transformed tissue ranged from about 0.5% (CRSP38) to 50% (GID1L3) of the geometric mean of the reference genes. The PR1 and CRSP38 genes in particular exhibited large variation in expression in control tissues. Nevertheless, transient overexpression led to statistically significantly increased expression of all genes (ANOVA $p < 0.05$). Moreover, expression driven by the CaMV 35s promoter consistently produced expression levels 2-7 times that of the mean of the housekeeping genes. The six target genes had varied basal expression levels, and therefore overexpression results in average fold changes ranging from roughly 10- to over 700-fold.

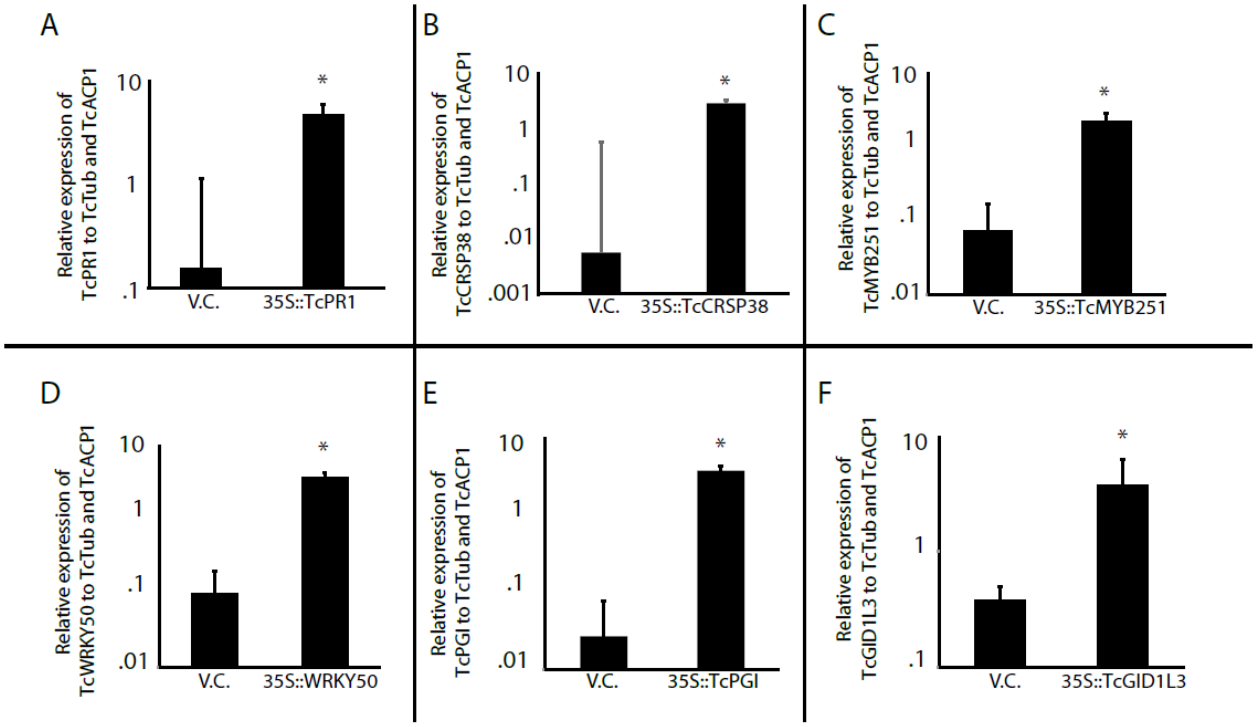


Figure A.2 - Relative expression of target genes to two housekeeping genes 48 hours after Agrobacterium infiltration. Panels represent the target genes: A) PR-1, B) CRSP38, C) Myb251, D) WRKY50, E) PGI, F) GID1L3. Error bars represent standard deviation from A) 12 replicates, B) 19 replicates, C) 7 replicates, D) 6 replicates, E) 6 replicates, F) 12 replicates.

Lesion size assessment

The effect of gene overexpression on defense was assessed by inoculation of leaves with *Phytophthora palmivora* mycelia grown on agar plugs (Fig. A.2). Successfully transformed leaves had media containing oomycete mycelia placed on the right side of the midvein and sterile agar plugs placed on the left side of the midvein. Lesion size was assessed 72 hours after inoculation. Despite successful overexpression of all genes, no statistically significant changes to lesion size were detected. In certain cases, for example with CRSP38, the first repetition of an experiment showed marginally decreased lesion size in overexpressing tissue. However, after subsequent repetitions, p values for lesion reduction increased. We repeated NPR1 overexpression experiments as a positive control, and this again resulted in statistically significant reduction in lesion size (t-test $p < 0.05$).

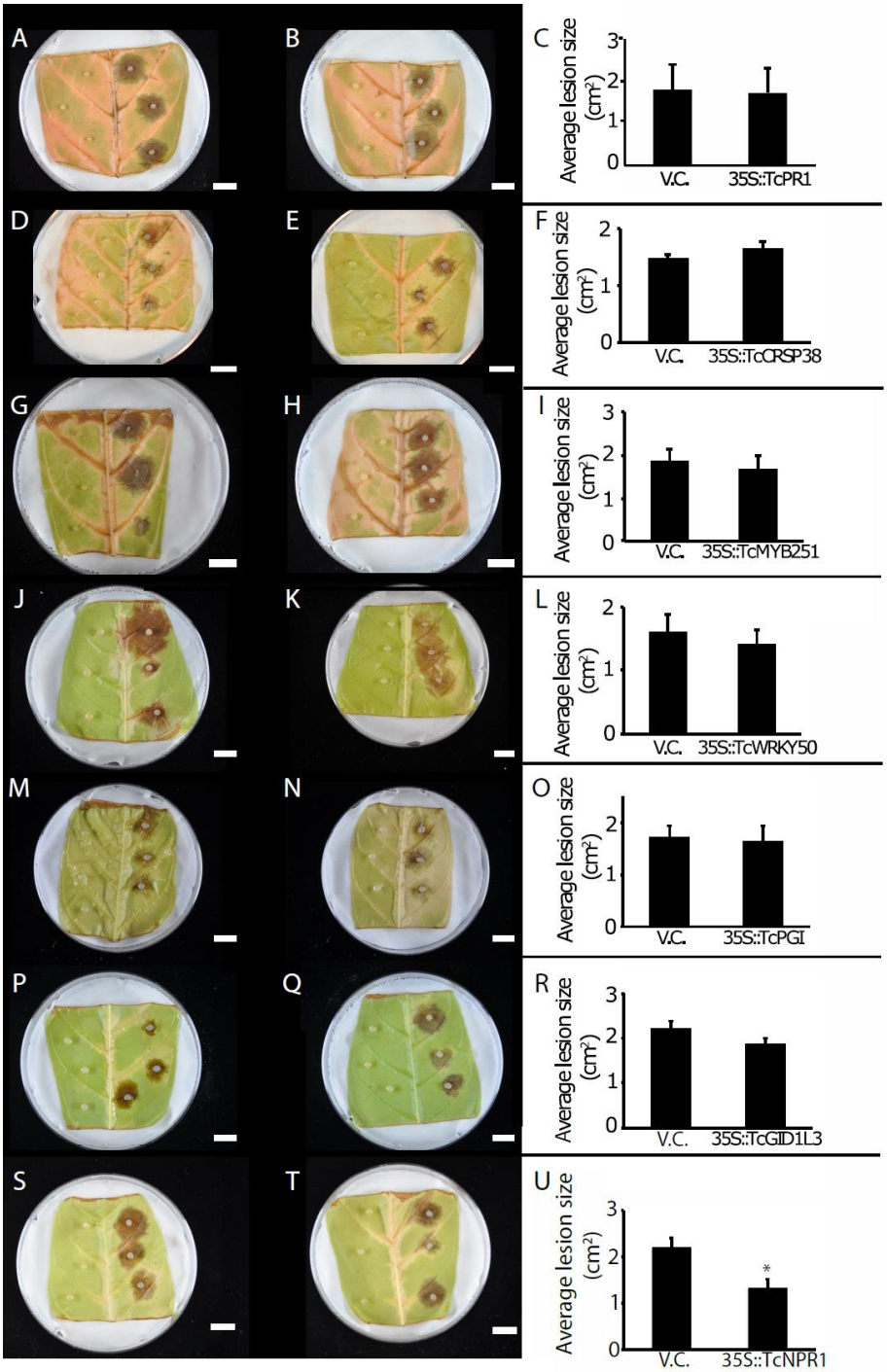


Figure A.3 - Lesion size assessment of target gene overexpression 72 hours after inoculation with *Phytophthora palmivora.* Rows represent experiments using different genes: Row 1) PR-1, Row2) CRSP38, Row 3) Myb251, Row 4) WRKY50, Row 5) PGI, Row 6) GID1L3, Row 7) NPR1. Left column panels are representative photos from vector control transformed tissue, center panels are representative photos from target gene overexpression. Right panels are bar graphs representing average lesion size, with error bars representing standard deviation from: PR1: 13 replicates, CRSP: 21 replicates, MYB251: 7 replicates, WRKY50: 6 replicates, PGI: 7 replicates, GID1L3: 14, NPR1, 8 replicates. Asterisk indicates t-test $p < 0.05$.

Evaluation of detachment effect

The variability of qRT-PCR measurement of transcript abundance in vector-control samples in our transient transformation experiment led us to suspect that wounding and detachment may induce expression of some genes. To test the hypothesis, expression of the PR-1 and CRSP38 genes were compared in fresh, detached, and detached then *Phytophthora palmivora*-infected leaves from Sca6 and ICS1 plants (Fig. A.4). Ultimately, we found that fresh tissue exhibited relatively low and stable expression of these two genes in both genotypes. Detachment alone significantly induced expression of both genes in both genotypes (t-test $p < 0.05$). Expression levels in detached samples were only significantly different from detached then infected samples in one case: CRSP38 in ICS1.

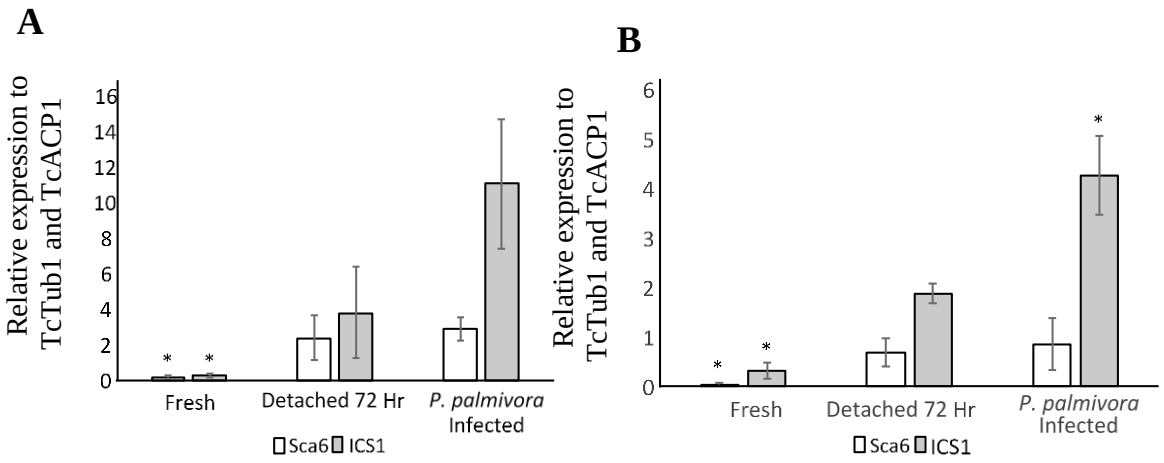


Fig. A.4 – qRT-PCR analysis of leaf detachment effect. **A)** Measurement of PR-1 and **B)** measurement of CRSP38. Bars represent means and error bars are standard deviation from 5 replicates. Asterisk denotes t-test $p < 0.05$ comparing the labelled bar to the Detached 72 Hr bar for the same genotype.

Discussion

In order to functionally screen the potentially thousands of genes with roles in defense, a system for prioritization is required. Here, we applied a strategy that took into account cacao gene expression data, QTL maps for resistance to disease, and comparative evolutionary genomics data, which in theory would be effective predictors of roles in defense. qRT-PCR validated our approach, showing that our transient expression system is reliably able to overexpress candidate genes. Nonetheless, transient overexpression of our target genes followed by pathogen inoculation did not demonstrate a lesion reduction phenotype.

In Chapters 4 and 5, we demonstrate that the transient transformation protocol can be used to reduce lesion size and pathogen biomass in cacao leaf tissue (Fister et al., 2015; Fister et

al., 2016). It is possible however that the genes described in those experiments play a greater individual role than any of those tested here; NPR1 is known to be a major defense regulator (Zhou et al., 2000; Dong, 2004; Yan and Dong, 2014), and chitinases are widely known to contribute to defense directly by degrading pathogen cell walls (Bishop et al., 2000; Grover, 2012). We cannot conclude that the genes tested here do not play a role in defense, but their overexpression in this interaction did not appear to reduce susceptibility.

The detached inoculation system may muddy the ability to perceive the effects of transient expression in our system. In the overexpression experiments shown here, some genes varied considerably in transcript abundance detected in vector control transformed samples. Subsequent experiments showed that detaching a leaf from the tree and preparing it for inoculation significantly increased expression of PR-1 and CRSP38, and level of expression in detached and mock inoculated leaves was in some cases no different than genes that were detached and inoculated with pathogen. Clearly, detachment triggers significant activation of the stress response, and it may modulate the defense system as a whole more than overexpression of a single gene using our transient transformation system. Activation of the plant's defenses resulting from detachment from the plant, therefore, could make detection of a phenotype caused by overexpression far more difficult using this system. Perhaps overexpression of only certain proteins, those with master regulatory roles in defense or critical, direct antimicrobial functions, such NPR1 and the chitinase tested in Chapter 5, can reduce lesion size in our system. Having found this flaw in our approach, future experiments will rely on transformation and inoculation of leaves that remain attached to the plant.

Another possible reason for the lack of effect is specificity of interaction between the plant genotype and the pathogen isolate. While we used Sca6 to clone genes because it confers resistance and was used to identify several important QTL, the microarray experiments assessing *P. palmivora* and *C. theobromicola* treatment on cacao seedlings used seedlings from an open-pollinated UF17 mother. Further the experiments used different *Phytophthora palmivora* isolates. Induction of certain genes, especially those with roles in binding and recognizing pathogen proteins, may only be effective for protecting the trees in specific plant-pathogen genotype pairs. While we are limited in which germplasm resources are available at Penn State, evaluation of genotype-isolate interaction elsewhere is possible and is vital to better understanding of specificity in defense.

Conclusions

Here we defined a strategy for defense gene prioritization based on gene expression data, QTL maps, and comparative genomics. Six defense genes were successfully cloned from cacao genomic DNA and overexpression vectors were constructed. *Agrobacterium*-mediated

transient overexpression successfully increased target gene transcript abundance, but overexpression did not produce a visible phenotype after assessing lesion size. While the approach is effective for defense gene overexpression, it may not be a broadly effective tool for functional analysis of all defense genes.

References

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JSS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of *Theobroma cacao*. Nature Genetics **43:** 101-108

Bishop JG, Dean AM, Mitchell-Olds T (2000) Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. Proceedings of the National Academy of Sciences of the United States of America **97:** 5322-5327

Britto DS, Pirovani CP, Andrade BS, dos Santos TP, Pungartnik C, Cascardo JCM, Micheli F, Gesteira AS (2013) Recombinant β-1,3-1,4-glucanase from Theobroma cacao impairs Moniliophthora perniciosa mycelial growth. Molecular Biology Reports **40:** 5417-5427

Brown JS, Schnell RJ, Motamayor JC, Lopes U, Kuhn DN, Borrone JW (2005) Resistance Gene Mapping for Witches' Broom Disease in *Theobroma cacao L.* in an F2 Population using SSR Markers and Candidate Genes. Journal of the American Society for Horticultural Science **130:** 366-373

Chae L, Kim T, Nilo-Poyanco R, Rhee SY (2014) Genomic Signatures of Specialized Metabolism in Plants. Science **344:** 510-513

De Bruyne L, Höfte M, De Vleesschauwer D (2014) Connecting Growth and Defense: The Emerging Roles of Brassinosteroids and Gibberellins in Plant Innate Immunity. Molecular Plant **7:** 943-959

De Lorenzo G, Ferrari S (2002) Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. Current Opinion in Plant Biology **5:** 295-299

Dong X (2004) NPR1, all things considered. Current Opinion in Plant Biology **7:** 547-552

Faleiro FG, Queiroz VT, Lopes UV, Guimarães CT, Pires JL, Yamada MM, Araújo IS, Pereira MG, Schnell R, de Souza Filho GA (2006) Mapping QTLs for witches' broom (Crinipellis perniciosa) resistance in cacao (Theobroma cacao L.). Euphytica **149:** 227-235

Fister AS, O'Neil ST, Shi Z, Zhang Y, Tyler BM, Guiltinan MJ, Maximova SN (2015) Two Theobroma cacao genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. J Exp Bot **66**

Fister AS, O'Neil ST, Shi Z, Zhang Y, Tyler BM, Guiltinan MJ, Maximova SN (2015) Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. Journal of Experimental Botany **66:** 6245-6258

Fister AS, Shi Z, Zhang Y, Helliwell EE, Maximova SN, Guiltinan MJ (2016) Protocol: transient expression system for functional genomics in the tropical tree Theobroma cacao L. Plant Methods **12:** 1-13

Freeling M (2009) Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. Annual Review of Plant Biology **60:** 433-453

Grover A (2012) Plant chitinases: genetic diversity and physiological roles. Critical Reviews in Plant Sciences **31:** 57-73

Gutiérrez OA, Campbell AS, Phillips-Mora W (2016) Breeding for Disease Resistance in Cacao. *In* AB Bailey, WL Meinhardt, eds, Cacao Diseases: A History of Old Enemies and New Encounters. Springer International Publishing, Cham, pp 567-609

Lanaud C, Fouet O, Clément D, Boccara M, Risterucci AM, Surujdeo-Maharaj S, Legavre T, Argout X (2009) A meta–QTL analysis of disease resistance traits of Theobroma cacao L. Molecular Breeding **24:** 361-374

Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods **25:** 402-408

Maximova S, Florez S, Shen X, Niemenak N, Zhang Y, Curtis W, Guiltinan M (2014) Genome-wide analysis reveals divergent patterns of gene expression during zygotic and somatic embryo maturation of *Theobroma cacao L*., the chocolate tree. BMC Plant Biology **14:** 185

Maximova S, Marelli J-P, Young A, Pishak S, Verica J, Guiltinan M (2006) Over-expression of a cacao class I chitinase gene in *Theobroma cacao L*. enhances resistance against the pathogen, *Colletotrichum gloeosporioides*. Planta **224:** 740-749

Mejía LC, Herre EA, Sparks JP, Winter K, García MN, Van Bael SA, Stitt J, Shi Z, Zhang Y, Guiltinan MJ, Maximova SN (2014) Pervasive effects of a dominant foliar endophytic fungus on host genetic and phenotypic expression in a tropical tree. Frontiers in Microbiology **5**

Meng X, Zhang S (2013) MAPK cascades in plant disease resistance signaling. Annu Rev Phytopathol **51**

Misas-Villamil JC, van der Hoorn RAL (2008) Enzyme–inhibitor interactions at the plant–pathogen interface. Current Opinion in Plant Biology **11:** 380-388

Miyakawa T, Hatano K-i, Miyauchi Y, Suwa Y-i, Sawano Y, Tanokura M (2014) A Secreted Protein with Plant-Specific Cysteine-Rich Motif Functions as a Mannose-Binding Lectin That Exhibits Antifungal Activity. Plant Physiology **166:** 766-778

Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biol **14**

Naseem M, Kaltdorf M, Dandekar T (2015) The nexus between growth and defense signalling: auxin and cytokinin modulate plant immune response pathways. Journal of Experimental Botany

O'Brien J, Daudi A, Butt V, Paul Bolwell G (2012) Reactive oxygen species and their role in plant defense and cell wall metabolism. Planta **236:** 765-779

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiology **150:** 1648-1655

Ploetz R (2016) The Impact of Diseases on Cacao Production: A Global Overview. *In* Cacao Diseases. Springer, pp 33-59

Pound FJ (1943) Cacao and witches' broom disease (Marasmius perniciosa). Report on a recent visit to the Amazon territory of Peru, September, 1942 - February, 1943. Government Print, Port of Spain

Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theoretical and Applied Genetics **108:** 168-174

Royaert S, Jansen J, da Silva DV, de Jesus Branco SM, Livingstone Iii DS, Mustiga G, Marelli J-P, Araújo IS, Corrêa RX, Motamayor JC (2016) Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of Theobroma cacao L. in Brazil. BMC Genomics **17:** 1-16

Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Research **38:** D814-D821

Sarowar S, Kim YJ, Kim EN, Kim KD, Hwang BK, Islam R, Shin JS (2005) Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. Plant Cell Reports **24:** 216-224

Sawano Y, Miyakawa T, Yamazaki H, Tanokura M, Hatano K-i (2007) Purification, characterization, and molecular gene cloning of an antifungal protein from Ginkgo biloba seeds. Biological Chemistry **388:** 273

Shi Z, Maximova S, Liu Y, Verica J, Guiltinan M (2010) Functional analysis of the *Theobroma cacao* NPR1 gene in arabidopsis. BMC Plant Biology **10:** 248

Teixeira PJPL, Thomazella DPdT, Reis O, do Prado PFV, do Rio MCS, Fiorin GL, José J, Costa GGL, Negri VA, Mondego JMC, Mieczkowski P, Pereira GAG (2014) High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. The Plant Cell **26:** 4245-4269

Ülker B, Somssich IE (2004) WRKY transcription factors: from DNA binding towards biological function. Current Opinion in Plant Biology **7:** 491-498

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology **44:** 135-162

van Verk MC, Bol JF, Linthorst HJ (2011) Prospecting for Genes involved in transcriptional regulation of plant defenses, a bioinformatics approach. BMC Plant Biology **11:** 1-12

van Verk MC, Bol JF, Linthorst HJM (2011) WRKY Transcription Factors Involved in Activation of SA Biosynthesis Genes. BMC Plant Biol **11**

van Verk MC, Gatz C, Linthorst HJM (2009) Transcriptional Regulation of Plant Defense Responses. *In* LC van Loon, ed, Adv Bot Res,

Yan S, Dong X (2014) Perception of the plant immune signal salicylic acid. Current Opinion in Plant Biology **20:** 64-68

Yao C, Conway WS, Ren R, Smith D, Ross GS, Sams CE (1999) Gene encoding polygalacturonase inhibitor in apple fruit is developmentally regulated and activated by wounding and fungal infection. Plant Molecular Biology **39:** 1231-1241

Zhou J-M, Trifa Y, Silva H, Pontier D, Lam E, Shah J, Klessig DF (2000) NPR1 Differentially Interacts with Members of the TGA/OBF Family of Transcription Factors That Bind an Element of the PR-1 Gene Required for Induction by Salicylic Acid. Molecular Plant-Microbe Interactions **13:** 191-202

Appendix B:

Supplemental Data

Chapter 2

Supplementary data are available at *BMC Genomics* online.

http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2693-3

Microarray data are available at NCBI (GEO: GSE73804).

Supplementary Data

Figure S2.1. Karyogram depicting the position of PR genes along the length of chromosomes based on the Matina genome sequence.

Figure S2.2. Maximum-likelihood phylogeny of Criollo and Arabidopsis PR-1 family members.

Figure S2.3. Maximum-likelihood phylogeny of Criollo and Arabidopsis PR-4 family members.

Figure S2.4. Whole genome gene expression profiles in microarray-analyzed leaves. Scatterplots of $\log_2$ normalized expression values for all probes on the microarray, comparing pathogen treatments with water treatment.

Figure S2.5. Representative photographs showing leaves 72 h after A) H_2O, B) *C. theobromicola* (with red lines indicating developing lesions), and C) *P. palmivora* treatment. Scale bars represent 1 cm.

Table S2.1. PR gene gamily type members: GenBank accession numbers for PR type member amino acid sequences used as BLASTp queries.

Table S2.2. Gene IDs and positions of Criollo PR genes mapped to the ten cacao chromosomes. Those not mapped to the ten chromosomes are appended to the bottom of the list without positional information.

Table S2.3. Gene IDs and positions of Matina PR genes mapped to the ten cacao chromosomes. Those not mapped to the ten chromosomes are appended to the bottom of the list without positional information.

Table S2.4. Gene IDs and BLASTp E-values for *Arabidopsis thaliana* PR loci.

Table S2.5. Gene IDs and BLASTp E-values for *Brachypodium distachyon* PR loci.

Table S2.6. Gene IDs and BLASTp E-values for *Oryza sativa* PR loci.

Table S2.7. Gene IDs and BLASTp E-values for *Populus trichocarpa* PR loci.

Table S2.8. Gene IDs and BLASTp E-values for *Vitis vinifera* PR loci.

Table S2.9. Percentage of PR genes in tandem arrays in the six analyzed plant species.

Table S2.10. Percent identities for Criollo PR-1 genes, color-coded to show tandem array members.

Table S2.11. Percent identities for Criollo PR-3 genes, color-coded to show tandem array members.

Table S2.12. Percent identities for Criollo PR-4 genes, color-coded to show tandem array members.

Table S2.13. Log_2 normalized expression values for all PR genes on microarray, with values averaged across five biological replicates.

Table S2.14. Log_2 fold change for all significantly regulated (Benjamini-Hochberg $p < 0.05$) PR genes on microarray.

Table S2.15. Sequences of qRT-PCR primers for validation of PR-1, PR-3, and PR-4 family expression.

Chapter 3

Supplementary data are available at *JXB* online.

https://jxb.oxfordjournals.org/content/early/2015/07/09/jxb.erv334

Supplementary Fig. S3.1. Vector map for the binary plasmid used for transient overexpression of TcNPR1.

Supplementary Table S3.1. All genes in the Sca6 genotype with differential regulation from SA treatment.

Supplementary Table S3.2. All genes in the ICS1 genotype with differential regulation from SA treatment.

Supplementary Table S3.3. Genes showing statistically significant differential regulation in both Sca6 and ICS1 genotypes.

Supplementary Table S3.4. Differential regulation of PR genes according to genotype, as detected by the microarray.

Supplementary Table S3.5. Enriched GO terms as detected by PAGE for Sca6.

Supplementary Table S3.6. Enriched GO terms as detected by PAGE for ICS1.

Supplementary Table S3.7. Enriched GO terms for SA effect when averaging all samples, across genotypes.

Supplementary Table S3.8. GO term enrichment by PAGE comparing basal (water-treated) and induced (SA-treated) states of Sca6.

Supplementary Table S3.9. GO term enrichment by PAGE comparing basal (water-treated) and induced (SA-treated) states of ICS1.

Supplementary Table S3.10. Primers used for qRT-PCR.

Chapter 4

Supplementary Table S4.1. Genotypes used for genotyping and population genetics analyses.

French Guiana			Nacional			Pound		
Geno-type	In genotyp-ing analysis?	In DNAsp analysis ?	Genotype	In genotypin g analysis?	In DNAsp analysis ?	Genotype	In genotyping analysis?	In DNAsp analysis ?
ELP1-A	X	X	AM 2/96 F5BI8I9T1 5 Stock	X	X	IMC 102	X	X
ELP11-A	X	X	CLEM/S-62-1 F5B I745 T4 Stock	X	X	IMC 12 1054	X	X
ELP16-S1	X	X	SLC 8 F5A A13 T11 Stock	X	X	IMC 3	X	X
ELP20-A	X	X	AM 1/28 A1 T4 Stock	X	X	IMC 44	X	X
ELP25-A	X	X	LV20 F5B B83 T12A	X	X	IMC 45	X	X
ELP34-A	X	X	MOQ 222 F6A B81 T9 SL	X	X	IMC 47	X	X
ELP7-S2	X	X	SLA 16 F5B D242 T13 DIL	X	X	IMC 49	X	X
ELP9-A	X	X	MOQ S5 P6A B82 T13 SL	X	X	IMC 50	X	X
GU100-A	X	X	MOQ 534 F4A D358 T2	X	X	IMC 73	X	X
GU102-A	X	X	CL 19/21 F5A A46 T14 Stock	X	X	IMC 76-1123	X	X
GU138-A	X	X	MOQ 423 F4A D339 T4	X	X	IMC 94-1080	X	X
GU142-A	X	X	LP 3/25 F5A C236 T13NL Stock	X	X	IMC 94-640	X	X
GU145-A	X	X	LP 45 F4A E478 T2	X	X	IMC 96	X	X

GU148-A	X	X	LX43 F5B C201 T1 Stock	X	X	MO 17	X	X
GU150-1	X	X	CL 27/50 F5B I744D T6 Stock	X	X	MO 3	X	X
GU153-M	X	X	B9/10-33 F5B I768 T13 Stock	X	X	MO 4	X	X
GU159-7	X	X	B13/7 F5B I728 T14 Stock	X	X	MO 81	X	X
GU161-7	X	X	AM 2/31 F4A F560 T1 Stock	X	X	MO 9	X	X
GU163-6	X	X	AM 1/53 F4A F551 T1 Stock	X	X	MO 90 F6A B125 T3 Stock	X	X
GU167-7	X	X	LP 329 F5B I797 T8	X	X	NA 12	X	X
GU170-L	X	X	AM 1/19 F5B I771 T3B Stock	X	X	NA 140	X	X
GU173-A	X	X	MOQ 693 F5B C220 T2	X	X	NA 144	X	X
GU180-3	X	X	LP 1/21 F5B I779 T4A Stock	X	X	NA 170 F4A D390 T2 Stock	X	X
GU182-A	X	X	LP 4/7 F5B A34 T8 Stock	X	X	NA 176 F4A D389 T4 Stock	X	X
GU186-A	X	X	AM 2/83 F5B B108 T12 Stock	X	X	NA 178 F4A D388 T2 Stock	X	X
GU189-6	X	X	B14/9 F6B F484 T2 Stock	X	X	NA 189 F5A D268 T1	X	X
GU192-6	X	X	CLM 100 F5A C160 T9 Stock	X	X	NA 246 F5B E404 T8 Stock	X	X
GU196-2	X	X	JA 3/4 F4A E466 T2 Stock	X	X	NA 312 F5B G614 T2	X	X

GU199-6	X	X	LX45 F5B C213 T1 Stock	X	X	NA 34	X	X
GU20-1	X	X	AM 2/18 F5B H679 T15 Stock	X	X	NA 387	X	X
GU203-2	X	X	LX38 F5B C206 T5 Stock	X	X	NA 399	X	X
GU205-1	X	X	LP 4/12 F5B I803 T4 Stock	X	X	NA 406 F4A D409 T2 Stock	X	X
GU208-A	X	X	LX2 F4A D325 T2	X	X	NA 43	X	X
GU214-A	X	X	B9/10-32 F5B I766 T2 Stock	X	X	NA 471 F6A B92 T4 Stock	X	X
GU216-6	X	X	MOQ 652 F4A D356 T2	X	X	NA 61 F4A D375 T4 Stock	X	X
GU220-6	X	X	B10/28 F5A B83 T14 Stock	X	X	NA 669 F4A D418 T4 Stock	X	X
GU232-5	X	X	B12/1 F6B F461 T13 Stock	X	X	NA 670	X	X
GU240-A	X	X	JA 5/2 F5A C170 T5 Stock	X	X	NA7/11	X	X
GU244-1	X	X	LX1 F4A D335 T3	X	X	NA 702 F5B G631 T3 NL Stock	X	X
GU254-1	X	X	MOQ 699 F5B C188 T6	X	X	NA 712	X	X
GU258-A	X	X	LP 1/45 F5B B96 T16 Stock	X	X	NA 756	X	X
GU262-A	X	X	B17/17 F5B I784 T2 Stock	X	X	NA 79	X	X
GU270-A	X	X	LZ 13 F6B D249 T8 Stock	X	X	NA 804	X	X
GU274-A	X	X	JA 5/10 F5A C236	X	X	NA 807	X	X

			T13NL Stock					
GU278-A	X	X	CL 15/197 F5B A73 T10 Stock	X	X	NA 90	X	X
GU280-A	X	X	LP 4/32 F5B I754 T14 Stock	X	X	NA 92	X	X
GU283-1	X	X	LX41 F4A D334 T3	X	X	PA 12	X	X
GU287-A	X	X	B23/2 F6A A11 T13 Stock	X	X	PA 120	X	X
GU291-2	X	X	MOQ 421 F4A D361 T2	X		PA 121	X	X
GU294-M	X	X				PA 124	X	X
GU317-A	X	X				PA 125 F4A F511 T3 Stock	X	X
GU325-7	X	X				PA 13	X	X
GU332-A	X	X				PA 137	X	X
GU336-1	X	X				PA 151	X	X
GU340-1	X	X				PA 191	X	X
GU344-A	X	X				PA 2	X	X
GU350-M	X	X				PA 35	X	X
GU39-1	X	X				PA 39	X	X
GU44-7	X	X				PA 67	X	X
GU46-8	X	X				IMC 57-1063	X	X
GU60-1	X	X				NA 370 F4A D405 T3 Stock	X	X
GU8-5	X	X				NA 702-T2 Stock	X	X
GU86-1	X	X				Sca 10 1065	X	X
GU98-A	X	X				Sca10-1182	X	X
KER11-1-L	X	X				Sca 12	X	X
KER2-D	X	X				Sca 16	X	X

YAL3	X	X				Sca 20	X	X
YAL6-S1	X	X				Sca 23-1082	X	X
						Sca 23-273	X	X
						Sca 27	X	X
						Sca 3-643	X	X
						Sca6	X	X
						MO 109	X	
						MO82	X	
						NA 110 F4A D395 T3 Stock	X	
						NA111 F4A E495 T3 Stock	X	
						NA33	X	

Chapter 5

Supplemental material available online at Plant Methods:

http://plantmethods.biomedcentral.com/articles/10.1186/s13007-016-0119-5

Fig. S5.1. Multiple genotype infection graph – Box and whisker plots displaying lesions sizes 72 hours after inoculation of stage C cacao leaves of 17 genotypes with Phytophthora tropicale mycelia.

Fig. S52. Photographs of infected leaf tissue of diverse genotypes – Representative photographs 72 hours after inoculation of leaf tissue with Phytophthora tropicale. Scale bars represent 1 cm.

Supplemental Methods – Descriptions of protocols used for plant growth, transformation and photography of the five leaf stages, the force to puncture test, transformation of the eight cacao genotypes, and evaluation of effects of TcChi1 overexpression.

CURRICULUM VITA

Andrew S. Fister

EDUCATION

2011-2016 **Ph.D. in Genetics, Guiltinan-Maximova Lab**
Huck Institutes of the Life Sciences
Pennsylvania State University, University Park, PA 16801

2007-2011 **B.A.: Double Major in Biology, concentration in Genetics and Development and English Literature, cum laude in Biology**
College of Arts and Sciences
Cornell University, Ithaca, NY 14850

AWARDS AND HONORS

2011 **University Graduate Fellow, Penn State University**

2013 **Huck Graduate Dissertation Research Grant**

TEACHING EXPERIENCE

Fall 2012 and Fall 2013 **Biology 110: Basic Concepts and Biodiversity**
Led two lab sections per semester

Spring 2013 and Spring 2014 **Biology 220: Populations and Communities**
Led two lab sections per semester

PUBLICATIONS

Fister AS, Mejia LC, Zhang Y, Herre EA, Maximova SN, Guiltinan MJ (2016) *Theobroma cacao L*. pathogenesis-related gene tandem array members show diverse expression dynamics in response to pathogen colonization. BMC Genomics 17: 1-16.

Fister, A. S., Shi, Z., Zhang, Y., Helliwell, E. E., Maximova, S. N., & Guiltinan, M. J. (2016). Protocol: transient expression system for functional genomics in the tropical tree *Theobroma cacao* L. *Plant methods*, *12*(1), 1.

Fister, A. S., O'Neil, S. T., Shi, Z., Zhang, Y., Tyler, B. M., Guiltinan, M. J., & Maximova, S. N. (2015). Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. *Journal of experimental botany*, erv334.

SEMINARS

Defining cacao's induced defenses
Frontiers in Science and Technology for Cacao Quality, Productivity, and Sustainability, 2016

Functional Genomics of the Defense Response in Theobroma Cacao
Plant & Animal Genome, 2015