Using the Bacterial Plant Pathogen *Pseudomonas syringae* pv. tomato as a Model to Study the Evolution and Mechanisms of Host Range and Virulence

Shuangchun Yan

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Plant Physiology

Boris Vinatzer

John Jelesko

João Setubal

Brett Tyler

Liqing Zhang

December 6th, 2010
Blacksburg, VA

Keywords: *Arabidopsis thaliana*, *Pseudomonas syringae*, molecular evolution, population genetics, microbial genomics, HopM1, callose, host-microbe interaction

Using the Bacterial Plant Pathogen *Pseudomonas syringae* pv. tomato as a Model to Study the Evolution and Mechanisms of Host Range and Virulence

Shuangchun Yan

ABSTRACT

Most plant pathogens are specialists where only few plant species are susceptible, while all other plants are resistant. Unraveling the mechanisms behind this can thus provide valuable information for breeding or engineering crops with durable disease resistance. A group of *Pseudomonas syringae* strains with different host ranges while still closely related were thus chosen for comparative study. We confirmed their close phylogenetic relationship. We found evidence supporting that these strains recombined during evolution. The *Arabidopsis thaliana* and tomato pathogen *P. syringae* pv. tomato (Pto) DC3000 was found to be an atypical tomato strain, distinct from the typical Pto strains commonly isolated in the field that do not cause disease in *A. thaliana,* such as Pto T1. Comparing *A. thaliana* defense responses to DC3000 and T1, we found that T1 is eliciting stronger responses than DC3000. T1 is likely lacking Type III effector genes necessary to suppress plant defense. To test this, we sequenced the genomes of strains that cause and do not cause disease in *A. thaliana*. Comparative genomics revealed candidate effector genes responsible for this host range difference. Effector genes conserved in strains pathogenic in *A. thaliana* were expressed in T1 to test whether

they would allow T1 to growth better in *A. thaliana*. Surprisingly, most of them reduced T1 growth. One of the effectors, HopM1, was of particular interest because it is disrupted in typical Pto strains. Although HopM1 has known virulence function in *A. thaliana*, HopM1 reduced T1 growth in both *A. thaliana* and tomato. HopM1 also increased the number of bacterial specks but reduced their average size in tomato. Our data suggest that HopM1 can trigger defenses in these plants. Additionally, transgenic detritivore *Pseudomonas fluorescens* that can secrete HopM1 shows dramatically increased growth *in planta*. The importance of genetic background of the pathogen for the functions of individual effectors is discussed. T1 cannot be manipulated to become an *A. thaliana* pathogen by deleting or adding individual genes. We now have a list of genes that can be studied in the future for the molecular basis of host range determination.

Dedication

This dissertation is dedicated to my parents, all the teachers I had, and all the people that inspired me to be a scientist. Without them, this dissertation would have never been possible.

Acknowledgement

It took me five years plus another semester of rotation to get to this day, when I am sitting in front of my computer and typing this acknowledgement, while the rest of the dissertation is almost complete.

I have been thinking about whom I would love to thank for a while now. I decided to dedicate this dissertation to all the teachers I ever had, because rationally and factually, I understand the importance and deeply appreciate their mentorship. It is not until a couple of weeks ago when I went to see the documentary Waiting for Superman at the Lyric theater, did I feel the overwhelming gratefulness in a very real form.

I can't list all the teachers, but I would like to list a few of them for what they have taught me: (in chronicle order)

Mrs. Hou Fang and **Mrs. Cheng Mingming**, especially for their care and enlightenment in music and performance art in pre-school. **Mrs. Zhang Xiaorong** taught Putonghua (Mandarin) so well in elementary school that I was immediately better than peers, and chose me to perform music and other art related affairs. **Mrs. Wen Jianmin** and **Mrs. Zhang Qingqing** for their inspiration in both music, dancing and art in elementary school.

In middle school, **Mrs. Wang Yuanlin** gave me constant care and excellent teaching in mathematics in middle school. **Mrs. Zhi Qin** was my first English

teacher who helped building great foundation for me. **Mrs. He Ping** and **Mrs. Huang Xuejuan** taught Biology classes in middle school, and I was always very intrigued in their classes. Special thanks for **Mrs. Huang Xuejuan** for publically announcing that I was her favorite student several times that my classmates all made fun of me. **My mother** taught me all kinds of stuff in life, and she was the best chemistry teacher in my middle school. **Mrs. Zhang Runyuan** for teaching me Chinese in middle school. I deeply enjoyed writing because of **Mrs. Zhang** for being my best audience.

Mr. Wan Yinghong in high school was such an inspiring role model and great chemistry teacher. **Mrs. Yang Xiaoling** for being the best English teacher ever, and for helping me learn the language the way I want to learn, and experiment with different learning techniques on my own without forcing me to listen to her lecture. I really cannot ask for more. **Mrs. Liu Ying** and **Mr. Fan Zhongli** for teaching me Biology in high school. Mrs. Liu was so inspiring and encouraging that she played the key role for helping me become a biologist. Without her organizing the biology competition I may not have majored in Biology in college. Both of them dedicated extra time to train me for the 2nd round competition at Sichuan University. And Mr. Fan traveled with me to Sichuan University for the competition. It was one of the best experiences in high school. **Mr. Zhao Deyou** who made my mathematics classes in high school less dreadful,

at least for a couple of semesters. I wished that I could have you as my mathematics teacher for the entire high school. I really enjoyed the problem solving classes; the feelings of being inspired were so awesome. **Dr. Wang Shenghua** let me work under him as undergraduate research assistant. It was great experience.

I want to thank **Dr. Boris A. Vinatzer** for picking out my application sent to the PPWS, and for deciding to bring me into MPS program, which was definitely the No.1 reason why I am here at Virginia Tech today. His advisory, open-minded teaching style, his candidate discussions with me about science, society and life, will all become precious memory of mine. The whole inter-disciplinary research project totally broadened my knowledge in biological science. I would also like to thank my degree committee members for their support, suggestions and help through out the whole process.

Other friends, people and non-human things I would like to thank because without them, I will never be who I am today and where I am (in chronicle order): Qian Yuan (you are the hipster!), Zhou Shi, Aijing (the first artistic role model), Englishtown.com, Yahoo!, Auli Pekkonon, Wenquxing eDictionary, Apple Computer Inc., 21st Centaury (newspaper) and other writers and books that showed me the world and the futuristic vision, Zhao Huajie, Hu Tao, Wu Yuan, Bjork, Xie Gang, Dong Ke, Wang Lu, Liu Yi, Li Jing, Naji R. Naccache, my uncle (Yu

Wenye), Guo Liheng, Li Yong, GNU/Linux and OSS community, Guo Zheying, Yang Yihong, Amos Desjardin, Sun Jian, Huang Mengsu, Lin Wen, Ted Chernesky, Xinhao Ye and Will Waddlington.

Whole-hearted thanks to all previous and current lab members, especially Toni Mohr, Jenny Jenrette, Megan Moore, Christopher Clarke, Rongman Cai, Haijie Liu, Ashley Pearson and Kimberly Duvall.

Special thanks to my colleagues in Latham Hall and previously Fralin Biotechnology Center. Especially to those who provided me with any assistance, free chat and free food. You know whom you are, but I am not going to try to make an exhaustive list.

I sincerely thank all the people that had long conversations with me. It is these conversations that get me through tough times, and expanded me mentally.

Finally I thank my family. Their love and support are priceless. I love you. And special thanks to my father's annoyed comments to the then little me when I asked him too many questions. You said "You kid like to get to the bottom of things by persistently asking many questions". Although you were totally annoyed at that time, I took that as encouragement.

Table of Contents

Chapter	Page
Abstract	**ii**
Dedication	**iii**
Acknowledgements	**iv**
List of Tables	**xiii**
List of Figures	**xiv**
List of Abbreviations	**xv**

Chapter 1 1

Introduction and Literature Review ***1***

a. General Background **2**

b. Diversity, Population Genetics and Evolution of *P. syringae* **6**

c. Overview of Virulence mechanisms of Pathogens and Plant Defenses 12

d. The T3SS and T3Es **16**

e. Determination of Host Range, and Evolution of Host Range and Virulence **19**

f. Overview of research **25**

g. References **25**

Chapter 2 42

Phylogenetics, Population Genetics and Molecular Evolution of a group of Pseudomonas syringae strains ***42***

a. Abstract **45**

b. Introduction **46**

c. Materials and Methods **48**

Bacterial isolates. 48

PCR and DNA sequencing of gene fragments. Primers were designed on 24 *P. syringae* genes. 48

Molecular evolutionary analysis. 49

Plant growth conditions and bacterial infections. 51

Nucleotide sequence accession numbers. 52

d. Results **52**

The sequenced PtoDC3000 belongs to a group of closely related isolates from cultivated and wild plants around the world. 52

Isolates very closely related to PtoDC3000 have different host ranges. 56

Recombination contributed to the evolution of the core genome of PtoDC3000 and closely related isolates. 58

Acquisition and loss of the *avrPto1* PAI and its role in host range evolution. 65

e. Discussion **68**

Importance of recombination in the evolution of closely related *P. syringae* strains. 68

Role of recombination in determining the distribution of the effector *avrPto1*. 70

Revealed identity of PtoDC3000. 71

f. Acknowledgements **72**

g. References **72**

h. Supplementary Materials **81**

Chapter 3 87

Genome Sequencing of Pseudomonas syringae strains and Identification of Genomic Differences ***87***

a. Abstract **90**

b. Introduction **91**

c. Results **94**

T1 is a nonpathogen of *A. thaliana* ecotypes and causes a HR. 94

T1 draft genome sequence obtained using Solexa sequencing and 454 pyrosequencing and compared with the DC3000 genome. 96

Alignment of the T1 genome with the three sequenced *P. syringae* genomes98

Comparison of protein complements between the four *P. syringae* genomes.98

Prediction of the T3SS effector repertoire of T1 and comparison with the DC3000 effector repertoire. 100

The presence of *avrRpt2* in the T1 genome is not sufficient to explain the inability of T1 to cause disease in *A. thaliana*. 104

Can gene for gene interactions other than *avrRpt2-RPS2* explain the inability of T1 to cause disease in *A. thaliana*? 105

d. Discussions **108**

e. Materials and Methods **111**

Bacterial strains and growth media. 111

Plant growth and infections. 112

Genome sequencing and assembly. 112

Annotation. 113

Comparative genomics. 114

Effector prediction, cloning, *P. syringae* transformation, and testing for translocation. 114

f. Acknowledgements **114**

g. Literature Cited **115**

AUTHOR-RECOMMENDED INTERNET RESOURCES 123

Chapter 4 124

Investigation of the Basis for Non-host Resistance of A. thaliana to Pto T1 and the Roles of Type III Effectors in Host Range Determination with HopM1 as Case Study ***124***

a. Abstract **125**

b. Introduction **127**

c. Results **130**

A. thaliana Resistance to Pto T1 is a kind of multi-layered non-host resistance 132

Comparisons of Arabidopsis defense responses to T1 and DC3000 135

Co-infection of DC3000 with T1 slightly lowers DC3000 grows and dramatically increases T1 growth 139

Genome sequencing of *P. syringae* strains closely related to both T1 and DC3000 revealed conserved effector repertoire differences 140

HopAS1 in T1 elicits ETI 145

DC3000 effectors do not significantly increase T1' or T1* growth in *A. thaliana* 145

The presence of *hopM1* is actively selected against in typical Pto strains 147

Truncated HopM1 alleles lost their ability to cause cell death in tomato and *N. benthamiana* 148

The inability of truncated HopM1 alleles to cause cell death is not due to different cellular localization 151

HopM1 reduces T1 growth in tomato, reduces bacterial speck size, and causes wilting of tomato 154

HopM1 alone allow *Pseudomonas fluorescens* to grow to pathogen-level population density in both *N. benthamiana* and *A. thaliana* 156

d. Discussions **158**

e. Materials and Methods **167**

Plant growth conditions and bacterial infections 167

Callose deposition 168

Construction of plasmids and genetic manipulation of bacterial strains 169

Genome Sequencing and Draft Genome Assembly 170

Effector Repertoire Prediction 170

HopM1 cloning, transient expression 171

Western blot 171

Confocal Microscopy and co-localization analysis 172

PCR and Real-time PCR 173

f. Author attributions 174

g. References 174

Supplementary Figures 183

Chapter 5 184

Conclusions 184

a. Summary and Conclusions 185

b. References 190

Appendix 193

Contribution to other manuscripts 193

List of Tables

Table 2.1: *P. syringae* strains used in this study. 53

Table 2.2: Strains belonging to the STs in the Bayesian consensus tree (Fig. 2.2) and the NeighborNet network (Fig. 2.4). 57

Table 2.3: Percent DNA identities between isolates and dN/dS ratio. 60

Table 2.4: Recombination breakpoints predicted in gene fragments. 63

Table 2.5: Estimate of ρ and θ. 64

Table 2.6: SH test of gene partitions. 66

Table 2.S1: Primer sequences. 86

Table 3.1: Comparison of the Pto T1 draft genome sequences obtained by two methods. 96

Table 3.2: Test the translocation of Pto T1 effector candidate-Cya fusion proteins. 103

Table 3.3: Repertoire of T3E and T3SS-independent genes in T1 versus DC3000. 103

Table 3.4: *Pseudomonas syringae* pv. *tomato* isolates used in this study. 108

Table 4.1: Strains and plasmid used in this study. 130

Table 4.2: Characteristics of genome assemblies. 143

Table 4.3: List of conserved effectors differences. 144

Table 4.4: Real-Time PCR primers. 174

List of Figures

Figure 2.1: Positions of all analyzed genes in the genome of DC3000. 55

Figure 2.2: Bayesian tree and host range test of representative isolates. 59

Figure 2.3: Virulence test of representative *P. syringae* isolates on tomato. 61

Figure 2.4: NeighborNet analysis of the nine concatenated housekeeping genes. 62

Figure 2.S1: Pictures of infected leaves and population sizes of bacteria in leaves. 81

Figure 2.S2: Phylogenetic networks. 84

Figure 2.S3: DNA alignments. 85

Figure 3.1: Comparison of in planta growth and symptoms of DC3000 and T1. 95

Figure 3.2: Pairwise alignments of the T1 draft genome and the DC3000 genome. 97

Figure 3.3: Venn Diagram comparing the protein repertoire of T1, DC3000, Psy B728a and Pph 1448A. 100

Figure 3.4: AvrRpt2 alone cannot explain *A. thaliana* resistance to T1. 106

Figure 3.5: Bacterial growth in *SGT1* and *RAR1* mutants. 107

Figure 4.1: *A. thaliana* resistance to T1 is non-host resistance and multi-genic. 134

Figure 4.2: Comparison of T1 and DC3000 after inoculation in *A. thaliana*. 138

Figure 4.3: Co-infection of T1 and DC3000. 140

Figure 4.4: Phylogenetic tree and the abilities to cause disease in *A. thaliana* for representative *P. syringae* strains. 142

Figure 4.5: Growth of T1 WT, T1’ and T1* with individual DC3000 effectors in plants. 146

Figure 4.6: Different HopM1 alleles have different abilities to cause cell death but the same cellular localization. 150

Figure 4.7: Cellular localization of HopM1. 153

Figure 4.8: HopM1 effects in tomato when delivered by either T1 or DC3000. 154

Figure 4.9: Growth of *P. fluorescens* Pf0-1 EtHAn with HopM1 and symptoms it causes in *A. thaliana* and *N. benthamiana*. 157

Figure 4.S1: Different callose deposition morphologies. 183

List of Abbreviations

HR: Hypersensitive reaction

T3E: Type III effector

T3SS: Type III secretion system

A. thaliana: *Arabidopsis thaliana*

N. benthamiana: *Nicotiana benthamiana*

P. syringae: *Pseudomonas syringae*

P. fluorescens: *Pseudomonas fluorescens*

COR: coronatine

CFU: colony forming units

CEL: conserved effector locus

PAI: pathogenicity island

Pto: *P. syringae* pv. tomato

Psy: *P. syringae* pv. syringae

Pph: *P. syringae* pv. Phaseolicola

hrp: HR and pathogenicity

hrc: HR and conserved

Chapter 1

Introduction and Literature Review

a. General Background

The field of host–pathogen interactions is an important subject of microbiology. The interaction between pathogens and their hosts is a type of parasitism while parasitism is a type of symbiotic relationship in which only the parasite benefits from the host (Newton, Fitt et al. 2010). Whether the parasite actively causes damage and/or suffering to the host, thereby causing disease, is the most important criterion whether a parasite is considered a pathogen (Newton, Fitt et al. 2010).

Diseases of humans and organisms closely associated with them, such as livestocks and domesticated crop plants, have a significant impact on many different aspects of society. Some pandemics of infectious diseases in human history wiped out a large percentage of the population, for example the plague epidemic caused by *Yersinia pestis* in medieval Europe (Drancourt and Raoult 2002; Morelli, Song et al. 2010). Indeed, many infectious diseases are either lethal, or cause tremendous suffering to the hosts. Infection of people, livestock and agricultural crops ultimately result in significant economic loss to society, sometimes resulting in, or contribute to, social instability (Council 2002; Friedman, Kippax et al. 2006), famine (Stukenbrock and McDonald 2008) and possibly extinction of civilization or loss of major portion of population, such as the smallpox and measles epidemics among Native Americans in the 16th century (Patterson and Runge 2002). Because most pathogens are microscopic, such as viral, bacterial and fungal pathogens, infectious diseases were mysterious to humans before the invention of the microscope. Since then, tremendous progress has been made towards diagnostics, prevention, and treatment.

During the past hundred years, biological weapons have been developed and used. An example is the dropping of bags filled with *Yersinia pestis*-infected fleas

over Chinese cities by the Unit 731 of Imperial Japanese Army during the second World War (Yamaguchi 2001; Barenblatt 2004; 凤凰卫视 2009). During the past decades, bioterrorism has also become a concern, exemplified by the anthrax attacks after 9/11 in the United States (Wills, Leikin et al. 2008). Additionally, emergence of new infectious diseases and increased cases of multi-drug resistant “superbugs” (Ferber 2010; Ippolito, Leone et al. 2010) are threats to public health and safety. Examples are HIV, Ebola virus, SARS virus, Avian influenza virus H5N1, or MRSA (methicilin resistant *Staphylococcus aureus*) (Ippolito, Leone et al. 2010). Therefore, large funds are invested into basic and applied research of human/animal infectious disease each year worldwide.

For agriculture, disease management and the cultivation and engineering of disease resistant crops have been priorities. The introduction of pathogens to new territories often causes significant losses, such as the recent introduction of the citrus greening pathogen *Liberibacter asiaticus* to Florida or the Sudden Oak Death pathogen *Phytophtora ramorum* to California (Rizzo, Garbelotto et al. 2005; Tyler, Tripathy et al. 2006; Duan, Zhou et al. 2009). Some plant pathogens not yet found in the USA are listed as select agents by the US Department of Agriculture (USDA), as they could potentially wipe out an entire industry once introduced into the USA.

Another concern is emergence of highly virulent pathogens, possibly from animal pathogens previously incompatible with human. Host jumps of pathogens sometimes lead to very deadly diseases that are not advantageous to the pathogen because the host population is quickly reduced. During co-evolution of host and pathogen, over time hosts are believed to become more resistant and pathogens less virulent (Lipsitch and Moxon 1997; Stukenbrock and McDonald 2008).

To researchers, some of the most intriguing questions about pathogen-host interactions are: 1) How do organisms evolve into pathogens? 2) How do pathogens adapt to their hosts? 3) How do pathogens and hosts co-evolve? 4) What are the pathogen virulence mechanisms? 5) What are durable strategies for preventing and treating infectious diseases? To help providing answers to these questions, plant pathologists and microbiologists have widely used the Gram-negative bacterial plant pathogen *Pseudomonas syringae* as a model and tool.

P. syringae is of great diversity and has been analyzed phylogenetically (Sarkar and Guttman 2004; Hwang, Morgan et al. 2005; Guttman, Gropp et al. 2006). It can be considered a species complex (Vinatzer and Bull 2009). Many strains known as *P. syringae* today were only intended to be part of this species temporarily since the species to which they originally belonged to were not included in the list of approved names in 1980 because 1) no physiological differences had been identified between them and the type strain of *P. syringae*, 2) no type strains had been deposited for these species (Vinatzer and Bull 2009). Therefore, the formerly named species, for example, *Pseudomonas tomato* became *Pseudomonas syringae* pathovar (pv.) *tomato* (Pto), where the pathovar concept was introduced to retain the information about the host of isolation (Dye, Bradbury et al. 1980). However, based on DNA similarity, the current *P. syringae* species could be divided into several different species (reviewed by Vinatzer & Bull, 2009).

Based on plant species of origin, plant pathologists have described over 50 *P. syringae* pathovars (Cunnac, Lindeberg et al. 2009). The most intensively studied pathovars include pv. *tomato*, pv. *phaseolicola* and pv. *syringae.* Three representative strains of *P. syringae* from these three pathovars have been completely sequenced and annotated (Buell, Joardar et al. 2003; Feil, Feil et al. 2005; Joardar, Lindeberg et al. 2005).

P. syringae strains have a Type III Secretion System (T3SS). The T3SS has been under intensive investigation during the last decade because of its importance in pathogenesis. *P. syringae* has been one of the model organisms for the study of the T3SS and the so-called "effector" proteins secreted by it (Alfano & Collmer, 2004, Greenberg & Vinatzer, 2003). The T3SS is also the key component of many other Gram-negative bacterial plant and animal pathogens, for example, *Yersinia*, *Escherichia*, *Salmonella*, *Shigella*, *Erwinia*, *Xanthomonas* and *Ralstonia* (Collmer *et al.*, 2000). T3SS effectors either function inside host cells, regulate the secretion of other effectors, or facilitate the secretion process (Greenberg and Vinatzer 2003). Collectively, these effectors suppress the plant defense responses, manipulate plant metabolism, allow pathogen population to grow exponentially, and ultimately cause disease. The T3SS and effectors may also be beneficial for competition with fungi or protists in the environment (Matz, Moreno et al. 2008; Munkvold, Martin et al. 2008).

Not only is the T3SS structurally conserved between bacterial plant and animal pathogens, effectors of animal and plant pathogens also share common virulence mechanisms (Galán 2009). For example, AvrPtoB from *P. syringae* and effectors in the SspH family from *Salmonella enterica* are both molecular mimicry of E3-ubiquitin ligase that targets proteins for degradation. Furthermore, animals and plants share common strategies of innate immunity. For instance, both animals and plants can recognize the flagellin of bacteria using receptor kinases (Gomez-Gomez and Boller 2000; Hayashi, Smith et al. 2001). It is plausible that these mechanisms of innate immunity evolved before animals and plants diverged, although it is more likely the result of convergent evolution (Ausubel 2005). Nonetheless, studying the plant immune system can still provide a whole body of knowledge about defense strategies and the biochemistry of the interactions between pathogen effectors and host proteins.

Some strains of *P. syringae* are able to cause disease on the model plant *Arabidopsis thaliana*. As a result, the *P. syringae–A. thaliana* system has become one of the most frequently used models for studying plant-microbe interactions and has led to many important discoveries, such as the understanding of plant resistance genes against pathogen effectors (Martin, Brommonschenkel et al. 1993), molecular mimicry and the utilization of a host apparatus for destroying host targets (Janjusevic, Abramovitch et al. 2006; Rosebrock, Zeng et al. 2007).

Because of the diversity of host range that exists within the *P. syringae* group, comparative studies of *P. syringae* pathovars can give insight into host range evolution and determination. Moreover, relatives of *P. syringae* in the *Pseudomonas* genus have adopted very different life styles. For example, *P. aeruginosa* is an opportunistic human pathogen, and some strains have been shown to cause disease symptoms in plants (Rahme, Stevens et al. 1995); *P. putida* is a saprophytic soil bacterium; *P. fluorescens* is primarily an inhabitant of soil and water but some *P. fluorescens* strains also cause soft rots in crops and some strains can be opportunistic pathogens of immune-compromised aquatic animals and humans (Hsueh, Teng et al. 1998; Zhang, Hu et al. 2009). Moreover, some *P. syringae* strains are lethal to aphids when infected (Stavrinides, McCloskey et al. 2009). This diversity makes the *Pseudomonas* genus an ideal object for comparative studies of plant and animal bacterial pathogenesis (Buell, Joardar et al. 2003).

b. Diversity, Population Genetics and Evolution of *P. syringae*

P. syringae, like many other bacterial species, has long been considered a clonal species (Denny, Gilmour et al. 1988; Sarkar and Guttman 2004). However,

Sarkar and Guttman came to the conclusion that *P. syringae* is a clonal species analyzing strains from distantly related lineages (Sarkar and Guttman 2004), which may not be appropriate for investigating the clonality of a species (discussed below). Recently, homologous recombination, a major mechanism for integrating foreign DNA and reshuffling genetic variation between bacteria, has been gaining more and more attention due to advances in our ability to detect recombination. One advancement was multi-locus sequence typing (MLST) proposed in 1998 (Maiden, Bygraves et al. 1998) and reviewed by Maiden (Maiden 2006). MLST makes it possible to differentiate and type bacteria at the sub-species level (Maiden 2006). Usually, 6-7 genes from a group of isolates are partially sequenced. Websites with MLST databases have been set up, for example pubmlst.org (http://pubmlst.org/). Other new computational tools have also been developed for population genetics and recombination detection, such as LDhat (McVean, Awadalla et al. 2002) GARD (Kosakovsky Pond, Posada et al. 2006), START (Jolley, Feil et al. 2001), TOPALi (Milne, Wright et al. 2004). When Goss and co-workers examined *Pseudomonas viridiflava*, a foliar pathogen of *A. thaliana,* they found more recombination within clades than between clades (Goss, Kreitman et al. 2005). Through experimental studies in *Bacillus*, recombination was found to decline as a function of sequence divergence (Majewski 2001). This effect has been found to be associated with the balance of RecA-mediated homologous recombination and MutS-mediated mismatch repair (Fraser, Hanage et al. 2007). Therefore, it is not surprising that Sarkar and Guttman did not find recombination between distantly related *P. syringae* strains (Sarkar and Guttman 2004). Moreover, many of these strains cause disease on different plant species, thus occupying different ecological niche, thereby further limiting the opportunity for recombination. Therefore, when studying population genetics and recombination in a species-complex such as *P. syringae*, it would be more appropriate to examine a

group of strains within this complex that have DNA similarity that is high enough to recombine and that share the same ecological niche, where they have the opportunity to recombine.

Genomic islands are regions of bacterial genomes that can be exchanged *en bloc* between bacteria by site directed recombination (Schubert, Rakin et al. 2004; Pallen and Wren 2007; Kelly, Vespermann et al. 2009). Pathogenicity islands (PAIs) are a particular type of genomic island that harbor virulence genes, such as effector genes. Since site-directed recombination only requires short stretches of sequences that often have two motifs with a partial inverted-repeat symmetry and range from 30 to 200bp in length, genomic islands can be transferred even between distantly related species with low overall sequence similarity. Since PAIs are frequently transferred between pathogenic bacteria, they are believed to be the main source of differences in virulence gene repertoires between strains (Sundin 2007).

Similar to the problem of detecting recombination between distantly related strains, attempts to compare the sequenced genomes of the distantly related strains Pto DC3000 (Buell, Joardar et al. 2003), Psy B728a (Feil, Feil et al. 2005) and Pph 1448A (Joardar, Lindeberg et al. 2005) have not generated many insights into the genes responsible for the different host specificity of these strains. Because many differences in gene content and many allelic differences exist between these strains, it is very challenging to correlate these differences with the different host range of these strains.

There are several factors, ranging from environmental and ecological factors to the biology of hosts and pathogens, that contribute to the spread of disease. Some factors include temperature, humidity, population size of available host and also the virulence of the pathogen. The population size of susceptible hosts is also an important factor in the context of pathogen-host co-evolution, and it is often

determined by host ecology, environmental conditions, and season. One concern in regard to infectious disease emergence is the rapid increase of human population density, which is providing optimal conditions for the evolution of highly virulent human pathogens. Similarly, large modern agricultural fields planted with genetically identical crops at very high density provide optimal conditions for the evolution of highly specialized and highly virulent crop pathogens. Similarly to the use of antibiotics in human medicine, which selects for resistant strains, the application of antibiotics and pesticides/fungicides in agriculture select for resistant plant pathogens. As a result, both animal and plant pathogens are under similar selective pressures. Using *P. syringae* as a model for studying the ecology and evolution of infectious disease is thus a plausible and practical complementation to the study of animal pathogens, and will allow a more comprehensive understanding of the emergence and evolution of pathogens in general.

Ecology of *P. syringae*. Previous studies of *P. syringae* have mainly focused on its interaction with plants in the context of agricultural habitats; strains were isolated from diseased crop plants and non-domesticated plants close to the planting field (Hirano and Upper 2000). This has significantly limited our understanding of the ecology of *P. syringae*. Recent discoveries in microbial ecology found that *P. syringae* and close relatives survive in various environments, including clouds, rivers and snow, suggesting a connection between water cycle and the *P. syringae* life cycle (Christner, Morris et al. 2008; Morris, Sands et al. 2008). However, unlike many animal or human pathogens whose complete life cycles have been well characterized, the life cycles of many bacterial plant pathogens are not completely documented. Also, *P. syringae* in the natural environment has not been systematically surveyed, nor are the number of isolates deposited in strain collections comparable to that of human pathogens, and some depositions are not well documented. There are several reasons for this. First of all,

some symptoms that plant pathogens cause are difficult to distinguish from abiotic damage or stress. Isolation from tissue and culturing are required before identification of bacterial pathogens, because, unlike fungi and oomycetes, they do not produce structures that can be seen by naked eye and used for identification (Goss and Bergelson 2006). Third, it is very difficult to distinguish different *P. syringae* strains from each other using non-molecular methods (Gardan, Shafik et al. 1999).

Ice-nucleation activity. Ice-nucleation was first discovered in the 1970s (Arny, Lindow et al. 1976) and is a very important aspect of *P. syringae* ecology. It is believed that *P. syringae* possesses this activity to inflict frost damage in plant leaves, which should facilitate its entry into intercellular spaces especially in temperate region (Sands, Langhans et al. 1982). It has also been speculated that *P. syringae*, with this activity, can actively participate in snow and rain formation (Christner, Morris et al. 2008), the impact of which can be beneficial to plants and entire ecosystems as part of the collective efforts by different organisms to actively change the environment (Margulis and Sagan 1997).

Biofilms. *P. syringae* can also form biofilms on epilithic surfaces in rivers and other water bodies (Morris, Sands et al. 2008). The formation of biofilms is one of the important mechanisms for single-cell organisms to survive in certain inhospitable environments and is particularly useful for attachment to the leaf surface while waiting for the environment to change to a more favorable condition for invasion and causing disease (Lee, Costerton et al. 2007; Morris, Sands et al. 2008). *P. aeruginosa*, a close relative of *P. syringae*, has been well studied for its formation of biofilm, which contributes to chronic infection and resistance to antibiotics (Pamp, Gjermansen et al. 2008). However, it is yet not quite clear when and where *P. syringae* forms biofilms, and whether biofilm formation contributes to pathogenesis and confers protection from plant defenses.

Distribution. *P. syringae* has been found all around the world in different environments (Morris, Sands et al. 2008; Morris, Sands et al. 2010). *P. syringae* stops growing at temperatures over 30°C (Janse 2005). Thus, damage to crops usually occurs in spring, and *P. syringae* populations stop growing or decline when it is hot and dry (Hirano and Upper 2000).

Transmission. Transmission of pathogens from natural reservoirs or hosts already infected to new hosts is one of the most important aspects of a pathogen's life cycle because successful transmission has to occur before pathogenesis, and because it is an important step in disease prevention (Morris, Sands et al. 2008). It has been suggested that bacterial plant pathogens are poor dispersers (Goss and Bergelson 2006). It is believed that their transmission heavily relies on rain splash, frost damage, leaf wounds and natural openings, such as stomata. Fitness and virulence of pathogens are linked to transmissibility between hosts (Lipsitch and Moxon 1997), but how *P. syringae* is transmitted is still not clearly understood. Poor knowledge in this area is a limiting factor for a comprehensive understanding of *P. syringae* virulence. Recently, *P. syringae* has been found to be able to grow to high titer inside aphids, suggesting a possible vector-born transmission mechanism (Stavrinides, McCloskey et al. 2009). McCarter and colleagues reported that *P. syringae* can survive on tomato seeds (McCarter, Jones et al. 1983), however the importance of seed transmission is still debatable. Being able to quickly swim to stomata or other openings may also play an important role in the transmission ability. Better chemotaxis capabilities may allow faster movement and more accurate guidance towards these openings, which can be assumed to release molecules that are sensed by bacterial chemotaxis receptors. Although it has been shown that *P. syringae* pv. tomato is attracted to organic acids in tomato, unfortunately overall research in this area essentially stagnated for the past 20 years (Cuppels 1988; Soby, Kirkpatrick et al. 1991).

c. Overview of Virulence mechanisms of Pathogens and Plant Defenses

Secretion systems. Protein secretion systems are fundamental virulence mechanisms of both, bacterial animal and plant pathogens, with the type III, type IV, and type VI secretion systems being the most important ones (Galán 2009). For *P. syringae*, currently, only the type III secretion system has been shown to be necessary for virulence (Block, Li et al. 2008). However, although *P. aeruginosa* can cause disease symptoms in a number of plant species in laboratory conditions (Mahajan-Miklos, Rahme et al. 2000; Starkey and Rahme 2009), its type III secretion system is not necessary for *in planta* growth and symptom formation (Miyata, Casey et al. 2003), highlighting how pathogens can use distinctive strategies for causing disease. Nonetheless, similarities between bacterial animal and plant pathogens have indeed exceeded the secretion system mechanism itself with discoveries of similar effectors delivered by both plant and animal pathogens (Hardt and Galán 1997; Mills, Boland et al. 1997; Monack, Mecsas et al. 1997; Mahajan-Miklos, Rahme et al. 2000). There are about 30 verified effectors in the Pto DC3000 genome that are translocated into plant cells (Munkvold, Martin et al. 2008). The functional redundancy of these effectors is well known, since strains with individual effector genes knocked out often exhibit no obvious phenotype (Badel, Nomura et al. 2003). Collectively, these effectors suppress plant defenses and subvert plant metabolisms to create an environment favorable for pathogen growth.

Toxins. In addition to effector proteins, toxins are also commonly utilized by pathogens. Toxins often act by themselves while effectors often act together. Toxins normally have direct, irreversible and dramatic effects on host cellular homeostasis by targeting specific host targets; effectors often have different subtle

effects, which cumulatively alter host metabolism in favor of pathogen growth by modifying various host cellular processes (Galán 2009). One of the better-studied toxins is coronatine produced by a number of *P. syringae* strains including DC3000 (Uppalapati, Ayoubi et al. 2005). It is a molecular mimic of the plant hormone jasmonic acid (JA) and thus down-regulates salicylic acid (SA) signaling in plants since these two pathways are antagonistic (Uppalapati, Ayoubi et al. 2005). Coronatine has been shown to reopen stomata after pathogen-induced closure (Melotto, Underwood et al. 2006). Other toxins produced by *P. syringae* pathovars include phaseolotoxin, syringomycin, syringopeptin and tabtoxin, which have different modes of actions (Bender, Alarcón-Chaidez et al. 1999; Gross and Loper 2009).

Plant defenses. Plants have multi-layered defenses which can successfully fend off most pathogens. In fact, the plant immune system is so effective that diseases are only the exception. The first layer of defenses consists in passive physical barriers such as cuticle and bark. When this first layer of defenses fail, *p*athogen-*a*ssociated *m*olecular *p*attern (PAMP)-*t*riggered *i*mmunity (PTI) (Chisholm, Coaker et al. 2006) and *e*ffector-*t*riggered *i*mmunity (ETI) step in (Guo, Tian et al. 2009).

Flagellin, elongation factor Tu (EF-Tu), lipopolysacchrides (LPS), and chitin are examples of PAMPs (Alfano and Collmer 2004; Chisholm, Coaker et al. 2006; Boller and He 2009). It is believed that plant PTI and animal PTI evolved independently as a result of convergent evolution (Ausubel 2005). However, plants and animals both use pattern recognition receptors (PRRs), which usually have extracellular leucine rich repeats (LRRs) domains and intracellular domains that activate downstream defense signaling pathways (Ausubel 2005).

Also type III effectors are recognized by plants leading to defenses (Alfano and Collmer 2004). The defenses triggered by effectors are collectively called

effector-triggered immunity, or ETI (Jones and Dangl 2006). ETI is often stronger than PTI and often includes a hypersensitive response (HR) (Boller and He 2009). Effector proteins that trigger HR were easier to identify than those with subtle effects. Because HR ultimately terminates the infection, especially for biotrophs (Wroblewski, Caldwell et al. 2009), these effectors were originally called avirulence effectors, hence their naming, for example, *avrA* (Carney and Denny 1990), *avrRpt2* (Innes, Bent et al. 1993), *avrPto1* (Ronald, Salmeron et al. 1992).

Downstream defense responses of both PTI and ETI include oxidative burst, callose deposition, release of phytoalexin, and defense protein accumulation.

The "guard hypothesis". The large number of PRRs in plants (about 140 in *A. thaliana* and over 500 in rice) may be a strategy to compensate for the lack of an adaptive immune system (Ausubel 2005). Moreover, the evolutionary rate of these genes is faster for these defense-related genes compared to other plant genes (Warren, Anandakrishnan et al. 2010). However, the number and variation of these PRRs would still be limited compared to the myriad of different pathogens. The "guard hypothesis" (Dangl and Jones 2001) proposes a solution to this dilemma: plant PRRs may monitor the limited number of host proteins often targeted by pathogens for modification rather than the thousands of effectors (Ausubel 2005). One well-studied example supporting the guard hypothesis is the RPS2-AvrRpt2 gene-for-gene resistance. The resistance protein RPS2 recognizes the AvrRpt2-mediated degradation of RIN4, resulting in an HR (Axtell and Staskawicz 2003; Yao, Eisfelder et al. 2004; Kim, da Cunha et al. 2005). Essentially, RIN4 is "guarded" by RPS2. Because RIN4 in *A. thaliana* is targeted by two additional known effectors, AvrRpm1 and AvrB1, this is an example of how guarding key targets of pathogens by the plant is more feasible than evolving R proteins that directly interact with a large number of effectors. In lettuce, the resistance to three effectors has been mapped to the same locus, suggesting there may be only one R

gene guarding a common target of these effectors (Wroblewski, Caldwell et al. 2009).

Multi-layered interactions. Successful pathogens have to suppress both PTI and ETI, and be able to withstand harsh environmental changes in the apoplast. For example, the type III effector AvrPto1 targets FLS2 to suppress the recognition of flagellin (Xiang, Zong et al. 2010); the fungal effector Ecp6 helps avoiding PTI by competing with the host receptor for chitin (de Jonge, van Esse et al. 2010); AvrB targets the host protein Rar1 (Shang, Li et al. 2006), which is involved in ETI by forming a complex with Sgt1 and heat shock proteins to stabilize R proteins (Azevedo, Sadanandom et al. 2002; Azevedo, Betsuyaku et al. 2006; Botër, Amigues et al. 2007). Pathogens can also mutate PAMPs and effectors to overcome recognition (Gassmann, Dahlbeck et al. 2000; Wichmann, Ritchie et al. 2005; Sun, Dunning et al. 2006). It is now believed that the downstream responses of both PTI and ETI share components, and PTI may act as priming mechanism, and ETI will reinforce the PTI defense responses (Göhre and Robatzek 2008). The AvrRpt2–RIN4–RPS2 interaction is an example of cross talk of PTI and ETI (Kim, da Cunha et al. 2005). The term “arms race” is often used to describe the interactions between host and pathogens (McCann and Guttman 2008). Perhaps the term “cold war” is also appropriate in some ways, since there is no “heat” but continuous pressure to upgrade weaponry.

Essentially, these complex interactions boil down to this: 1) Plant defenses are effective if multiple defense responses at different levels are activated successfully, thus 2) in order to be successful, pathogens have to block or disrupt multiple defense signaling pathways in a timely manner, and prevent the activation of plant defenses, which is necessary ahead of 3) modifying plant metabolism, promote the leakage of nutrition to favor pathogen growth for biotrophs, or 4) injecting deadly toxins to kill plant cells which is a common strategy for

necrotrophs. For successful *P. syringae* strains, and other bacterial plant pathogens such as *Xanthomonas* (Kay and Bonas 2009), the control of plant cells is primarily accomplished by the type III secretion system (T3SS) and type III effectors (T3Es).

d. The T3SS and T3Es

Structure of the machinery. It is believed that the T3SS evolved from the flagellum based on its similar structure and because many genes from these two systems are homologues (McCann and Guttman 2008). It is likely that the ancestor of *P. syringae* acquired a T3SS by horizontal gene transfer (Staskawicz, Dahlbeck et al. 1987). The T3SS acts as needle to deliver T3Es into plant cells and/or intercellular spaces to suppress plant defenses and hijack plant metabolism in favor of pathogen growth. The main structural components include a basal body that is embedded in the bacterial inner and outer cell membranes, and a *hrp* pilus, through which proteins are delivered. Besides the structural conservation of the T3SS, effectors from *P. syringae* can be secreted by *Yersinia* using a heterogeneous T3SS, probably by utilizing conserved targeting sequences (Anderson, Fouts et al. 1999). It was later shown that T3E proteins have a modular structure (Mudgett, Chesnokova et al. 2000; Guttman and Greenberg 2001; Stavrinides, Ma et al. 2006). The amino-terminal regions seem to contain secretion targeting signals, and the carboxy-terminals mostly determine effector function (Cornelis and Van Gijsegem 2000). These modules have been shown to be swappable, which can accelerate evolution of effectors (Guttman, Vinatzer et al. 2002). The challenge for studying T3E is that most of these proteins do not have domains of known function.

The study of effectors. It is relatively more difficult to study T3Es of bacterial plant pathogens compared to bacterial animal pathogens. This lies in the

dramatic quantitative differences in which these effectors are secreted in culture. For example, *Yersinia* effectors are secreted in large quantities in culture and are thus readily detectable (Trülzsch, Sporleder et al. 2004; Davis and Mecsas 2007). In contrast, the T3Es of *P. syringae* are difficult to detect even when very sensitive immunoblots are used (Collmer *et al.*, 2002). As a result, many effectors first identified were those with "avirulence" (avr) activity that cause the easily detectable HR (Greenberg, 1997). The HR is generally only caused by potential pathogens (Collmer et al., 2000). For example, *E. coli* does not cause an HR on tobacco when infiltrated at high concentration, except when heterogenous *P. syringae* effectors from potential pathogenic strains of tobacco are expressed and secreted through a *P. syringae* secretion system expressed from a plasmid (Collmer et al., 2000). This suggests that the HR is only triggered by potential pathogens.

T3E repertoires of different *P. syringae* strains are very different from each other (Sarkar, Gordon et al. 2006). The differences in the T3E repertoire are believed to explain many of the host range differences between *P. syringae* strains (Sarkar, Gordon et al. 2006). Contribution of individual effectors to host range is difficult to discern as dozens of other effectors create a very "noisy" background. Nonetheless, a few avirulence effectors restrict host range by eliciting strong ETI. For example, removing *hopQ1* from DC3000 lifts the restriction of its growth in *N. benthamiana* (Wei, Kvitko et al. 2007).

Several groups have attempted to develop large-scale assays for studying effectors. Rohmer and colleagues examined 44 known or candidate T3E genes from a large number of strains to look at phylogenetics, evidence for horizontal gene transfer, and diversifying selection on these effector genes (Rohmer, Guttman et al. 2004). DNA-microarray analysis of 91 *P. syringae* strains identified several sets of candidate genes, such as T3E genes, possibly important for host range determination based on their association with host of isolation (Sarkar, Gordon et

al. 2006). However, Sarkar and colleagues used 53 validated and putative DC3000 effectors as the basis for the microarray design (Sarkar, Gordon et al. 2006), as a result the analysis was extremely biased and did not consider the effectors absent from DC3000. Either a microarray design based on many more effectors genes from a larger number of *P. syringae* strains is needed, or whole genome sequencing is needed for identification of all effector differences.

In another study, Tadeusz and colleagues cloned 171 effector genes from several pathovars of *P. syringae* and *Ralstonia* and expressed them in different crop species one by one using Agrobacterium-mediated transient assays in the absence of any other effectors, which eliminated the type of "noise" mentioned above. They tested the ability of effectors to induce necrosis or chlorosis as an indication of an HR, aiming at dissecting the complex co-evolution between different hosts and effectors (Wroblewski, Caldwell et al. 2009). However, due to the nature of transient expression *in planta*, the large quantity of the effector proteins expressed may obscure their true function in plant cells. Nonetheless, this study was a powerful one that discovered effectors that potentially trigger an HR in a number of plant species. Another study used *P. fluorescens* expressing a *P. syringae* T3SS from a plasmid and individual effectors of strain DC3000. In this study, most DC3000 effectors were found to be able to suppress the HR triggered by the effector HopA1(Guo, Tian et al. 2009). However, these results are controversial, for example, they showed that both HopM1 and HopT1 are able to completely suppress HopA1-elicited HR in tobacco although it has been shown that these two effectors elicit cell death in tobacco and *N. benthamiana* respectively (Vinatzer, Teitzel et al. 2006; Wroblewski, Caldwell et al. 2009). Nonetheless, numerous studies have shown that many individual effectors can suppress plant defenses, suggesting the main role of T3Es is to suppress plant defenses.

Despite of these large-scale studies, to date, most new discoveries about effector functions are still done by more traditional molecular biology and cell biology methods. Since the cloning of the first avirulent effector gene *avrA* from *P. syringae* pv. *glycinea* (Staskawicz, Dahlbeck et al. 1984), virulence functions of many effectors have been uncovered (Mudgett 2005; Göhre and Robatzek 2008; Cunnac, Lindeberg et al. 2009; Galán 2009). Here I just list a few examples. Several effector proteins are proteases that directly degrade host target proteins. For example, AvrRpt2, AvrRxv, AvrBsT, and AvrXv4 are cysteine proteases and YopJ-like SUMO proteases (Rohmer, Guttman et al. 2004; Göhre and Robatzek 2008). Indeed, these cysteine proteases have been shown to target host protein for cleavage or degradation (Hotson and Mudgett 2004; Kim, Desveaux et al. 2005). AvrBs3, on the other hand, is one example of effectors that function as transcription factors. The transcription of the resistance gene *Bs3* in pepper is activated by this effector resulting in resistance (Kay and Bonas 2009). These transcription activator like effectors (TAL) from *Xanthomonas* are of high interest to biotechnology because they can be engineered to bind to specific DNA sequences, which allows for engineering of resistance genes specifically activated by these effectors (Bogdanove, Schornack et al. 2010).

e. Determination of Host Range, and Evolution of Host Range and Virulence

Host range. Host range limited to a single host could be a risky strategy for any plant pathogen in natural mixed plant communities. However, one could imagine that the situation changes once a pathogen adapts to the agricultural monoculture (Stukenbrock and McDonald 2008). A similar evolutionary history has been found for some human pathogens, which had broad host ranges before

humans started to live at high density in larger groups, but then evolved into highly virulent, exclusively human, pathogens. An example is *Bordetella pertussis*, which causes the deadly "wooping cough" disease only in humans but evolved from its ancestor, which probably lived almost asymptomatically in the upper respiratory tract of a wide range of animals as the *B pertussis* relative *B. bronchiseptica* (Bjornstad and Harvill 2005). It is reasonable to hypothesize that, under the selection of large-scale monoculture, a bacterial pathogen may similarly restrict its host range and increase its virulence accordingly. However, this hypothesis has not been tested yet.

Mechanisms for host range determination. Avirulence effectors, such as HopQ1, AvrRpt2, AvrPto1 and AvrPtoB, have been shown to restrict host range due to recognition by many plant species (Ronald, Salmeron et al. 1992; Axtell and Staskawicz 2003; Wei, Kvitko et al. 2007). Thus gene-for-gene resistance is a key mechanism for host range determination. Additionally, more non-host plant species react to transiently expressed effectors than host plants, suggesting that non-host resistance could be largely determined by the recognition of multiple avirulence effectors (Wroblewski, Caldwell et al. 2009).

Host range evolution. There are a few challenges for studying host range evolution. First, it is difficult to determine host range by experiments in lab conditions compared to field conditions. Second, a precise phylogenetic tree that can be used for interference is difficult to build. Third, the mutation rate of *P. syringae* in the field is unknown. Fourth, to date, it is not fully understood how host range is determined besides that it is often determined by multiple factors from both the host and the pathogen. As a result, only certain questions can be asked and tested experimentally to study host range evolution of *P. syringae*.

Origin of pathogens. Understanding the origin of pathogens first can help us understand the evolution of pathogens. For many proteobacterial pathogens, the

T3SS is essential for causing disease, so the origin of the T3SS might have been seminal for these sophisticated pathogens. The fact that gene clusters that encode the T3SS are often located in PAIs suggest that T3SS are acquired by horizontal gene transfer (McCann and Guttman 2008). Thus the origin of pathogenic strains may start with the acquisition of a T3SS. The next step would be to acquire virulence genes. It is now believed that many microbes can be naturally competent under certain environmental conditions. Consequently, in a mixed bacterial community in the environment, emerging pathogens may relatively easily acquire new DNA and integrate it into their genomes by site-specific or homologous recombination (Arnold, Jackson et al. 2007; Lovell, Mansfield et al. 2009).

Gain and loss of host range. From the perspective of bacterial pathogens, gain of host range is likely dependent on gain of a number of virulence factors and evasion of plant immune responses through mutation of PAMPs or loss of effectors with avirulence activity. One example is that, when DC3000 loses *hopQ1*, it becomes compatible with *N. benthamiana* (Wei, Kvitko et al. 2007). Gain of a new host may be accompanied by loss of other hosts since gain of additional effectors that may be virulence factors on the new host may have avirulence activity on other one.

The plant hosts are also constantly updating their defense mechanisms. Mutation in, and gain of, receptors for PAMPs, duplication and gain of function of *R* genes, can all confer resistance to pathogens thus restrict host range of some pathogen strains.For example, by human intervention, *Pto* resistance gene bred into tomato cultivars confers resistance to race 0 *P. syringae* pv tomato strains (Ronald, Salmeron et al. 1992; Kunkeaw, Tan et al. 2010). Adding PAMP receptor gene to plants can also increase resistance (Lacombe, Rougon-Cardoso et al. 2010). Different cultivars or ecotypes often exhibit different level of resistance to the same pathogen, suggesting rapid evolution of defense at cultivar or ecotype level.

Mechanisms. It is now generally believed that horizontal gene transfer (HGT) played an important role in the evolution of T3E repertoires and thus in the evolution of host range (Rohmer, Guttman et al. 2004; Sarkar, Gordon et al. 2006). Evidence includes: the association of effector genes with PAIs (Lindeberg, Myers et al. 2008); highly similar effectors distributed across distantly related pathovars that have phylogenies incongruent with the rest of the genome (Rohmer, Guttman et al. 2004; Sarkar, Gordon et al. 2006); *P. syringae* T3E that are shared with other species (Rohmer, Guttman et al. 2004).

Some mechanisms of HGT have been examined in both plant and animal pathogen systems (Basim et al., 1999, Dogra et al., 2004, Daane et al., 1996, Patil & Sonti, 2004). Evidence for HGT of effectors has also been found in sequenced *P. syringae* genomes based on GC content, nearby mobile elements, and phylogenetic incongruencies (Rohmer, Guttman et al. 2004; Sarkar, Gordon et al. 2006).

Bacteriophages represent one kind of mobile element that contribute to the dynamics of bacteria genomes (Kim, Charkowski et al. 1998; Pallen and Wren 2007). The insertion of phages into genomes has been shown to confer virulence or expansion of host range (O'Brien, Newland et al. 1984; Karaolis, Somara et al. 1999; Miao and Miller 1999). The excision of phages from a genome, or homologous recombination events that are facilitated by the presence of phage remnant sequences, allow the loss of T3Es in the phage gene cluster, or exchange and introduction of T3Es. These are classical examples of bacterial pathogen genome dynamics. Many PAIs are associated with transposable elements. One example of an intensively studied PAI is PPHGI-1 of *P. syringae* pv. *phaseolicola*, which can be excised, integrated, or lost during plant infection (Pitman, Jackson et al. 2005; Lovell, Mansfield et al. 2009).

One additional mechanism for pathogens to modulate virulence and host range is called pathoadaptation (Stavrinides, Ma et al. 2006). It consists in swapping domains between effector genes and other genes or only effector genes and leads to new effectors with possible new virulence or avirulence activities.

Loss of T3Es that trigger strong ETI is one of the mechanisms that expands host range. Mechanisms for losing T3Es that trigger ETI include transposon insertion and single nucleotide mutation that result in early termination, and excision of large PAIs (Arnold, Jackson et al. 2007). It has been proposed that stress resulted from plant defenses can accelerate genome rearrangement in bacterial pathogens (Arnold, Jackson et al. 2007) and has been demonstrated (Pitman, Jackson et al. 2005; Lovell, Mansfield et al. 2009). Although it has not been proven, it is entirely possible that some mobile elements can be under control of certain regulators that are activated upon perception of stress, similar to the regulation of the T3SS. As a result, when a bacterial strain acquires an effector that elicits stronger ETI, or when the host has evolved a resistance gene for this effector, this strain may mutate or completely lose the gene that encodes the effector protein. This could be one key mechanism for the pathogen to quickly evade newly introduced ETI in the field.

Influence of plant defenses. Plant defenses can induce higher HGT efficiency in *P. syringae*, including the incorporation of DNA fragments, for instance, PAIs that contain effectors beneficial to the pathogens (Lovell, Mansfield et al. 2009).

Introduction of R genes into cultivated crops also puts selective pressure on the evolution T3E repertoires. *P. syringae* pv. tomato race 1 strains, which have overcome the *Pto* resistance gene introduced into commercial tomato cultivars, are now the dominating isolates in the field in California (Kunkeaw, Tan et al. 2010). Similarly, the *Bs2* resistance gene introduced into pepper varieties has also

significantly influenced the evolution of *avrBs2* alleles of *Xanthomonas axonopodis* pv. *vesicatoria* strains in the field (Gassmann, Dahlbeck et al. 2000; Wichmann, Ritchie et al. 2005). These cases demonstrate the fast adaptability of bacterial plant pathogens. The need for the engineering of crops that have more durable resistance is pressing.

Invasion. The ability to invade plant tissue is necessary for pathogenesis of many plant pathogens. *P. syringae* and *Xanthomonas* have to pass through wounds or stomata on the leaf surface to enter the intercellular space before they can cause disease. *Ralstonia solanacearum* has to invade the plant root tissue and eventually enter xylem before it can spread systematically and cause wilting (VASSE, FREY et al. 1995). Inability to invade plant tissue will result in incompatibility. Because the leaf surface is a tough environment due to limited resources and exposure to UV (Beattie and Lindow 1995; Hirano and Upper 2000), the faster some strains can enter the intercellular space, the better their chance for survival, causing disease and passing on the genes to next generations. Chemotaxis is one key mechanism that controls directional movement of flagellum-equipped bacterial plant pathogens. *P. syringae* has been shown to reopen stomata, swim towards and gather around stomata, supporting the notion that stomata are important sites for entry (Melotto, Underwood et al. 2006). Compounds released from stomata, such as organic acids, may guide pathogens towards these openings by chemotaxis (Cuppels 1988). While it has been shown that chemotaxis is important for *Ralstonia* virulence and for zoospores of *Phytophthora* species (Tyler 2002; Yao and Allen 2006), less is known in this aspect for *P. syringae*. For fungal pathogens, the abilities to penetrate the leaf cuticle or germination of spores towards stomata are required for host compatibility (Tucker and Talbot 2001; Van Damme, Andel et al. 2005). *A. thaliana PENETRATION* mutants, such as *PEN1*, *PEN2* and *PEN3* have been shown to be compromised in resistance against

invasion by non-host powdery mildew *Blumeria graminis* f.sp. *hordei* and *Erysiphe pisi* (Collins, Thordal-Christensen et al. 2003; Lipka, Dittgen et al. 2005; Stein, Dittgen et al. 2006). These findings highlight the importance of invasion for host range.

f. Overview of research

Comparative studies of *P. syringae* strains which are more closely related to each other than the three complete sequenced strains have been rare. I believe such research is essential for understanding host range evolution and determination. For this dissertation, I had three main objectives:

Characterization of a group of closely related strains with different host ranges and virulence.

Identification of genomic difference between closely related strains characterized above.

Further investigation of the roles some specific genomic differences play during host range determination.

g. References

Alfano, J. R. and A. Collmer (2004). "Type III secretion system effector proteins: double agents in bacterial disease and plant defense." Annu Rev Phytopathol **42**: 385-414.

Anderson, D. M., D. E. Fouts, et al. (1999). "Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals." Proc Natl Acad Sci U S A **96**(22): 12839-12843.

Arnold, D. L., R. W. Jackson, et al. (2007). "Evolution of microbial virulence: the benefits of stress." Trends Genet. **23**(6): 293-300.

Arny, D. C., S. E. Lindow, et al. (1976). "Frost sensitivity of Zea mays increased by application of *Pseudomonas syringae*." Nature **262**(5566): 282-284.

Ausubel, F. M. (2005). "Are innate immune signaling pathways in plants and animals conserved?" Nat Immunol **6**(10): 973-979.

Axtell, M. J. and B. J. Staskawicz (2003). "Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4." Cell **112**(3): 369-377.

Azevedo, C., S. Betsuyaku, et al. (2006). "Role of SGT1 in resistance protein accumulation in plant immunity." EMBO J **25**(9): 2007-2016.

Azevedo, C., A. Sadanandom, et al. (2002). "The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance." Science **295**(5562): 2073-2076.

Badel, J. L., K. Nomura, et al. (2003). "*Pseudomonas syringae* pv. *tomato* DC3000 HopPtoM (CEL ORF3) is important for lesion formation but not growth in tomato and is secreted and translocated by the Hrp type III secretion system in a chaperone-dependent manner." Mol Microbiol **49**(5): 1239-1251.

Barenblatt, D. (2004). A plague upon humanity : the secret genocide of Axis Japan's germ warfare operation. New York, HarperCollins.

Beattie, G. A. and S. E. Lindow (1995). "The secret life of foliar bacterial pathogens on leaves." Annu Rev Phytopathol **33**: 145-172.

Bender, C. L., F. Alarcón-Chaidez, et al. (1999). "*Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases." Microbiol Mol Biol Rev **63**(2): 266-292.

Bjornstad, O. N. and E. T. Harvill (2005). "Evolution and emergence of *Bordetella* in humans." Trends Microbiol **13**(8): 355-359.

Block, A., G. Li, et al. (2008). "Phytopathogen type III effector weaponry and their plant targets." Curr. Opin. Plant Biol. **11**(4): 396-403.

Bogdanove, A. J., S. Schornack, et al. (2010). "TAL effectors: finding plant genes for disease and defense." Curr Opin Plant Biol **13**(4): 394-401.

Boller, T. and S. Y. He (2009). "Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens." Science **324**(5928): 742-744.

Botër, M., B. Amigues, et al. (2007). "Structural and functional analysis of SGT1 reveals that its interaction with HSP90 is required for the accumulation of Rx, an R protein involved in plant immunity." Plant Cell **19**(11): 3791-3804.

Buell, C. R., V. Joardar, et al. (2003). "The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000." Proc Natl Acad Sci U S A **100**(18): 10181-10186.

Carney, B. F. and T. P. Denny (1990). "A cloned avirulence gene from *Pseudomonas solanacearum* determines incompatibility on *Nicotiana tabacum* at the host species level." Journal of bacteriology **172**(9): 4836-4843.

Chisholm, S. T., G. Coaker, et al. (2006). "Host-microbe interactions: shaping the evolution of the plant immune response." Cell **124**(4): 803-814.

Christner, B. C., C. E. Morris, et al. (2008). "Ubiquity of biological ice nucleators in snowfall." Science **319**(5867): 1214.

Collins, N. C., H. Thordal-Christensen, et al. (2003). "SNARE-protein-mediated disease resistance at the plant cell wall." Nature **425**(6961): 973-977.

Cornelis, G. R. and F. Van Gijsegem (2000). "Assembly and function of type III secretory systems." Annu Rev Microbiol **54**: 735-774.

Council, U. S. N. I. (2002). "National intelligence estimate: the global infectious disease threat and its implications for the United States." Environ Change Secur Proj Rep(6): 33-65.

Cunnac, S., M. Lindeberg, et al. (2009). "*Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions." Curr Opin Microbiol **12**(1): 53-60.

Cuppels, D. A. (1988). "Chemotaxis by *Pseudomonas syringae* pv. *tomato*." Appl Environ Microbiol **54**(3): 629-632.

Dangl, J. L. and J. D. Jones (2001). "Plant pathogens and integrated defence responses to infection." Nature **411**(6839): 826-833.

Davis, A. J. and J. Mecsas (2007). "Mutations in the *Yersinia pseudotuberculosis* type III secretion system needle protein, YscF, that specifically abrogate effector translocation into host cells." J Bacteriol **189**(1): 83-97.

de Jonge, R., H. P. van Esse, et al. (2010). "Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants." Science **329**(5994): 953-955.

Denny, T. P., M. N. Gilmour, et al. (1988). "Genetic diversity and relationships of two pathovars of *Pseudomonas syringae*." J Gen Microbiol **134**(7): 1949-1960.

Drancourt, M. and D. Raoult (2002). "Molecular insights into the history of plague." Microbes Infect **4**(1): 105-109.

Duan, Y., L. Zhou, et al. (2009). "Complete genome sequence of citrus huanglongbing bacterium, 'Candidatus Liberibacter asiaticus' obtained through metagenomics." Mol Plant Microbe Interact **22**(8): 1011-1020.

Dye, D. W., J. F. Bradbury, et al. (1980). "International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains." Rev. Plant Pathol. **59**: 153-168.

Feil, H., W. S. Feil, et al. (2005). "Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000." Proc Natl Acad Sci U S A **102**(31): 11064-11069.

Ferber, D. (2010). "Infectious disease. From pigs to people: the emergence of a new superbug." Science **329**(5995): 1010-1011.

Fraser, C., W. P. Hanage, et al. (2007). "Recombination and the nature of bacterial speciation." Science **315**(5811): 476-480.

Friedman, S. R., S. C. Kippax, et al. (2006). "Emerging future issues in HIV/AIDS social research." AIDS **20**(7): 959-965.

Galán, J. E. (2009). "Common themes in the design and function of bacterial effectors." Cell Host Microbe **5**(6): 571-579.

Gardan, L., H. Shafik, et al. (1999). "DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959)." Int J Syst Bacteriol **49 Pt 2**: 469-478.

Gassmann, W., D. Dahlbeck, et al. (2000). "Molecular evolution of virulence in natural field strains of *Xanthomonas campestris* pv. *vesicatoria*." J Bacteriol **182**(24): 7053-7059.

Göhre, V. and S. Robatzek (2008). "Breaking the barriers: microbial effector molecules subvert plant immunity." Annu Rev Phytopathol **46**: 189-215.

Gomez-Gomez, L. and T. Boller (2000). "FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis." Mol Cell **5**(6): 1003-1011.

Goss, E. M. and J. Bergelson (2006). "Variation in resistance and virulence in the interaction between *Arabidopsis thaliana* and a bacterial pathogen." Evolution Int J Org Evolution **60**(8): 1562-1573.

Goss, E. M., M. Kreitman, et al. (2005). "Genetic diversity, recombination and cryptic clades in *Pseudomonas viridiflava* infecting natural populations of *Arabidopsis thaliana*." Genetics **169**(1): 21-35.

Greenberg, J. T. and B. A. Vinatzer (2003). "Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells." Curr Opin Microbiol **6**(1): 20-28.

Gross, H. and J. E. Loper (2009). "Genomics of secondary metabolite production by *Pseudomonas* spp." Nat Prod Rep **26**(11): 1408-1446.

Guo, M., F. Tian, et al. (2009). "The majority of the type III effector inventory of *Pseudomonas syringae* pv. *tomato* DC3000 can suppress plant immunity." Mol Plant Microbe Interact **22**(9): 1069-1080.

Guttman, D. S. and J. T. Greenberg (2001). "Functional analysis of the type III effectors AvrRpt2 and AvrRpm1 of *Pseudomonas syringae* with the use of a single-copy genomic integration system." Mol Plant Microbe Interact **14**(2): 145-155.

Guttman, D. S., S. J. Gropp, et al. (2006). "Diversifying Selection Drives the Evolution of the Type III Secretion System Pilus of *Pseudomonas syringae*." Mol Biol Evol **23**(12): 2342-2354.

Guttman, D. S., B. A. Vinatzer, et al. (2002). "A functional screen for the type III (Hrp) secretome of the plant pathogen Pseudomonas syringae." Science **295**(5560): 1722-1726.

Hardt, W. D. and J. E. Galán (1997). "A secreted Salmonella protein with homology to an avirulence determinant of plant pathogenic bacteria." Proc Natl Acad Sci USA **94**(18): 9887-9892.

Hayashi, F., K. D. Smith, et al. (2001). "The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5." Nature **410**(6832): 1099-1103.

Hirano, S. S. and C. D. Upper (2000). "Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*-a pathogen, ice nucleus, and epiphyte." Microbiol Mol Biol Rev **64**(3): 624-653.

Hotson, A. and M. B. Mudgett (2004). "Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity." Curr. Opin. Plant Biol. **7**(4): 384-390.

Hsueh, P. R., L. J. Teng, et al. (1998). "Outbreak of *Pseudomonas fluorescens* bacteremia among oncology patients." J Clin Microbiol **36**(10): 2914-2917.

Hwang, M. S. H., R. L. Morgan, et al. (2005). "Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*." Appl Environ Microbiol **71**(9): 5182-5191.

Innes, R. W., A. F. Bent, et al. (1993). "Molecular analysis of avirulence gene *avrRpt2* and identification of a putative regulatory sequence common to all known *Pseudomonas syringae* avirulence genes." J Bacteriol **175**(15): 4859-4869.

Ippolito, G., S. Leone, et al. (2010). "Methicillin-resistant *Staphylococcus aureus*: the superbug." International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases **14S4**: S7-S11.

Janjusevic, R., R. B. Abramovitch, et al. (2006). "A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase." Science **311**(5758): 222-226.

Janse, J. D. (2005). Phytobacteriology : principles and practice. Wallingford, UK ; Cambridge, MA, CABI Pub.

Joardar, V., M. Lindeberg, et al. (2005). "Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among

pathovars in genes involved in virulence and transposition." J Bacteriol **187**(18): 6488-6498.

Jolley, K. A., E. J. Feil, et al. (2001). "Sequence type analysis and recombinational tests (START)." Bioinformatics **17**(12): 1230-1231.

Jones, J. D. and J. L. Dangl (2006). "The plant immune system." Nature **444**(7117): 323-329.

Karaolis, D. K., S. Somara, et al. (1999). "A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria." Nature **399**(6734): 375-379.

Kay, S. and U. Bonas (2009). "How *Xanthomonas* type III effectors manipulate the host plant." Curr Opin Microbiol **12**(1): 37-43.

Kelly, B. G., A. Vespermann, et al. (2009). "The role of horizontal gene transfer in the evolution of selected foodborne bacterial pathogens." Food Chem Toxicol **47**(5): 951-968.

Kim, H.-S., D. Desveaux, et al. (2005). "The *Pseudomonas syringae* effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from Arabidopsis membranes to block RPM1 activation." Proc Natl Acad Sci USA **102**(18): 6496-6501.

Kim, J., A. Charkowski, et al. (1998). "Sequences Related to Transposable Elements and Bacteriophages Flank Avirulence Genes of *Pseudomonas syringae*." Molecular Plant-Microbe Interactions **11**(12): 1247-1252.

Kim, M. G., L. da Cunha, et al. (2005). "Two *Pseudomonas syringae* type III effectors inhibit RIN4-regulated basal defense in Arabidopsis." Cell **121**(5): 749-759.

Kosakovsky Pond, S. L., D. Posada, et al. (2006). "Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm." Mol Biol Evol **23**(10): 1891-1901.

Kunkeaw, S., S. Tan, et al. (2010). "Molecular and evolutionary analyses of *Pseudomonas syringae* pv. *tomato* race 1." Mol Plant Microbe Interact **23**(4): 415-424.

Lacombe, S., A. Rougon-Cardoso, et al. (2010). "Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance." Nature biotechnology.

Lee, K., J. W. Costerton, et al. (2007). "Phenotypic and functional characterization of Bacillus anthracis biofilms." Microbiology **153**(Pt 6): 1693-1701.

Lindeberg, M., C. R. Myers, et al. (2008). "Roadmap to New Virulence Determinants in *Pseudomonas syringae*: Insights from Comparative Genomics and Genome Organization." Molecular Plant-Microbe Interactions.

Lipka, V., J. Dittgen, et al. (2005). "Pre- and postinvasion defenses both contribute to nonhost resistance in Arabidopsis." Science **310**(5751): 1180-1183.

Lipsitch, M. and E. R. Moxon (1997). "Virulence and transmissibility of pathogens: what is the relationship?" Trends Microbiol **5**(1): 31-37.

Lovell, H. C., J. W. Mansfield, et al. (2009). "Bacterial evolution by genomic island transfer occurs via DNA transformation in planta." Curr Biol **19**(18): 1586-1590.

Mahajan-Miklos, S., L. G. Rahme, et al. (2000). "Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts." Mol Microbiol **37**(5): 981-988.

Maiden, M. C. (2006). "Multilocus Sequence Typing of Bacteria." Annu Rev Microbiol.

Maiden, M. C., J. A. Bygraves, et al. (1998). "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." Proc Natl Acad Sci U S A **95**(6): 3140-3145.

Majewski, J. (2001). "Sexual isolation in bacteria." FEMS Microbiol Lett **199**(2): 161-169.

Margulis, L. and D. Sagan (1997). Slanted truths : essays on Gaia, symbiosis, and evolution. New York, Copernicus.

Martin, G. B., S. H. Brommonschenkel, et al. (1993). "Map-based cloning of a protein kinase gene conferring disease resistance in tomato." Science **262**(5138): 1432-1436.

Matz, C., A. M. Moreno, et al. (2008). "*Pseudomonas aeruginosa* uses type III secretion system to kill biofilm-associated amoebae." ISME J **2**(8): 843-852.

McCann, H. C. and D. S. Guttman (2008). "Evolution of the type III secretion system and its effectors in plant-microbe interactions." New Phytol **177**(1): 33-47.

McCarter, S. M., J. B. Jones, et al. (1983). "Survival of *Pseudomonas syringae* pv. *tomato* in Association with Tomato Seeds, Soil, Host Tissue, and Epiphytic Weed Hosts in Georgia." Phytopathology **73**(10): 1393-1398.

McVean, G., P. Awadalla, et al. (2002). "A coalescent-based method for detecting and estimating recombination from gene sequences." Genetics **160**(3): 1231-1241.

Melotto, M., W. Underwood, et al. (2006). "Plant stomata function in innate immunity against bacterial invasion." Cell **126**(5): 969-980.

Miao, E. A. and S. I. Miller (1999). "Bacteriophages in the evolution of pathogen-host interactions." Proc Natl Acad Sci USA **96**(17): 9452-9454.

Mills, S. D., A. Boland, et al. (1997). "*Yersinia enterocolitica* induces apoptosis in macrophages by a process requiring functional type III secretion and translocation mechanisms and involving YopP, presumably acting as an effector protein." Proc Natl Acad Sci USA **94**(23): 12638-12643.

Milne, I., F. Wright, et al. (2004). "TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments." Bioinformatics **20**(11): 1806-1807.

Miyata, S., M. Casey, et al. (2003). "Use of the *Galleria mellonella* Caterpillar as a Model Host To Study the Role of the Type III Secretion System in *Pseudomonas aeruginosa* Pathogenesis." Infection and immunity **71**(5): 2404.

Monack, D. M., J. Mecsas, et al. (1997). "*Yersinia* signals macrophages to undergo apoptosis and YopJ is necessary for this cell death." Proc Natl Acad Sci USA **94**(19): 10385-10390.

Morelli, G., Y. Song, et al. (2010). "*Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity." Nature genetics.

Morris, C. E., D. C. Sands, et al. (2010). "Inferring the Evolutionary History of the Plant Pathogen *Pseudomonas syringae* from Its Biogeography in Headwaters of Rivers in North America, Europe, and New Zealand." MBio **1**(3).

Morris, C. E., D. C. Sands, et al. (2008). "The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle." ISME J.

Mudgett, M. B. (2005). "New insights to the function of phytopathogenic bacterial type III effectors in plants." Annu Rev Plant Biol **56**: 509-531.

Mudgett, M. B., O. Chesnokova, et al. (2000). "Molecular signals required for type III secretion and translocation of the *Xanthomonas campestris* AvrBs2 protein to pepper plants." Proc Natl Acad Sci U S A **97**(24): 13324-13329.

Munkvold, K. R., M. E. Martin, et al. (2008). "A survey of the *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion system effector repertoire reveals several effectors that are deleterious when expressed in *Saccharomyces cerevisiae*." Mol Plant Microbe Interact **21**(4): 490-502.

Newton, A. C., B. D. L. Fitt, et al. (2010). "Pathogenesis, parasitism and mutualism in the trophic space of microbe-plant interactions." Trends Microbiol **18**(8): 365-373.

O'Brien, A. D., J. W. Newland, et al. (1984). "Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea." Science **226**(4675): 694-696.

Pallen, M. J. and B. W. Wren (2007). "Bacterial pathogenomics." Nature **449**(7164): 835-842.

Pamp, S. J., M. Gjermansen, et al. (2008). "Tolerance to the antimicrobial peptide colistin in *Pseudomonas aeruginosa* biofilms is linked to metabolically active cells, and depends on the pmr and mexAB-oprM genes." Mol Microbiol **68**(1): 223-240.

Patterson, K. B. and T. Runge (2002). "Smallpox and the Native American." Am J Med Sci **323**(4): 216-222.

Pitman, A. R., R. W. Jackson, et al. (2005). "Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants." Curr Biol **15**(24): 2230-2235.

Rahme, L. G., E. J. Stevens, et al. (1995). "Common virulence factors for bacterial pathogenicity in plants and animals." Science **268**(5219): 1899-1902.

Rizzo, D. M., M. Garbelotto, et al. (2005). "Phytophthora ramorum: integrative research and management of an emerging pathogen in California and Oregon forests." Annu Rev Phytopathol **43**: 309-335.

Rohmer, L., D. S. Guttman, et al. (2004). "Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*." Genetics **167**(3): 1341-1360.

Ronald, P. C., J. M. Salmeron, et al. (1992). "The cloned avirulence gene *avrPto* induces disease resistance in tomato cultivars containing the *Pto* resistance gene." J Bacteriol **174**(5): 1604-1611.

Rosebrock, T. R., L. Zeng, et al. (2007). "A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity." Nature **448**(7151): 370-374.

Sands, D. C., V. E. Langhans, et al. (1982). "The association between bacteria and rain and possible resultant meteorological implications." Journal of the Hungarian Meteorological Service **86**(2-4): 5.

Sarkar, S. F., J. S. Gordon, et al. (2006). "Comparative genomics of host-specific virulence in *Pseudomonas syringae*." Genetics **174**(2): 1041-1056.

Sarkar, S. F. and D. S. Guttman (2004). "Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen." Appl Environ Microbiol **70**(4): 1999-2012.

Schubert, S., A. Rakin, et al. (2004). "The *Yersinia* high-pathogenicity island (HPI): evolutionary and functional aspects." Int J Med Microbiol **294**(2-3): 83-94.

Shang, Y., X. Li, et al. (2006). "RAR1, a central player in plant immunity, is targeted by *Pseudomonas syringae* effector AvrB." Proc Natl Acad Sci USA **103**(50): 19200-19205.

Soby, S., B. Kirkpatrick, et al. (1991). "Chemotaxis of *Pseudomonas syringae* subsp. *savastanoi* Virulence Mutants." Appl Environ Microbiol **57**(10): 2918-2920.

Starkey, M. and L. G. Rahme (2009). "Modeling *Pseudomonas aeruginosa* pathogenesis in plant hosts." Nat Protoc **4**(2): 117-124.

Staskawicz, B., D. Dahlbeck, et al. (1987). "Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*." J. Bacteriol. **169**(12): 5789-5794.

Staskawicz, B. J., D. Dahlbeck, et al. (1984). "Cloned avirulence gene of *Pseudomonas syringae* pv. *glycinea* determines race-specific incompatibility on Glycine max (L.) Merr." Proc Natl Acad Sci USA **81**(19): 6024-6028.

Stavrinides, J., W. Ma, et al. (2006). "Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens." PLoS Pathog **2**(10): e104.

Stavrinides, J., J. K. McCloskey, et al. (2009). "Pea aphid as both host and vector for the phytopathogenic bacterium *Pseudomonas syringae*." Appl Environ Microbiol **75**(7): 2230-2235.

Stein, M., J. Dittgen, et al. (2006). "Arabidopsis PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration." Plant Cell **18**(3): 731-746.

Stukenbrock, E. H. and B. A. McDonald (2008). "The origins of plant pathogens in agro-ecosystems." Annu Rev Phytopathol **46**: 75-100.

Sun, W., F. M. Dunning, et al. (2006). "Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses." Plant Cell **18**(3): 764-779.

Sundin, G. W. (2007). "Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts." Annu Rev Phytopathol **45**: 129-151.

Trülzsch, K., T. Sporleder, et al. (2004). "Contribution of the major secreted yops of *Yersinia enterocolitica* O:8 to pathogenicity in the mouse infection model." Infection and immunity **72**(9): 5227-5234.

Tucker, S. L. and N. J. Talbot (2001). "Surface attachment and pre-penetration stage development by plant pathogenic fungi." Annu Rev Phytopathol **39**: 385-417.

Tyler, B. M. (2002). "Molecular basis of recognition between *phytophthora* pathogens and their hosts." Annu Rev Phytopathol **40**: 137-167.

Tyler, B. M., S. Tripathy, et al. (2006). "*Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis." Science **313**(5791): 1261-1266.

Uppalapati, S. R., P. Ayoubi, et al. (2005). "The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato." Plant J **42**(2): 201-217.

USDA. "Agricultural Select Agent." from http://www.aphis.usda.gov/programs/ag_selectagent/ag_bioterr_toxinlist.shtml.

Van Damme, M., A. Andel, et al. (2005). "Identification of arabidopsis loci required for susceptibility to the downy mildew pathogen *Hyaloperonospora parasitica*." Mol Plant Microbe Interact **18**(6): 583-592.

VASSE, J., P. FREY, et al. (1995). "MICROSCOPIC STUDIES OF INTERCELLULAR INFECTION AND PROTOXYLEM INVASION OF TOMATO ROOTS BY *PSEUDOMONAS-SOLANACEARUM*." Mol Plant Microbe In **8**(2): 241-251.

Vinatzer, B. A. and C. T. Bull (2009). The impact of genomic approaches on our understanding of diversity and taxonomy of plant pathogenic bacteria. Plant Pathogenic Bacteria: Genomics and Molecular Biology. R. W. Jackson, Caister Academic Press.

Vinatzer, B. A., G. M. Teitzel, et al. (2006). "The type III effector repertoire of *Pseudomonas syringae* pv. *syringae* B728a and its role in survival and disease on host and non-host plants." Mol Microbiol **62**(1): 26-44.

Warren, A. S., R. Anandakrishnan, et al. (2010). "Functional bias in molecular evolution rate of *Arabidopsis thaliana*." BMC Evol Biol **10**: 125.

Wei, C. F., B. H. Kvitko, et al. (2007). "A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*." Plant J **51**(1): 32-46.

Wichmann, G., D. Ritchie, et al. (2005). "Reduced genetic variation occurs among genes of the highly clonal plant pathogen *Xanthomonas axonopodis* pv. *vesicatoria*, including the effector gene *avrBs2*." Appl Environ Microbiol **71**(5): 2418-2432.

Wills, B., J. Leikin, et al. (2008). "Analysis of suspicious powders following the post 9/11 anthrax scare." J Med Toxicol **4**(2): 93-95.

Wroblewski, T., K. S. Caldwell, et al. (2009). "Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of *Pseudomonas* and *Ralstonia*." Plant Physiol **150**(4): 1733-1749.

Xiang, T., N. Zong, et al. (2010). "FLS2, but not BAK1, is a target of the *Pseudomonas syringae* effector AvrPto." Mol Plant Microbe Interact.

Yamaguchi, M. (2001). Japan triggered bubonic plague outbreak, doctor claims. The Associated Press, The Independent.

Yao, J. and C. Allen (2006). "Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*." J Bacteriol **188**(10): 3697-3708.

Yao, N., B. J. Eisfelder, et al. (2004). "The mitochondrion--an organelle commonly involved in programmed cell death in *Arabidopsis thaliana*." The Plant journal : for cell and molecular biology **40**(4): 596-610.

Zhang, W.-w., Y.-h. Hu, et al. (2009). "Identification and characterization of a virulence-associated protease from a pathogenic *Pseudomonas fluorescens* strain." Vet Microbiol **139**(1-2): 183-188.

凤凰卫视. (2009). "华日军组建 731 部队 系统性使用生化武器." Retrieved 08/26/2009, 2009, from http://news.ifeng.com/history/phtv/dsy/200908/0827_5712_1322763.shtml.

Chapter 2

Phylogenetics, Population Genetics and Molecular Evolution of a group of *Pseudomonas syringae* strains

Copyright Statement for Chapter 2:
Copyright © 2008, American Society for Microbiology. All Rights Reserved.
APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2008, p. 3171–3181 Vol. 74, No. 10 0099-2240/08/$08.00+0 doi:10.1128/AEM.00180-08

Title: Role of Recombination in the Evolution of the Model Plant Pathogen *Pseudomonas syringae* pv. *tomato* DC3000, a Very Atypical Tomato Strain▿†

Authors and author affiliations:

Shuangchun Yan,[1] Haijie Liu,[1] Toni J. Mohr,[1] Jenny Jenrette,[1] Rossella Chiodini,[1] Massimo Zaccardelli,[2] João C. Setubal,[3] and Boris A. Vinatzer[1*]

Department of Plant Pathology, Physiology, and Weed Science, Virginia Polytechnic Institute and State University, Latham Hall, Ag Quad Lane, Blacksburg, Virginia 24061[1]; CRA-Centro di Ricerca per l'Orticoltura Sede di Battipaglia, I-84091 Battipaglia (SA), Italy[2]; and Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061[3]

Received 20 January 2008/Accepted 19 March 2008

Footnote:

[*] Corresponding author. Mailing address: Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Latham Hall, Room 551, Ag Quad Lane, Blacksburg, VA 24060. Phone: (540) 231- 2126. Fax: (540) 231-3347. E-mail: vinatzer@vt.edu.

† Supplemental material for this article may be found at http://aem .asm.org/.

∇ Published ahead of print on 31 March 2008.

a. Abstract

Pseudomonas syringae pv. *tomato* strain DC3000 (PtoDC3000) is one of the most intensively studied bacterial plant pathogens today. Here we report a thorough investigation into PtoDC3000 and close relatives isolated from *Antirrhinum majus* (snapdragon), *Apium graveolens* (celery), and Solanaceae and Brassicaceae species. Multilocus sequence typing (MLST) was used to resolve the precise phylogenetic relationship between isolates and to determine the importance of recombination in their evolution. MLST data were correlated with an analysis of the locus coding for the type III secreted (T3S) effector AvrPto1 to investigate the role of recombination in the evolution of effector repertoires. Host range tests were performed to determine if closely related isolates from different plants have different host ranges. It was found that PtoDC3000 is located in the same phylogenetic cluster as isolates from several Brassicaceae and Solanaceae species and that these isolates have a relatively wide host range that includes tomato, *Arabidopsis thaliana*, and cauliflower. All other analyzed tomato isolates from three different continents form a distinct cluster and are pathogenic only on tomato. Therefore, PtoDC3000 is a very unusual tomato isolate. Several recombination breakpoints were detected within sequenced gene fragments, and population genetic tests indicate that recombination contributed more than mutation to the variation between isolates. Moreover, recombination may play an important role in the reassortment of T3S effectors between strains. The data are finally discussed from a taxonomic standpoint, and *P. syringae* pv. *tomato* is proposed to be divided into two pathovars.

b. Introduction

Pseudomonas syringae pv. *tomato* DC3000 (PtoDC3000) is one of the most intensively studied plant pathogen isolates today. It was completely sequenced (6), and a large part of what is known about the plant immune system has been learned by studying the interaction of PtoDC3000 with its hosts *Arabidopsis thaliana* and tomato (*Solanum lycopersicum*), as can be seen from many recent high-profile publications (see references 39 and 47 for examples). However, much less is known about how PtoDC3000 relates to other *P. syringae* strains. Although PtoDC3000 is a rifampin-resistant derivative of the type strain of *P. syringae* pv. *tomato* (9; D. Cuppels, personal communication), its host range (which includes tomato, cauliflower [*Brassica oleracea* var. *botrytis*], and *A. thaliana*) was found to be more similar to that of pathovar maculicola isolates from Brassicaceae species than to the host range of other *P. syringae* pv. *tomato* strains (which are limited to tomato) (10, 58). Also, based on physiological (10) and molecular analyses (10, 63), PtoDC3000 was suggested to be more similar to pathovar *maculicola* strains than to other pathovar tomato strains. However, since strains of pathovars tomato, maculicola, antirrhini (isolated from ornamental snapdragon, *Antirrhinum majus*), and apii (isolated from celery, *Apium graveolens*) were all found to be closely related (18, 25), the precise relationship of PtoDC3000 with strains of these pathovars has been difficult to resolve. Therefore, PtoDC3000 is still generally considered to be a member of pathovar tomato.

Multilocus sequence typing (MLST) is a powerful approach to resolve the phylogenetic relationship at the inter- and intraspecies levels (34, 35). MLST is based on the sequencing of six or more fragments of housekeeping genes that are under purifying selection. Isolates with identical alleles at each locus are then grouped into sequence types (STs). Alternatively, sequences at all analyzed loci are

concatenated, and phylogenetic trees can be constructed on the concatenated sequences. MLST allows for the determination of the contribution of homologous recombination (as a consequence of conjugation, transduction, or transformation [41] to variation between strains of a species). While homologous recombination has been found to play an important role in the evolution of several human pathogen species (54), MLST analysis of the plant pathogens *P. syringae*, *Xylella fastidiosa*, and *Ralstonia solanacearum* (see references 49, 50, and 7, respectively) indicated that these species are mostly clonal; i.e., the variation between strains of these species appears to be caused more by mutation than by recombination. However, closely related strains of the plant pathogen species *Pseudomonas viridiflava* were found to recombine frequently (19). Since only a relatively small number of closely related strains of *P. syringae* had been analyzed, it was suggested that recombination may also be found to occur at a high rate in *P. syringae* and other plant pathogen species if a larger number of closely related strains were to be analyzed (19).

What might be the role of homologous recombination in the evolution of plant-pathogenic bacteria? It is well known that horizontal gene transfer and the loss of plasmids and pathogenicity islands (PAIs) (22, 24) by conjugation and site-directed recombination play an important role in the evolution of virulence gene repertoires in plant-pathogenic bacteria (1, 43), in particular, in the acquisition and loss of genes coding for so-called effector proteins that are translocated from many plant-pathogenic bacteria into plant cells through the type III secretion (T3S) system and that have an important function in virulence (20). If recombination were found to be frequent between closely related *P. syringae* strains, homologous recombination may also play a similarly important role in reshuffling virulence genes between strains.

Here we report an MLST study and host range analysis of PtoDC3000 and a worldwide collection of closely related strains of pathovars tomato, maculicola, apii, and antirrhini that made it possible for us to precisely resolve their phylogenetic relationship. Moreover, recombination analysis suggests that homologous recombination significantly contributed to the variation between strains and to the evolution of T3S effector repertoires.

c. Materials and Methods

Bacterial isolates.

The *P. syringae* isolates used in this study are given in Table 1. We are very grateful to our colleagues (Table 1) who generously shared their isolates with us.

PCR and DNA sequencing of gene fragments. Primers were designed on 24 *P. syringae* genes.

Gene sequences of the three sequenced *P. syringae* genomes (6, 16, 28) were aligned in SeqMan (Lasergene; DNAStar, Madison, WI). Fifty- to 100-bp-long regions with high sequence identity between the three sequenced genomes at an approximate distance of 500 to 800 bp from each other were chosen as locations for forward and reverse primers. *Pto*DC3000 sequences from these regions were used for primer design in Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). While some primers annealed to all three genomes without the need for degeneration, other primers had to be degenerated to anneal to all three sequenced genomes. Primer sequences are given in Table S1 in the supplemental material.

Gene fragments were amplified from genomic DNA of *P. syringae* isolates extracted with the Puregene DNA purification system cell and tissue kit (Gentra Systems, Minneapolis, MN). The following DNA polymerases were used for PCRs: Eppendorf HotMaster Taq DNA polymerase (Brinkmann, Bestbury, NY) and Qiagen HotStarTaq and Qiagen Taq (Valencia, CA). Most primer pairs were used with a 58°C annealing temperature and 1 min extension time. For some primer pairs on some bacterial isolates, the annealing temperature was lowered or raised for optimal results. Instructions from polymerase manufacturers were followed for all other cycling conditions. All PCRs were performed on Eppendorf Mastercycler ep gradient thermocyclers (Brinkmann, Bestbury, NY). A total of 10 μl of PCR mixtures was cleaned for sequencing by using 1 unit shrimp alkaline phosphatase (USB Corp., Cleveland, OH) and 1 unit exonuclease I (USB Corp., Cleveland, OH).

DNA sequencing was carried out at the University of Chicago Cancer Research Center DNA Sequencing Facility. Chromatograms were reviewed and edited with SeqMan (Lasergene; DNAStar, Madison, WI).

Molecular evolutionary analysis.

Edited sequences were aligned in BioX 1.0b2 to 1.1b1 (E. Lagercrantz [http://www.lagercrantz.name/software/biox/]) using ClustalW 1.83 as the backend with default parameters. BioX is a graphical user interface for the eBiotools software package (http://www.ebioinformatics.org).

Bayesian trees were generated in MrBayes 3.1.2 (26, 46) using the Markov chain Monte Carlo method. The evolutionary model was set to GTR (general time reversible) with gamma-distributed rate variation across sites and a proportion of invariable sites. The program was run for 300,000 generations, which was long

enough to ensure the standard deviation of split frequencies to be below 0.01. The sample frequency was 10. When summarizing the substitution model parameters and trees, 7,500 samples were used for the burn-in. Potential scale reduction factor values were all close to 1.0. The whole process was independently repeated three times to ensure convergence on the same tree. The Bayesian tree was rooted with *P. syringae* pv. *syringae* strain B728a (*Psy*B728a) as the outgroup in TreeView PPC 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

In order to determine the importance of recombination in the evolution of the analyzed isolates, genes with more than 10 informative sites were tested using the homoplasy test (37) in START 1.0.8 (29). DNA sequences were concatenated in frame for the homoplasy test. Phylogenetic networks were generated in Splits-Tree 4.6 (27), using the NeighborNet (5) algorithm. The Web-based service GARD (genetic algorithm for recombination detection) (30) was employed to detect and locate recombination breakpoints. GARD's built-in tool was used to predict evolutionary models. The Shimodaira-Hasegawa (SH) test (52), implemented in PAUP* 4.0b10 (55), was performed on partitions flanking breakpoints predicted by GARD to determine their significance. The SH test determines the likelihood of a data set, given alternative trees.

Population recombination and mutation rates were estimated using the composite likelihood method in LDhat 2.0 (38). LDhat's built-in likelihood permutation test was used to test for the presence of recombination. The ratios of nonsynonymous to synonymous evolutionary changes for detecting positive selection were estimated by codonml of the PAML 3.15 package (61). Pair-wise sequence percent identities were calculated with MegAlign (Lasergene; DNA-Star, Madison, WI).

BioX was used to convert sequence data file formats. Tree files were converted to SVG format with TreeView X 0.5.0

(http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/). Phylogenetic trees and network graphics were scaled and edited with Inkscape 0.45 (http://www.inkscape.org) and Illustrator CS3 (Adobe Systems Inc.) for publishing. ModelTest, MrBayes, LDhat, and PAML were compiled from the source in Mac OS X 10.4 to be run on Intel-based Mac computers.

Plant growth conditions and bacterial infections.

Plants were grown in a Percival Scientific CU-32L growth chamber (Perry, IA) in a 1:1 mixture of Pro-Mix BX and Pro-Mix PGX (Premier Horticulture Inc., Quakertown, PA). Plants were grown under 16-h days at 22°C and infected when 3 to 4 weeks old.

Isolates of *P. syringae* were streaked from glycerol stocks onto King's broth plates and grown at 30°C for 24 to 48 h. Bacteria were then restreaked onto new plates, covering the entire plate, and grown for another 24 h. Bacteria were scraped off the plates, resuspended, and diluted in sterile 10 mM MgSO4 for infections.

For the determination of disease symptoms, tomato cultivars Sunpride and Rio Grande, *A. thaliana* ecotype Columbia *rps2*, and cauliflower cultivar Early Snowball A were spray inoculated. Plants were placed into plastic bags and watered with 50 ml deionized water. Bags were sealed to maintain high humidity. Twenty-four hours later, the plants were sprayed with 20 ml of 10 mM MgSO4 containing 1.2 × 108 bacteria/ml. Twenty-four hours after inoculation, the plants were removed from the bags. Leaves were photographed 1 week after the date of infection. All pictures were taken with an Olympus Camedia C-765 digital camera. For the measurement of bacterial populations, plants were infected by dipping (tomato and cauliflower) or spraying (*A. thaliana*), including 0.02% of Silwet in

the bacterial suspensions in order to obtain a uniform distribution of bacteria on the leaf surfaces.

Nucleotide sequence accession numbers.

All ST sequences were deposited in GenBank under the accession numbers EU296540 to EU296598.

d. Results

The sequenced PtoDC3000 belongs to a group of closely related isolates from cultivated and wild plants around the world.

To confirm the close relationships between strains of pathovars tomato, maculicola, antirrhini, and apii reported in the literature, more than 100 isolates of these pathovars were assembled. A preliminary sequence analysis of the *gyrB* gene (data not shown) revealed that 83 isolates, including PtoDC3000, were closely related to each other, with DNA identities of more than 98.7%. In contrast, these isolates were only 92.3% identical, on average, to the other sequenced *P. syringae* isolates PsyB728a (16) and *P. syringae* pv. *phaseolicola* 1448A (Pph1448A) (28). Fifty-two representative isolates of the initial 83 isolates were chosen for further analysis to determine their phylogenetic relationship with PtoDC3000 and to test the hypothesis that closely related *P. syringae* isolates have high rates of recombination. Isolates were chosen based on preliminary sequence analysis of a subset of gene fragments, the geographic locations, the year of isolation, and the host of isolation, maximizing the diversity of the analyzed sample. Isolates are given in Table 1, including plant species, the year of collection, and the geographic location of isolation, when available.

TABLE 1. *P. syringae* strains used in this study (listed in the same order as in Table 2)

Pathovar	Strain name	Host of isolation (common name)	Host of isolation (scientific name)	Location	Collector of strain	Yr of collection	Source of strain	Reference
Antirrhini	126	Snapdragon	*A. majus*		M. Moffett	1965	D. Cuppels	
Antirrhini	152E	Snapdragon	*A. majus*	United Kingdom	J. Taylor	1960	D. Arnold	
Antirrhini	4303	Snapdragon	*A. majus*	United Kingdom	G. Jones	1965	D. Arnold	
Tomato	T1	Tomato	*S. lycopersicum*	Canada	G. Bonn		T. Denny	45
Tomato	Max1	Tomato	*S. lycopersicum*	Italy	M. Zaccardelli		M. Zaccardelli	62
Tomato	Max13	Tomato	*S. lycopersicum*	France			M. Zaccardelli	62
Tomato	PST6	Tomato	*S. lycopersicum*	Canada	D. Cuppels		T. Denny	13
Tomato	PT13	Tomato	*S. lycopersicum*		Gitaitis		J. Jones	
Tomato	PT14	Tomato	*S. lycopersicum*		G. Bonn		J. Jones	
Tomato	PT18	Tomato	*S. lycopersicum*	CA	C. Kado		T. Denny	14
Tomato	PT2	Tomato	*S. lycopersicum*	GA	S. McCarter		T. Denny	14
Tomato	PT21	Tomato	*S. lycopersicum*	FL	T. Howe	1990	J. Jones	
Tomato	PT26	Tomato	*S. lycopersicum*		M. Ricker	1990	J. Jones	
Tomato	PT32	Tomato	*S. lycopersicum*	FL	J. Jones	1993	J. Jones	
Tomato	NCPPB1108	Tomato	*S. lycopersicum*	Jersey, United Kingdom	R. A. Lelliott	1961	D. Arnold	
Tomato	B181	Tomato	*S. lycopersicum*	GA	S. McCarter	1981	T. Denny	14
Tomato	1318	Tomato	*S. lycopersicum*	Switzerland			D. Cuppels	10
Tomato	487	Tomato	*S. lycopersicum*	Greece			D. Cuppels	10
Tomato	KS127	Tomato	*S. lycopersicum*	Tanzania	K. C. Shenge	2004	M. Zaccardelli	51
Tomato	Max14	Tomato	*S. lycopersicum*	Spain			M. Zaccardelli	62
Tomato	JL1065	Tomato	*S. lycopersicum*	CA	J. Lindemann		R. Jackson	57
Tomato	JL1031	Tomato	*S. lycopersicum*	CA	J. Lindemann	1983	T. Denny	14
Tomato	PT28	Tomato	*S. lycopersicum*	Mexico	J. Jones	1992	J. Jones	
Tomato	PT29	Tomato	*S. lycopersicum*	Mexico	J. Jones	1992	J. Jones	
Tomato	PT30	Tomato	*S. lycopersicum*	Mexico	J. Jones	1992	J. Jones	
Tomato	PST26L	Tomato	*S. lycopersicum*	South Africa	M. Hattingh		D. Cuppels	10
Tomato	KS112	Tomato	*S. lycopersicum*	Tanzania	K. C. Shenge	2004	M. Zaccardelli	51
Maculicola	F1	Spinach mustard	*Brassica rapa* var. *perviridis*	OK		1995	J. Damicone	63
Maculicola	F7	Spinach mustard	*B. rapa* var. *perviridis*	OK		1995	J. Damicone	63
Apii	1089	Celery	*A. graveolens*	CA	D. A. Cooksey		D. Arnold	
Maculicola	F15	Turnip	*B. rapa* var. *rapifera*	OK		1995	J. Damicone	63
Maculicola	M3	Cauliflower	*Brassica oleracea* var. *botrytis*	United States	W. Burkholder	1937	J. Greenberg	12
Maculicola	M1	Cauliflower	*B. oleracea* var. *botrytis*	United Kingdom	R. Lelliott	1965	J. Greenberg	12
Maculicola	M2	Cauliflower	*B. oleracea* var. *botrytis*	New Zealand	D. Shackleton	1965	J. Greenberg	12
Maculicola	M6	Cauliflower	*B. oleracea* var. *botrytis*	United Kingdom	G. E. Jones	1965	J. Greenberg	12
Maculicola	M8	Kale	*B. oleracea* var. *acephala*	United Kingdom	J. Taylor		J. Greenberg	12
Maculicola	1766	Cauliflower	*B. oleracea* var. *botrytis*	United Kingdom	G. E. Jones	1965	D. Cuppels	10
Tomato	ICMP3443	Woolly nightshade	*S. mauritianum*	New Zealand	D. R. W. Watson	1972	J. Young	
Tomato	ICMP3449	Woolly nightshade	*S. mauritianum*	New Zealand	D. R. W. Watson	1972	J. Young	
Tomato	DC3000[a]	Tomato	*S. lycopersicum*	Guernsey, United Kingdom	R. A. Lelliott	1961	J. Greenberg	10
Tomato	OH314	Nettle	*S. carolinense*[b]	OH	D. Coplin	1978	D. Cuppels	10
Maculicola	F6	Kale	*B. oleracea* var. *acephala*	OK		1995	J. Damicone	63
Maculicola	F9	Spinach mustard	*B. rapa* var. *perviridis*	OK		1995	J. Damicone	63
Maculicola	F10A	Turnip	*B. rapa* var. *rapifera*	OK		1995	J. Damicone	63
Maculicola	F18	Kale	*B. oleracea* var. *acephala*	OK		1995	J. Damicone	63
Maculicola	F19	Turnip	*B. rapa* var. *rapifera*	OK		1996	J. Damicone	63
Maculicola	F16	Turnip	*B. rapa* var. *rapifera*	OK		1995	J. Damicone	63
Maculicola	F17	Spinach mustard	*B. rapa* var. *perviridis*	OK		1995	J. Damicone	63
Tomato	ICMP3435	Woolly nightshade	*S. mauritianum*	New Zealand	D. R. W. Watson	1972	J. Young	57
Tomato	ICMP3455	Woolly nightshade	*S. mauritianum*	New Zealand	D. R. W. Watson	1972	J. Young	57
Tomato	ICMP9305	Woolly nightshade	*S. mauritianum*	New Zealand	D. R. W. Watson	1987	J. Young	
Maculicola	84-59	Cauliflower	*B. oleracea* var. *botrytis*	CA	W. Wiebe		D. Cuppels	58

[a] PtoDC3000 is a rifampin-resistant derivative of NCPPB1106 (ICMP2844, CFBP2212), which is the pathotype strain of *P. syringae* pv. tomato.

[b] The botanical species was not determined at the time of collection (D. Coplin, personal communication). We believe it is probably *S. carolinense* (horse nettle), which is a close relative of tomato.

The *gyrB* gene was one of seven core genome genes (*gyrB*, *rpoD*, *fruK*, *pgi*, *gap1*, *gltA*, and *acnB*) used in a previous MLST study of *P. syringae* by Sarkar and Guttman (49). We used six of these genes and an additional three core genome genes located in proximity to *avrPto1* to investigate the phylogeny of our collection in more detail; *fruK* was excluded from the analysis because of its high level of conservation. The positions of the nine analyzed genes in the PtoDC3000 genome are shown in Fig. 1. Table S1 in the supplemental material lists the primers used to amplify fragments of these genes. Fourteen different unique allele profiles were identified among the isolates (from now on called STs). Table 2 lists all isolates belonging to each ST. The sequences of the nine genes of each isolate were also concatenated in the order they are found in the PtoDC3000 genome. A Bayesian tree, a maximum likelihood (ML) tree, and a neighbor-joining tree were constructed with the concatenated sequences and rooted with the sequenced PsyB728a (and/or Pph1448A) isolate as the outgroup. The Bayesian consensus tree is shown in Fig. 2. The neighbor-joining tree (not shown) and the ML tree (not shown) have the same overall topology and similar branch lengths. Two main groups of isolates (I and II in Fig. 2) can be distinguished within all three trees. Group I contains the two subgroups Ia and Ib. Subgroup Ia contains all pathovar tomato isolates from tomato (with the exception of PtoDC3000), the *P. syringae* pv. *antirrhini* isolates, and two *P. syringae* pv. *maculicola* strains with identical STs (*P. syringae* pv. *maculicola* F1 [PmaF1] is shown in the tree). Several pathovar maculicola isolates and two *Solanum mauritianum* (wooly nightshade) isolates form group Ib. Group II contains two isolates with identical MLST profiles, PtoDC3000 from tomato and the PtoOH314 isolate from nettle (possibly horse nettle, *Solanum carolinense*, but unfortunately the species was not determined at the time of collection in 1978; D. Coplin, personal communication). Group II also contains 11 other almost-identical isolates (only one nucleotide

difference in almost 6,000 bp compared to PtoDC3000), eight pathovar maculicola isolates from *Brassicaceae* and three *S. mauritianum* isolates. The *P. syringae* pv. *apii* 1089 isolate (Pap1089) is on a branch by itself but with low statistical support, making its placement outside or inside group I uncertain.

An ML tree built on only the six housekeeping genes *acnB*, *gap1*, *gltA*,

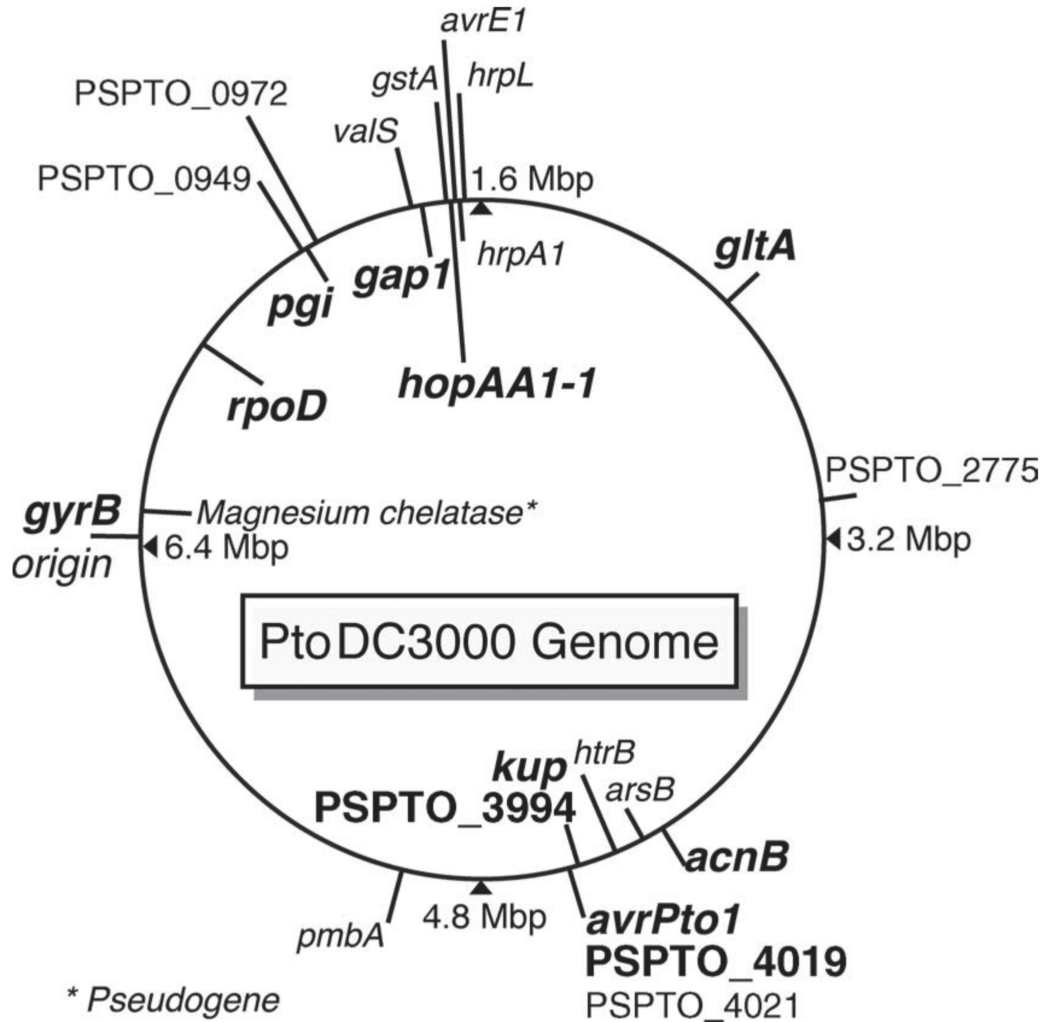


FIG. 1. Positions of all analyzed genes in the genome of PtoDC3000. Genes that were sequenced for all isolates given in Table 1 are bold and in larger font. All other genes were sequenced only in five isolates used for the analysis shown in Fig. S3 in the supplemental material and in Table 6. For the exact gene location, primers used, and lengths of sequenced fragments, see Table S1 in the supplemental material.

gyrB, *pgi*, and *rpoD* had a topology that was very similar to the trees built on all nine genes, showing that the three genes *kup*, PSPTO_3994, and PSPTO_4019 close to the *avrPto1* locus did not significantly alter tree topology (data not shown).

Isolates very closely related to PtoDC3000 have different host ranges.

Since the analyzed isolates are all very closely related to each other—some of the pathovar maculicola isolates were reported to also cause disease on tomato (10, 58) and PtoDC3000 is well known to cause disease on *A. thaliana*, cauliflower, and tomato (6, 10)—representative isolates were inoculated under controlled conditions to determine whether host range differences existed between isolates. Inoculations were performed by spraying bacteria onto leaf surfaces without the addition of any surfactant to make them resemble natural infections as much as possible. In all cases, symptoms always appeared at least 3 days after infection, making it very unlikely that symptoms were caused by a hypersensitive defense response (21), which usually becomes macroscopically visible only when bacteria are directly infiltrated into leaves at high doses and appears within a day after infection. Figure 3 shows that the pathovar tomato isolates of ST3, -4 and -5 (see Table 2 for the list of isolates belonging to each ST) caused the most severe symptoms on tomato cultivar Sunpride. The isolates of ST6, -8, -9, and -11 to -15, including PtoDC3000, caused mild symptoms. Only when the concentration of the inoculum was increased 10-fold did we obtain disease symptoms on tomato, and these isolates were similar to those caused by the pathovar tomato isolates of ST3, -4, and -5 (data not shown). ST1, -2, -7, and -10 did not cause any disease symptoms on tomato. Inoculations of the tomato cultivar Rio Grande caused

similar results, although differences in symptom severity between ST3, -4, and -5 on one hand and ST6, -8, -9, and -11 to -15 on the other hand were less pronounced (data not shown).

TABLE 2. Strains belonging to the STs in the Bayesian consensus tree (Fig. 2) and the NeighborNet network (Fig. 4)

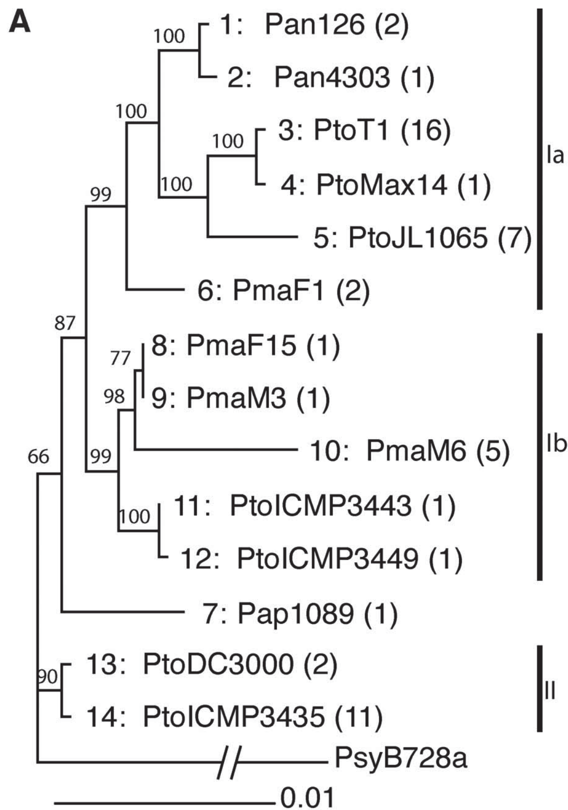
ST no.	Isolate(s)[a]
1	**Pan126**, Pan152E
2	**Pan4303**
3	**PtoT1**, PtoMax1, PtoMax13, PtoPST6, PtoPT13, PtoPT14, PtoPT18, PtoPT2, PtoPT21, PtoPT26, PtoPT32, PtoNCPP1108, PtoB181, Pto1318, Pto487, PtoKS127
4	**PtoMax14**
5	**PtoJL1065**, PtoJL1031, PtoPT28, PtoPT29, PtoPT30, PtoPST26L, PtoKS112
6	**PmaF1**, PmaF7
7	**Pap1089**
8	**PmaF15**
9	**PmaM3**
10	**PmaM6**, PmaM1, PmaM2, PmaM8, Pma1766
11	**PtoICMP3443**
12	**PtoICMP3449**
13	**PtoDC3000**, PtoOH314
14	**PmaF9**, PmaF6, PmaF10A, PmaF16, PmaF17, PmaF18, PmaF19, Pma84-59, PtoICMP3435, PtoICMP3455, PtoICMP9305

[a] Representative strains shown in the tree in Fig. 2 are bold.

Cauliflower and *A. thaliana* infections (see Fig. S1A and S1C in the supplemental material) also revealed remarkable differences in symptoms caused by the different isolates, which are also summarized in Fig. 2. For example, the ability of PtoDC3000 to cause disease on *A. thaliana* and on cauliflower, reported

in the literature, was confirmed. *P. syringae* pv. *tomato* isolates of ST3, -4, and -5 were found to be unable to cause any symptoms on these two plant species. Using an *A. thaliana* 'Columbia' accession mutated in the *RPS2* resistance gene (40), it was shown that the inability of *P. syringae* pv. *tomato* isolates with ST3, -4, and -5 to cause disease on *A. thaliana* was not due to the long-known gene-for-gene interaction between the cognate *A. thaliana* resistance gene *RPS2* (57) and the T3S effector gene *avrRpt2* (which was confirmed by PCR to be present in all *P. syringae* pv. *tomato* isolates of ST3, -4, and -5, with the exception of isolate PtoMax13 [data not shown]). Moreover, these STs were even unable to cause any disease on the *A. thaliana* sgt1 (2), sid2 (59), and pad4 (64) (data not shown) defense mutants, suggesting multigenic non-host resistance. A subset of isolates was tested on snapdragon and celery, which confirmed host range differences on these plant species as well (data not shown). For selected *P. syringae* pv. *tomato* and *P. syringae* pv. *maculicola* isolates, bacterial populations were measured on tomato, cauliflower, and *A. thaliana* 3 days after infection, revealing that symptom severity correlated well with bacterial population size (see Fig. S1B, S1D, and S1E in the supplemental material). For example, on tomato, the strains PtoT1 and PtoMax1 that caused the most severe symptoms grew to almost 100-fold-higher population densities than PtoDC3000 and PmaF18, which had caused only mild symptoms.

Recombination contributed to the evolution of the core genome of PtoDC3000 and closely related isolates.



B

Strain	ST	Origin	*avrPto1*	Cauliflower	*A. thaliana rps2*	Tomato
Pan126	1	Snapdragon	−	−	−	−
PtoMax1	3	Tomato	+	−	−	++
PtoJL1065	5	Tomato	+	−	−	++
PmaF1	6	Spinach mustard	+	++	+	+
Pap1089	7	Celery	−	−	−	−
PmaF15	8	Kale	+	+	+	+
PmaM3	9	Cauliflower	+	+	+	+
PmaM6	10	Cauliflower	−	+	−	−
PmaM2	10	Caulifllower	−	+	−	−
ICMP3443	11	Woolly nightshade	+	++	++	+
PtoDC3000	13	Tomato	+	++	++	+
PtoOH314	13	Nettle	+	++	++	+
PmaF18	14	Turnip	+	++	++	+
ICMP3435	14	Woolly nightshade	+	++	++	+

FIG. 2. Bayesian tree of the concatenated gene fragments and host range of representative isolates with regard to *A. thaliana*, cauliflower, and tomato. (A) Each ST is identified by its number (preceding the colon) and by a representative isolate. The number of isolates belonging to each ST is indicated in parentheses. See Table 2 for a complete list of isolates

Before using the nine sequenced gene fragments for recombination analyses, we determined that all genes are under purifying selection (Table 3) and are thus well suited for evaluating the importance of recombination in the evolution of the analyzed isolates.

Since the homoplasy test (37) is well suited for very closely related isolates with a DNA identity of more than 98% (44), it was the first test applied to our data. Homoplasies are defined as mutations shared between different branches of a phylogenetic tree that have not been directly inherited from an ancestor. The homoplasy test calculates the ratio between homoplasies minus the expected homoplasies in the case of no recombination and calculates the expected homoplasies in the case of free recombination minus the expected homoplasies in the case of no recombination. The closer this ratio is to 1, the more recombination can be inferred. The homoplasy ratio for the genes *rpoD*, *gyrB*, and *kup* and for the

belonging to each ST. The credibility values (×100) of clades, indicating the statistical significance of groupings, are given in front of each node. The sequenced isolate PsyB728a (16) was used as the outgroup. (B) The host of isolation (origin), the experimentally determined host range on selected plant species (−, no disease symptoms; +, mild disease symptoms; ++, severe disease symptoms), and the presence of the *avrPto1* gene are indicated for representative isolates of most STs (some STs that differ from another ST by only a single nucleotide were not included in host range tests). *A. thaliana rps2* is a mutant of *A. thaliana* with a nonfunctional *RPS2* gene (40).

concatenated sequence of all nine genes were found to be 0.408, 0.160, 0.623, and 0.413, respectively. The obtained ratios indicate the presence of recombination in two of three genes and in the concatenated sequence. It was not possible to calculate the homoplasy ratio for all other genes, since only *rpoD*, *gyrB*, and *kup* have the 10 or more informative sites required for this test in the START package (29).

Because the homoplasy test indicated recombination, the phylogenetic tree shown in Fig. 2 may not be an accurate representation of the phylogeny of the analyzed isolates. Since alternative topologies cannot be represented in a tree, a tree cannot reflect recombination. A tree is always only a "compromise" of the different possible trees that can be built on an alignment when conflicting signals

TABLE 3. Percent DNA identities between isolates used in this study and between the three sequenced *P. syringae* isolates and ratio of nonsynonymous to synonymous mutations for all analyzed genes

Gene	% DNA identity				*dN*/*dS* ratio[b]
	Avg	Minimum	Maximum	3S Avg[a]	
acnB	99.1	98.2	99.8	94.0	0
gap1	99.7	99.3	99.8	92.4	0.0775
gltA	99.3	99.2	99.6	95.9	0
gyrB	99.3	98.7	99.9	92.3	0
pgi	99.4	98.9	99.8	90.9	0.0313
kup	98.7	97.2	99.9	91.1	0.0056
PSPTO_3994	99.5	99.0	99.9	84.43	NA
PSPTO_4019	98.8	98.3	99.4	92.6	0
rpoD	98.8	98.0	99.8	93.6	0
All	99.34	98.80	99.99	NA	NA

[a] Percent DNA identities between the three completely sequenced *P. syringae* genomes (PtoDC3000, PsyB728a, and Pph1448A).

[b] Ratios of nonsynonymous to synonymous evolutionary changes were calculated based on the sequences of all closely related isolates used in this study by using codonml of the PAML 3.15 package (61). NA, not applicable.

are present. Phylogenetic networks have been developed to overcome this inherent shortcoming of trees (17). In a network, alternative phylogenies are represented by splits. The more splits (or reticulations) there are in a network, the more conflicting signals— possibly due to recombination—exist in the data. We built a phylogenetic network on the concatenated sequences of the nine core genome genes using the NeighborNet algorithm (5), which is similar to the commonly used splits decomposition algorithm but is better suited for handling large data sets, and the generated networks are generally more resolved (5). Figure 4 shows that there are a large number of splits in the network built on the concatenated sequence, indicative of conflicting phylogenetic signals.

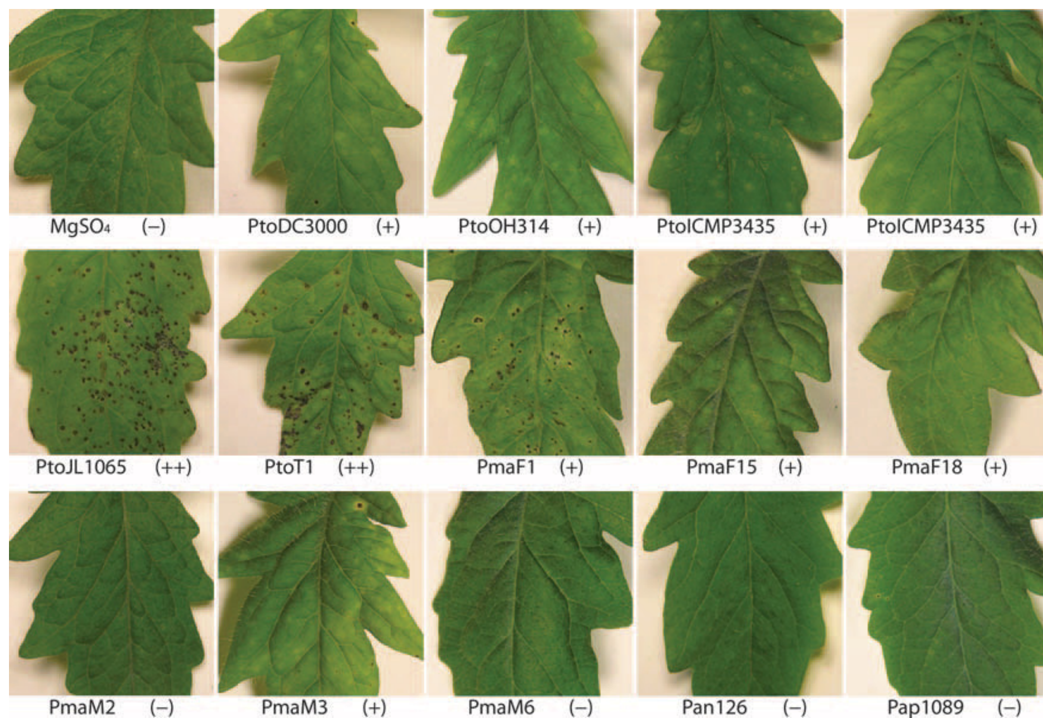


FIG. 3. Virulence tests of representative *P. syringae* isolates on tomato. Entire tomato plants were sprayed with bacterial suspensions of *P. syringae* isolates. Photographs of representative leaves were taken 1 week after infection. ++, highly virulent *P. syringae* isolates causing severe disease symptoms (large numbers of necrotic pits with chlorotic haloes); +, isolates causing mild disease symptoms (few pits and diffused chlorosis); −, isolates causing no symptoms.

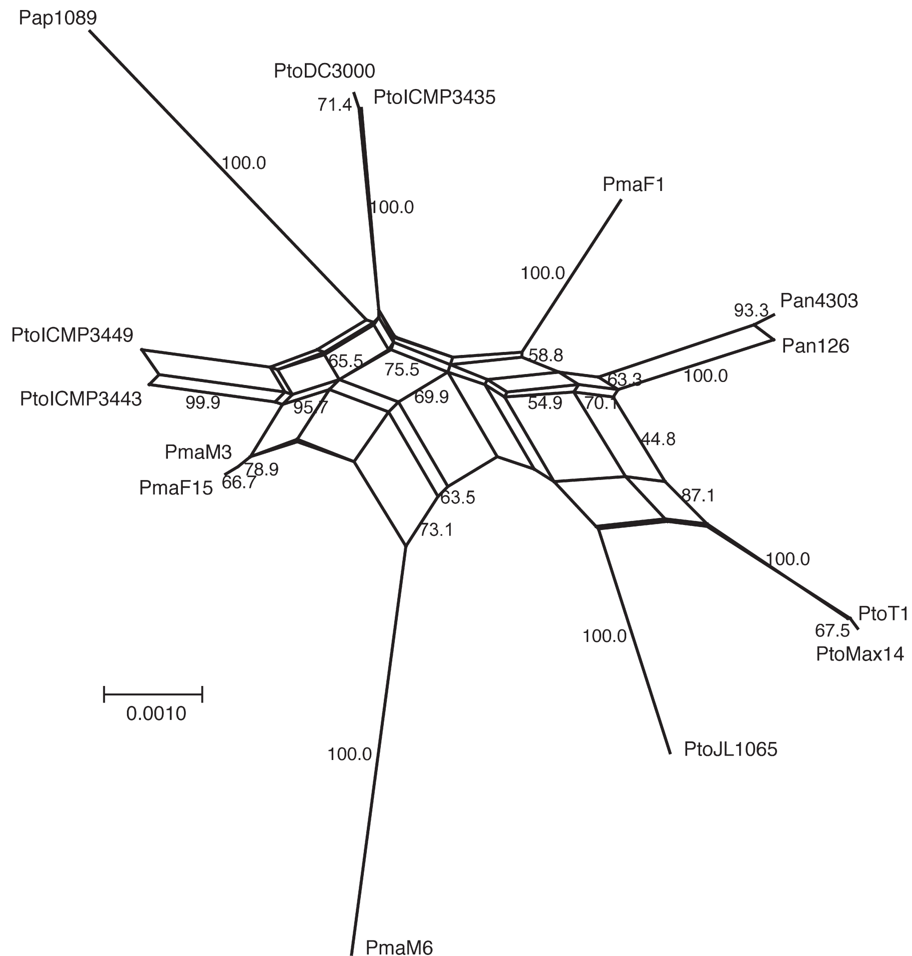


FIG. 4. NeighborNet analysis of the nine concatenated housekeeping genes *gyrB*, *rpoD*, *pgi*, *gap1*, *gltA*, *acnB*, *kup*, PSPTO_3994, and PSPTO_4019. One representative isolate of each ST and bootstrap values higher than 40 are shown.

To look more carefully at the individual gene level, NeighborNet networks were built for each gene fragment (see Fig. S2 in the supplemental material). Six of the nine sequences show various degrees of reticulation. For the three gene fragments with the greatest percentage of DNA identity between isolates, no reticulation was identified, possibly because the number of nucleotide differences in these fragments was insufficient. To identify recombination breakpoints and subsequently determine if recombination could explain the reticulation in the other six genes, we used the GARD program (30), which is a likelihood-based model selection procedure. A breakpoint was predicted based on the improvement of the Akaike information criterion score for trees constructed on the partitions flanking the predicted breakpoint compared to a tree constructed on the entire sequence

(30). Table 4 shows the GARD output for all genes. As suggested by the NeighborNet networks, GARD found breakpoints only in the most divergent genes. NeighborNet networks were then built on all predicted partitions of these genes. In many cases, NeighborNet did not find any reticulations in the gene partitions predicted by GARD, confirming the prediction of the recombination breakpoints (Table 4).

TABLE 4. Recombination breakpoints predicted in analyzed gene fragments

Gene	No. of breakpoints[a]	Length of fragments for each partition (nt)	NeighborNet results for partitions[c]	No. of detectable breakpoints/ total no. of breakpoints ($P < 0.05$)[d]
gyrB	1	358, 343	No reticulation	1/1
rpoD	1	404, 241	Fewer splits for both partitions	1/1
pgi	0	NA[e]	NA	NA
gap1	0	NA	NA	NA
gltA	0	NA	NA	NA
acnB	0	NA	NA	NA
kup and PSPTO_3994[b]	3	720, 273, 138, 812	2/4[c]	3/3
PSPTO_4019	3	233, 198, 189, 100	2/4[c]	1/3

[a] Number of breakpoints predicted by GARD (30).
[b] *kup* and PSPTO_3994 is the sequence of part of the *kup* gene and part of the PSPTO_3994 gene, including the intergenic region between the two genes.
[c] Number of partitions that show no reticulation out of the total number of partitions.
[d] A breakpoint was considered detectable by the SH test if at least one flanking tree was significantly worse at fitting the partition on the other side of the breakpoint. Values represent detectable breakpoints over the number of breakpoints tested.
[e] NA, not applicable.

The significance of the predicted breakpoints was further analyzed by performing the SH test (52) for all partition pairs separated by a breakpoint. The SH test is widely used to determine the significance of differences between tree topologies. If a tree built on one sequence alignment is statistically significantly different from the data derived from a second sequence alignment, recombination between the sequences can be inferred, as long as the sequences are under the same selection pressure. The SH test does not detect recombination events that change branch lengths without changing branching patterns (30). Nonetheless, the SH test showed that for six of eight breakpoints that were predicted by GARD (Table 4), trees built on flanking partitions were significantly different from each other, supporting the conclusion that recombination breakpoints exist in many analyzed genes.

TABLE 5. Estimates of ρ and θ

Gene	ρ[a]	Per site ρ	θ	Per site θ	ρ/θ	GARD[b]
gyrB	14.0***	0.0200	4.949	0.00706	2.829	Yes
rpoD	100.0***	0.1550	7.328	0.01136	13.646	Yes
pgi	9.5	0.0167	4.380	0.00770	2.169	No
gap1	13.5	0.0225	1.920	0.00320	7.031	No
gltA	5.0*	0.0099	3.333	0.00657	1.500	No
acnB	7.0	0.00116	5.255	0.00867	1.332	No
kup	4.0***	0.0038	15.45	0.01450	0.259	Yes
PSPTO_3994	70.5	0.0754	5.014	0.00536	14.061	Yes
PSPTO_4019	80.0**	0.1111	8.160	0.01133	9.804	Yes
Mean	33.722	0.046	6.198	0.008	5.847	

[a] Statistical significance for the presence of recombination determined by applying the likelihood permutation test. *, $P < 0.10$; **, $P < 0.05$; ***, $P < 0.01$; ρ, population recombination rate; θ, population mutation rate.

[b] Yes, recombination breakpoints were found by the program GARD (30); No, recombination breakpoints were not detected by GARD.

The best way to compare recombination rates between species is to use a population genetics approach by expressing recombination rates in relation to mutation rates. The ratio between the population recombination rate (ρ) and the population mutation rate (θ) is often used for this purpose. ρ and θ can be calculated using a coalescent theory-based method developed by McVean and coworkers and implemented in the program LDhat (38). Applying the likelihood permutation test within LDhat, we found ρ to be significantly different from zero for several of the genes and ρ/θ values ranged from 0.259 for kup to 14.061 for PSPTO_3994 (Table 5). When the ρ/θ values obtained for the analyzed genes were simply averaged, the contribution of recombination was estimated to be more than five times greater than the contribution of mutation to variation between isolates. In comparison, by doing the same calculation using a mix of closely and distantly related isolates of *P. syringae*, the mean ρ/θ was found to be only 0.252 (49). The mean ρ/θ for *P. viridiflava* was 0.48 overall but 2.38 and 10.16, respectively, for

closely related isolates in clades A and B (19). ρ/θ was found to be zero for *Escherichia coli* overall (42) but 2.139 for clade D of *E. coli*, which contains closely related, highly virulent isolates (60). Therefore, as for *P. viridiflava* and *E. coli*, recombination appears to greatly contribute to the variation between closely related isolates of *P. syringae*, while more distantly related isolates appear to mainly differ from each other because of mutation.

Acquisition and loss of the *avrPto1* PAI and its role in host range evolution.

AvrPto1 is a well-studied T3S effector protein. It contributes to virulence on susceptible tomato cultivars (8, 31, 33), although it induces plant defenses on tomato plants that carry the *Pto* resistance gene (45). Comparing the three sequenced *P. syringae* genomes by using the multiple genome alignment program MAUVE (11), the PtoDC3000 *avrPto1* gene is located on a 23,532-bp-long, PtoDC3000-specific region between nucleotides 4506925 and 4530456. Besides avrPto1, the PtoDC3000-specific region contains several transposase genes and a defective prophage genome (6). A similar prophage genome is present adjacent to the *avrPto1* gene in PsyB728a (16) but in a different genomic context. Together with the fact that *avrPto1* is present in several *P. syringae* strains that are only distantly related to each other (48), this suggests that *avrPto1* was acquired independently by several *P. syringae* strains, possibly as part of a bacteriophage. However, by correlating the phylogenetic tree with the distribution of *avrPto1* in the PtoDC3000 relatives in Fig. 2, the absence of *avrPto1* from the isolates of the *P. syringae* pv. *antirrhini*, Pap1089, and PmaM6 STs can be best explained through a one-time acquisition event of *avrPto1* by an ancestor of all analyzed isolates and the later loss of *avrPto1* from the ancestor of the *P. syringae* pv.

TABLE 6. SH test of gene partitions flanking the *avrPto1* prophage region compared to other regions in the *P. syringae* genome

	trees																					genes[c]
p.[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**16[b]**	17	18	**19[b]**	20	21	
1																						*gyrB*
2																						*gyrB*
3																						*mag. chlt*,[d] *rpoD*
4																						PSPTO_0949
5																						PSPTO_0949, *pgi*
6																						*pgi*
7																						*pgi*, PSPTO_0972
8																						PSPTO_0972
9																						*valS, gap1, gstA*
10																						*hopAA1-1*
11																						*hopAA1-1, hrpA1*
12																						*hrpL, gltA*, PSPTO_2775, *acnB*
13																						*arsB*
14																						*htrB*
15																						*kup*
16[2]																						***kup***
17																						***kup***
18[2]																						*kup*
19[2]																						**PSPTO_3994, PSPTO_4019**
20																						PSPTO_4019, PSPTO_4021, *pmbA*
21																						*pmbA*

[a] p., partitions identified with the program GARD (30) in the concatenated sequence.
[b] Partitions 16 and 19 are significantly different (dark grey, significantly different at a P of <0.01; light grey, significantly different at a P of <0.05) from most other partitions when data are compared with trees and trees are compared with data. Partition 19 is immediately flanking the *avrPto1* prophage region on both sides. Other partitions (for example, partition 12) are significantly different from other genes only when data are compared with trees but not when trees are compared with data.
[c] Genes of which the partitions are a part.
[d] *mag. chlt*, magnesium chelatase.

antirrhini isolates, the ancestor of Pap1089, and the ancestor of the PmaM6 ST isolates (three separate events). Other evolutionary scenarios would require a greater number of events.

If *avrPto1* were lost during evolution by the deletion of the *avrPto1* gene alone or by the excision of the entire *avrPto1* prophage region, similar to the excision of the PPHGI-1 PAI observed in *P. syringae* pv. phaseolicola 1301 (43),

the regions flanking the excised region would not be affected by this event. However, the genes *kup* and PSPTO_3994 immediately up-stream of the 23,532-bp-long, prophagelike PtoDC3000 *avrPto1* region and the gene PSTPO_4019 immediately down-stream of the same region have a very unusual nucleotide substitution pattern and contain several recombination breakpoints (Table 4). For most analyzed genes, the PmaM3 and PmaM6 alleles are either nearly identical to each other but different from the PtoDC3000 alleles or nearly identical to each other and nearly identical to the PtoDC3000 alleles. Only in the partitions flanking the *avrPto1* region are PmaM3 and PtoDC3000 nearly identical to each other but different from PmaM6. We extended the sequence analysis for the five isolates PtoT1, PtoJl1065, PmaM3, PmaM6, and PtoDC3000 to 13 more genes in the genome and confirmed this observation (see Fig. 1 for the location of these genes in the PtoDC3000 genome and see Fig. S3 in the supplemental material for all nucleotide differences between alleles). We used GARD (30) to identify breakpoints in the concatenated sequence of all 23 genes. The SH test (52) shows that even when only the five analyzed isolates were used, the trees built on most gene partitions flanking the *avrPto1* prophage region (in particular, partitions 16 and 19) are significantly different from almost all other analyzed genes in the genome (Table 6). Considering these results, a likely explanation for the loss of avrPto1 from an ancestor of the PmaM6 ST isolates is homologous recombination in core genome genes flanking the *avrPto1* prophage region, during which this region was replaced with a genomic region of a donor strain that did not contain *avrPto1*. Note that this explanation assumes that *avrPto1* was present in the same locus in the PmaM3-like ancestor of PmaM6 as it was in PtoDC3000, which is likely because of their close relationship but impossible to confirm based on our current data. Interestingly, PtoDC3000 and most of its relatives that cause weak or severe disease symptoms on tomato (ST3, -4, -5, -6, -8, -9, and -11 to -15) contain

avrPto1 (as determined by PCR [data not shown]), while the isolates of ST1, -2, -7, and -10 (including PmaM6) that do not cause disease symptoms on tomato do not contain *avrPto1* (Fig. 2), suggesting that the loss of *avrPto1* from the PmaM6 ancestor may have been involved in host range evolution (see Discussion) below.

e. Discussion

Importance of recombination in the evolution of closely related *P. syringae* strains.

While all bacteria were assumed to be clonal, John Maynard Smith (53) revolutionized the field of microbial evolution by developing molecular evolutionary tests that showed that some bacterial species recombine frequently. However, only recently has it become clear that recombination is an important evolutionary mechanism in numerous bacterial species. A prominent example is *E. coli*, a species previously considered to be clonal. Extensive sequence analysis of many isolates has now confirmed earlier evidence (23) of frequent recombination in some *E. coli* clades (60). Recombination in many bacterial species was overlooked before the advent of MLST because recombination can efficiently occur only when DNA sequences of the donor and the recipient are very similar (36); however, it can be detected only when DNA sequences of the donor and the recipient are sufficiently different from each other for recombination events to be inferred from sequence data with statistical significance (44).

In the case of *P. syringae*, little recombination was detected when a mix of distantly and closely related isolates was analyzed, and mutation was found to contribute approximately four times more than recombination to variation (49). The reason for the relatively small contribution of recombination to variation

between the analyzed isolates probably lies in the percentage of DNA sequence identity, which is as low as 91% between distantly related *P. syringae* isolates. Also, the adaptation of *P. syringae* pathovars to different plant species probably makes physical contact between distantly related isolates more difficult. The situation dramatically changed in the current study when closely related isolates with a DNA sequence identity of approximately 99% were analyzed; recombination was found to contribute 5.8 times more than mutation to variation between isolates (using the same population genetics test [38] that was used in the study by Sarkar and Guttman [49]). Moreover, many isolates that we studied have partially overlapping host ranges, which can be expected to give ample opportunity for recombination. Even in cases where a common host between isolates was not identified, isolates may have evolved only recently from ancestors with overlapping host ranges. We were also able to confirm high rates of recombination between closely related *P. syringae* isolates by identifying recombination breakpoints within several genes and by applying various recombination tests. However, for some gene fragments, the DNA percent identity between isolates was too high (more than 99.3%) to allow detection of recombination. The use of additional gene fragments with lower DNA percent identity (between 98% and 99%) will be advantageous in future studies.

Not only *P. syringae* (49) but also the plant pathogen species *X. fastidiosa* and *R. solanacearum* were found to evolve mainly by mutation rather than recombination when a mix of closely and distantly related isolates were analyzed (7, 50). However, when a large number of closely related isolates was analyzed, Goss and colleagues (19) found high recombination rates in *P. viridiflava*; *R. solanacearum* was found to be naturally competent during infection, potentially allowing a very high rate of recombination (4); and recombination rates were high enough in *Xanthomonas* to be detected in controlled coinfections (3). The

combination of these results and our results showing that high recombination rates exist between closely related *P. syringae* isolates suggests that recombination probably plays an important role in the evolution of many bacterial plant pathogen species but that this is easily overlooked when an insufficient number of closely related isolates is sampled.

Role of recombination in determining the distribution of the effector *avrPto1*.

We found that *avrPto1* was probably already present in an ancestor of all analyzed PtoDC3000 relatives and may have been lost later by the ancestor of PmaM6 through homologous recombination in genes flanking the *avrPto1* prophage region. What could have led to the loss of *avrPto1*? The *P. syringae* ancestor containing *avrPto1* can be assumed to have existed before the evolution of the plant resistance gene Pto, which elicits defenses upon recognition of *avrPto1* (45), or to have been a pathogen of plants that did not carry the *Pto* resistance gene. Therefore, *avrPto1* conferred a fitness advantage. However, after the evolution of the *Pto* resistance gene, it became a fitness advantage to lose *avrPto1* when infecting a plant that expressed *Pto*. This is also supported by the fact that several *P. syringae* strains that contain *avrPto* but that are not tomato pathogens can cause disease only on tomato plants that do not carry the *Pto* resistance gene (32). Therefore, recombination leading to a replacement of the genomic region containing *avrPto1* with a region from a *P. syringae* donor that did not contain *avrPto1* (but that possibly contained a different virulence gene) became advantageous.

However, the loss of *avrPto1* cannot explain why PmaM6 does not cause disease on tomato while other isolates which contain *avrPto1* do. The

presence/absence of *avrPto1* is just one of several differences in the T3S effector repertoire between the analyzed isolates. Dot blot experiments revealed that PmaM6 has at least 16 differences in its effector repertoire compared to that of PtoDC3000 (data not shown). Moreover, PtoT1, a pathovar tomato isolate without *avrPto1*, was isolated from tomato and is still pathogenic on tomato (45). Thus, we believe that the loss of *avrPto1* from an ancestor of PmaM6 was only one of the events in the adaptation of PmaM6 to plant species on which avrPto1 is detected by a *Pto*-like resistance gene.

Revealed identity of PtoDC3000.

In spite of the detected recombination between PtoDC3000 and its close relatives, it was possible to clearly resolve their phylogenetic relationship. Correlating the obtained phylogenetic data, the hosts of isolation, and the results from host range tests, it becomes clear that PtoDC3000 is not a typical *P. syringae* pv. tomato strain. Typical *P. syringae* pv. tomato strains form a distinct phylogenetic clade apart from PtoDC3000, have all been isolated from tomato, are more virulent on tomato than PtoDC3000, and do not cause disease on either *A. thaliana* or cauliflower. On the other hand, PtoDC3000 is part of a mixed group of almost identical *P. syringae* pv. maculicola and *P. syringae* pv. tomato isolates from cultivated *Brassicacaeae* and wild *Solanaceae* species that cause disease on tomato, cauliflower, and *A. thaliana*. Since, based on the definition of pathovar, strains are grouped into pathovars based on the host of isolation and the host range (15, 56), the typical *P. syringae* pv. tomato strains and the PtoDC3000-like strains should be separated into two distinct pathovars. In accordance with the rules on naming pathovars (15), the pathovar tomato name would need to remain associated with its deposited type strain, i.e., PtoDC3000/ PtoNCPPB1106. In fact,

associating pathovar tomato with a new type strain would create ambiguity. Therefore, PtoDC3000-like strains would continue to be part of pathovar tomato, and a new pathovar with a new name could be introduced for the typical *P. syringae* pv. tomato strains from tomato that cause disease only on tomato.

f. Acknowledgements

This work was funded through Virginia Tech Start Up funds to Boris A. Vinatzer and a Virginia Tech ASPIRES grant to Boris A. Vinatzer and João C. Setubal.

We thank our generous colleagues (Table 1) for sharing the isolates and thank the undergraduate students Leiya Williams, Nina Long, Eric Hall, and Douglas Chandler for their help with PCR analysis, sequence analysis, and plant infections.

g. References

1. Arnold, D. L., A. Pitman, and R. W. Jackson. 2003. Pathogenicity and other genomic islands in plant pathogenic bacteria. Mol. Plant Pathol. 4:407–420.
2. Austin, M. J., P. Muskett, K. Kahn, B. J. Feys, J. D. Jones, and J. E. Parker. 2002. Regulatory role of SGT1 in early R gene-mediated plant defenses. Science 295:2077–2080.
3. Basim, H., R. E. Stall, G. V. Minsavage, and J. B. Jones. 1999. Chromosomal gene transfer by conjugation in the plant pathogen *Xanthomonas axonopodis* pv. vesicatoria. Phytopathology 89:1044–1049.
4. Bertolla, F., A. Frostegard, B. Brito, X. Nesme, and P. Simonet. 1999. During infection of its host, the plant pathogen *Ralstonia solanacearum* naturally

develops a state of competence and exchanges genetic material. Mol. Plant-Microbe Interact. 12:467–472.

5. Bryant, D., and V. Moulton. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

6. Buell, C. R., V. Joardar, M. Lindeberg, J. Selengut, I. T. Paulsen, M. L. Gwinn, R. J. Dodson, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, S. Daugh- erty, L. Brinkac, M. J. Beanan, D. H. Haft, W. C. Nelson, T. Davidsen, N. Zafar, L. Zhou, J. Liu, Q. Yuan, H. Khouri, N. Fedorova, B. Tran, D. Russell, K. Berry, T. Utterback, S. E. Van Aken, T. V. Feldblyum, M. D'Ascenzo, W. L. Deng, A. R. Ramos, J. R. Alfano, S. Cartinhour, A. K. Chatterjee, T. P. Delaney, S. G. Lazarowitz, G. B. Martin, D. J. Schneider, X. Tang, C. L. Bender, O. White, C. M. Fraser, and A. Collmer. 2003. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. Proc. Natl. Acad. Sci. USA 100:10181–10186.

7. Castillo, J. A., and J. T. Greenberg. 2007. Evolutionary dynamics of *Ralstonia solanacearum*. Appl. Environ. Microbiol. 73:1225–1238.

8. Chang, J. H., J. P. Rathjen, A. J. Bernal, B. J. Staskawicz, and R. W. Michelmore. 2000. *avrPto* enhances growth and necrosis caused by *Pseudomonas syringae* pv. tomato in tomato lines lacking either Pto or Prf. Mol. Plant-Microbe Interact. 13:568–571.

9. Cuppels,D.A 1986.Generation and characterization of Tn5 insertion mutations in *Pseudomonas syringae* pv. tomato. Appl. Environ. Microbiol. 51:323–327.

10. Cuppels, D. A., and T. Ainsworth. 1995. Molecular and physiological char-acterization of *Pseudomonas syringae* pv. tomato and *Pseudomonas syringae* pv. maculicola strains that produce the phytotoxin coronatine. Appl. Envi- ron. Microbiol. 61:3530–3536.

11. Darling, A. C., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.
12. Debener, T., H. Lehnackers, M. Arnold, and J. L. Dangl. 1991. Identification and molecular mapping of a single *Arabidopsis thaliana* locus determining resistance to a phytopathogenic *Pseudomonas syringae* isolate. Plant J. 1:289–302.
13. Denny, T. P. 1988. Differentiation of *Pseudomonas syringae* pv. tomato from *P. syringae* with a DNA hybridization probe. Phytopathology 78:1186–1193.
14. Denny, T. P., M. N. Gilmour, and R. K. Selander. 1988. Genetic diversity and relationships of two pathovars of *Pseudomonas syringae*. J. Gen. Microbiol.134:1949–1960.
15. Dye, D. W., J. F. Bradbury, M. Goto, A. C. Hayward, R. A. Lelliott, and M. N. Schroth. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. Rev. Plant Pathol. 142:153–158.
16. Feil, H., W. S. Feil, P. Chain, F. Larimer, G. DiBartolo, A. Copeland, A. Lykidis, S. Trong, M. Nolan, E. Goltsman, J. Thiel, S. Malfatti, J. E. Loper, A. Lapidus, J. C. Detter, M. Land, P. M. Richardson, N. C. Kyrpides, N. Ivanova, and S. E. Lindow. 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. Proc. Natl. Acad. Sci. USA 102:11064–11069.
17. Fitch,W.M.1997.Networks and viral evolution. J.Mol.Evol.44(Suppl.1):S65–S75.
18. Gardan, L., H. Shafik, S. Belouin, R. Broch, F. Grimont, and P. A. Grimont. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and

description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). Int. J. Syst. Bacteriol. 49:469–478.

19. Goss, E. M., M. Kreitman, and J. Bergelson. 2005. Genetic diversity, recombination and cryptic clades in *Pseudomonas viridiflava* infecting natural populations of *Arabidopsis thaliana*. Genetics 169:21–35.
20. Grant, S. R., E. J. Fisher, J. H. Chang, B. M. Mole, and J. L. Dangl. 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. Annu. Rev. Microbiol. 60:425–449.
21. Greenberg, J. T., and N. Yao. 2004. The role and regulation of programmed cell death in plant-pathogen interactions. Cell. Microbiol. 6:201–211.
22. Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell 87:791–794.
23. Guttman, D. S., and D. E. Dykhuizen. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science 266:1380–1383.
24. Hacker, J., and J. B. Kaper. 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54:641–679.
25. Hendson, M., D. C. Hildebrand, and M. N. Schroth. 1992. Relatedness of *Pseudomonas syringae* pv. tomato, *Pseudomonas syringae* pv. maculicola and *Pseudomonas syringae* pv. antirrhini. J. Appl. Bacteriol. 73:455–464.
26. Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.
27. Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.
28. Joardar, V., M. Lindeberg, R. W. Jackson, J. Selengut, R. Dodson, L. M. Brinkac, S. C. Daugherty, R. DeBoy, A. S. Durkin, M. G. Giglio, R. Madupu, W. C. Nelson, M. J. Rosovitz, S. Sullivan, J. Crabtree, T. Creasy, T. David- sen, D. H. Haft, N. Zafar, L. Zhou, R. Halpin, T. Holley, H.

Khouri, T. Feldblyum, O. White, C. M. Fraser, A. K. Chatterjee, S. Cartinhour, D. J. Schneider, J. Mansfield, A. Collmer, and C. R. Buell. 2005. Whole-genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. J. Bacteriol. 187:6488–6498.

29. Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230–1231.
30. Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23:1891–1901.
31. Lin, N. C., and G. B. Martin. 2005. An *avrPto*/*avrPtoB* mutant of *Pseudomonas syringae* pv. tomato DC3000 does not elicit Pto-mediated resistance and is less virulent on tomato. Mol. Plant-Microbe Interact. 18:43–51.
32. Lin, N. C., and G. B. Martin. 2007. Pto- and Prf-mediated recognition of AvrPto and AvrPtoB restricts the ability of diverse *Pseudomonas syringae* pathovars to infect tomato. Mol. Plant-Microbe Interact. 20:806–815.
33. Macho, A. P., A. Zumaquero, I. Ortiz-Martin, and C. R. Beuzon. 2007. Competitive index in mixed infections: a sensitive and accurate assay for the genetic analysis of *Pseudomonas syringae*-plant interactions. Mol. Plant Pathol. 8:437–450.
34. Maiden, M. C. 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60:561–588.
35. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the

identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA 95:3140–3145.

36. Majewski, J., and F. M. Cohan. 1999. DNA sequence similarity requirements for interspecific recombination in Bacillus. Genetics 153:1525–1533.

37. Maynard Smith, J., and N. H. Smith. 1998. Detecting recombination from gene trees. Mol. Biol. Evol. 15:590–599.

38. McVean, G., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160:1231–1241.

39. Melotto, M., W. Underwood, J. Koczan, K. Nomura, and S. Y. He. 2006. Plant stomata function in innate immunity against bacterial invasion. Cell 126:969–980.

40. Mindrinos, M., F. Katagiri, G. L. Yu, and F. M. Ausubel. 1994. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089–1099.

41. Narra, H. P., and H. Ochman. 2006. Of what use is sex to bacteria? Curr. Biol. 16:R705–R710.

42. Perez-Losada, M., E. B. Browne, A. Madsen, T. Wirth, R. P. Viscidi, and K. A. Crandall. 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. 6:97–112.

43. Pitman, A. R., R. W. Jackson, J. W. Mansfield, V. Kaitell, R. Thwaites, and D. L. Arnold. 2005. Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. Curr. Biol. 15:2230–2235.

44. Posada, D., K. A. Crandall, and E. C. Holmes. 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36:75–97.

45. Ronald, P. C., J. M. Salmeron, F. M. Carland, and B. J. Staskawicz. 1992. The cloned avirulence gene *avrPto* induces disease resistance in tomato cultivars containing the *Pto* resistance gene. J. Bacteriol. 174:1604–1611.
46. Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
47. Rosebrock, T. R., L. Zeng, J. J. Brady, R. B. Abramovitch, F. Xiao, and G. B. Martin. 2007. A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. Nature 448:370–374.
48. Sarkar, S. F., J. S. Gordon, G. B. Martin, and D. S. Guttman. 2006. Comparative genomics of host-specific virulence in *Pseudomonas syringae*. Genetics 174:1041–1056.
49. Sarkar, S. F., and D. S. Guttman. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. Appl. Environ. Microbiol. 70:1999–2012.
50. Scally, M., E. L. Schuenzel, R. Stouthamer, and L. Nunney. 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. Appl. Environ. Microbiol. 71:8491–8499.
51. Shenge, K. C., R. B. Mabagala, C. N. Mortensen, D. Stephan, and K. Wydra. 2007. First report of bacterial speck of tomato caused by *Pseudomonas syringae* pv. tomato in Tanzania. Plant Dis. 91:462.
52. Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.
53. Smith, J. M., N. H. Smith, M. O’Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA 90:4384–4388.

54. Spratt, B. G., W. P. Hanage, and E. J. Feil. 2001. The relative contribution of recombination and point mutation to the diversification of bacterial clones. Curr. Opin. Microbiol. 4:602–606.

55. Swofford, D. L. 2003. PAUP* phylogenetic analysis using parsimony (and other methods), 4th ed. Sinauer Associates, Sunderland, MA.

56. Vinatzer, B. A., and C. T. Bull. The impact of genomics approaches on our understanding of diversity and taxonomy of plant pathogenic bacteria. In R. W. Jackson (ed.), Plant pathogenic bacteria: genomics and molecular biology, in press. Horizon Scientific Press, Norfolk, United Kingdom.

57. Whalen, M. C., R. W. Innes, A. F. Bent,and B. J. Staskawicz. 1991. Identificationof *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. Plant Cell 3:49–59.

58. Wiebe, W. L., and R. N. Campbell. 1993. Characterization of *Pseudomonas syringae* pv. maculicola and comparison with P. s. tomato. Plant Dis. 77:414–419.

59. Wildermuth, M. C., J. Dewdney, G. Wu, and F. M. Ausubel. 2001. Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature 414:562–565.

60. Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. Maiden, H. Ochman, and M. Achtman. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol. 60:1136–1151.

61. Yang, Z. 1997. A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

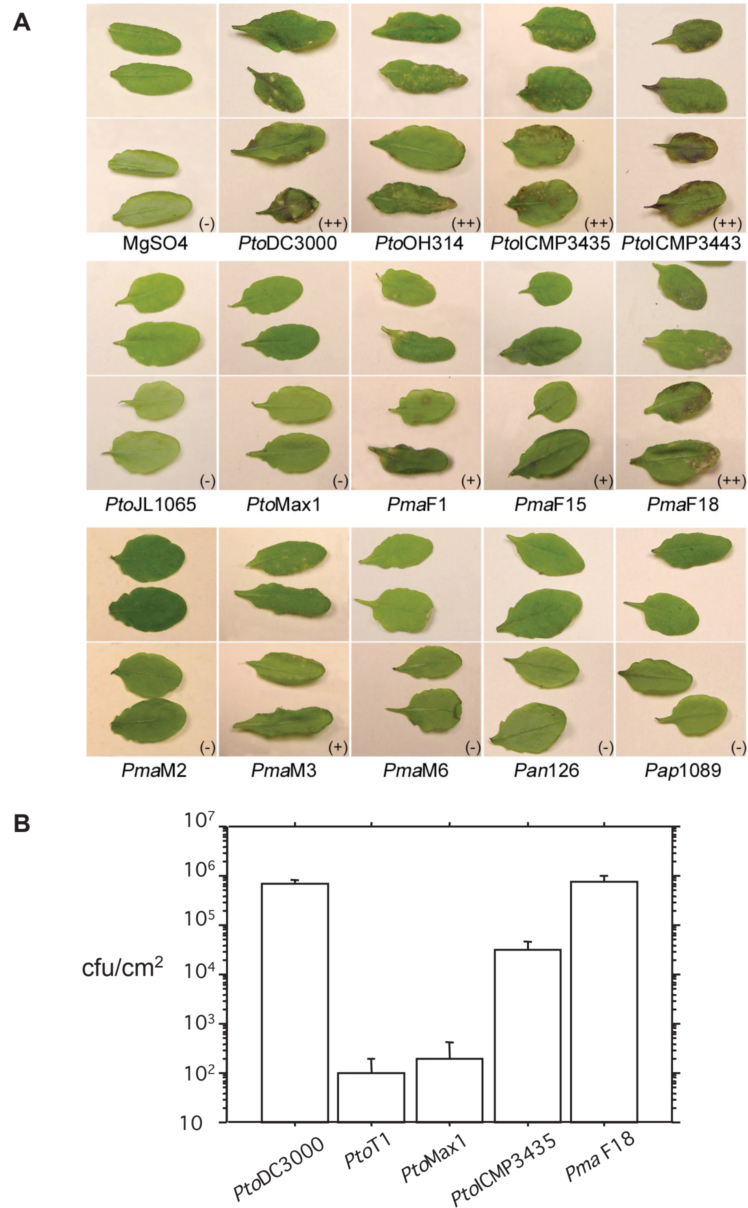
62. Zaccardelli, M., A. Spasiano, C. Bazzi, and M. Merighi. 2005. Identification and in planta detection of *Pseudomonas syringae* pv. tomato using PCR amplification of *hrpZ*$_{Pst}$. Eur. J. Plant Pathol. 111:85–90.

63. Zhao, Y., J. P. Damicone, D. H. Demezas, V. Rangaswamy, and C. L. Bender. 2000. Bacterial leaf spot of leafy crucifers in Oklahoma caused by *Pseudomonas syringae* pv. *maculicola*. Plant Dis. 84:1015–1020.

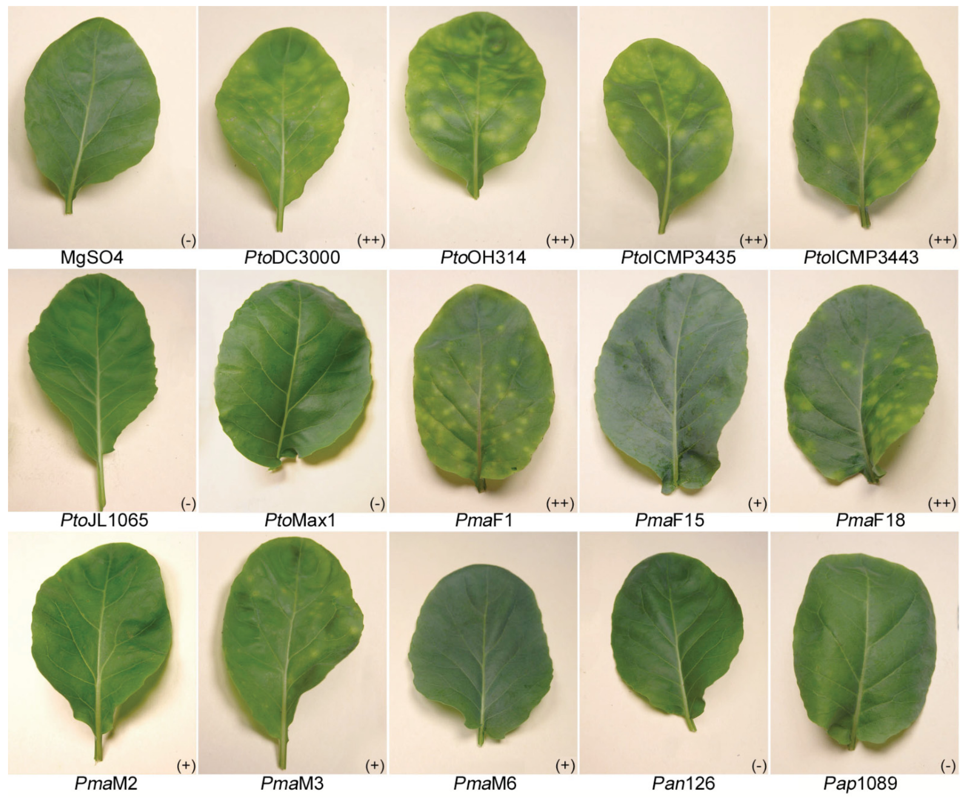
64. Zhou, N., T. L. Tootle, F. Tsui, D. F. Klessig, and J. Glazebrook. 1998. *PAD4* functions upstream from salicylic acid to control defense responses in Arabidopsis. Plant Cell 10:1021–1030.

h. Supplementary Materials

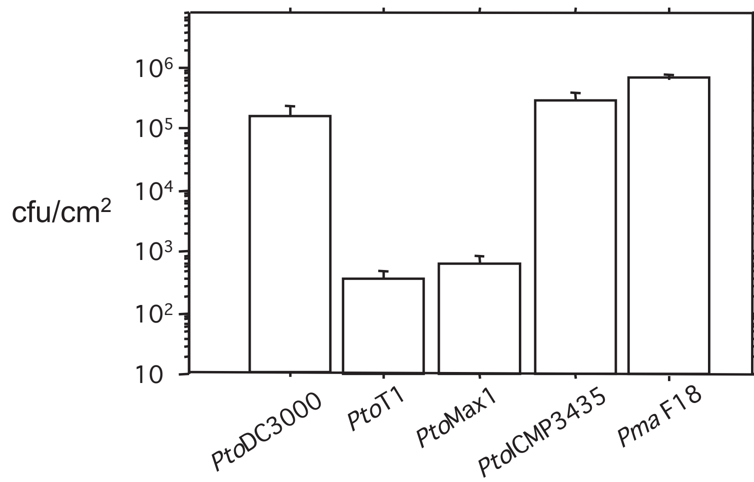
Pictures of infected leaves and population sizes of bacteria in leaves (Fig. S1).



C



D



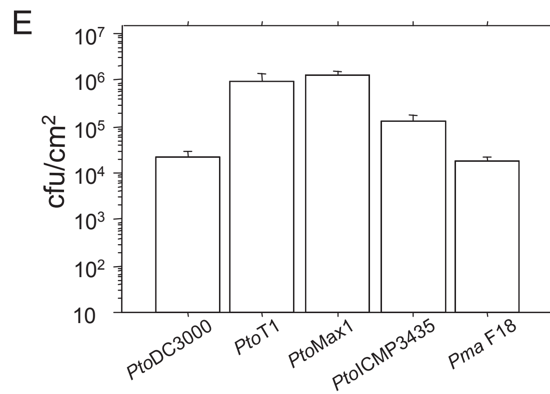
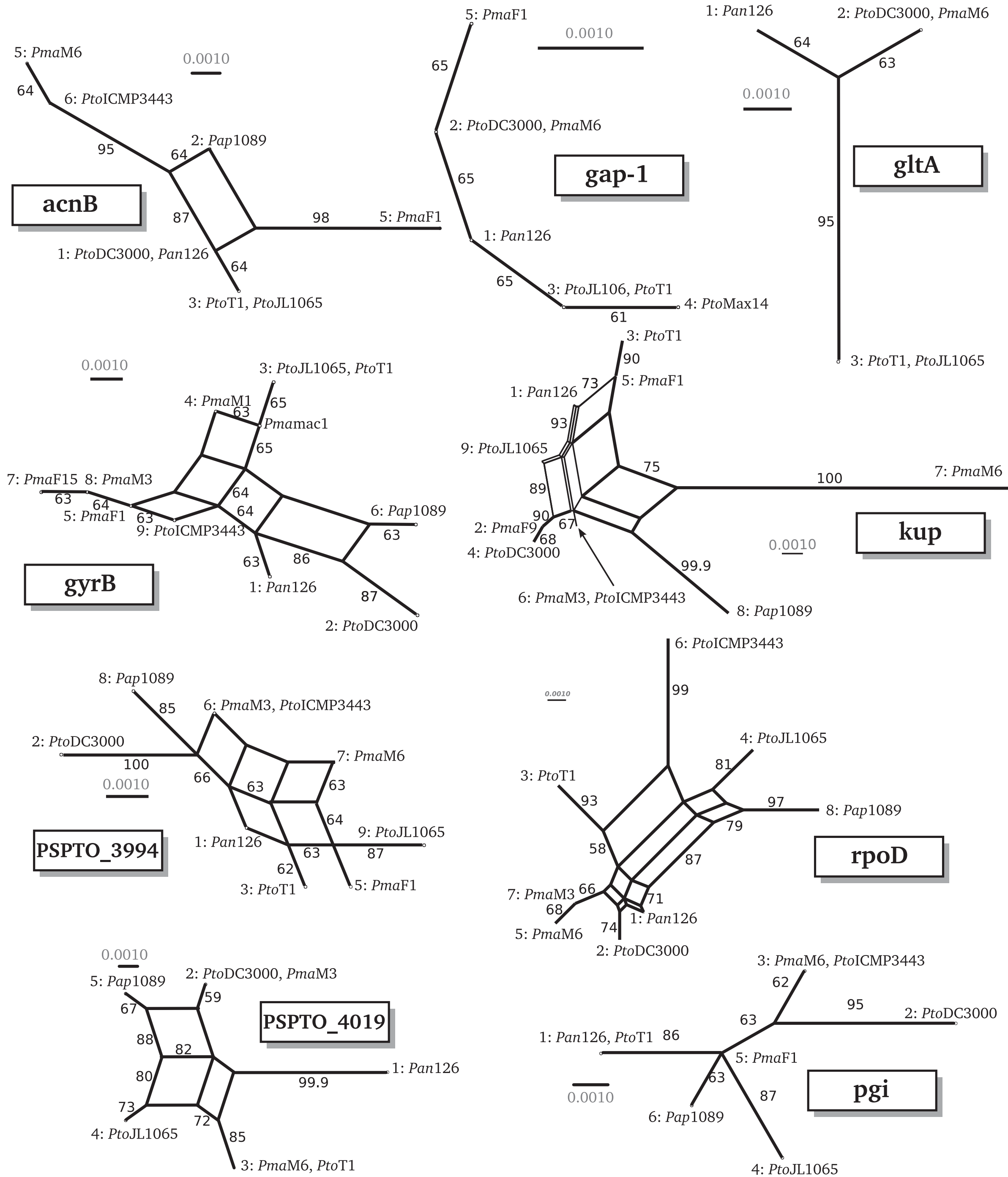


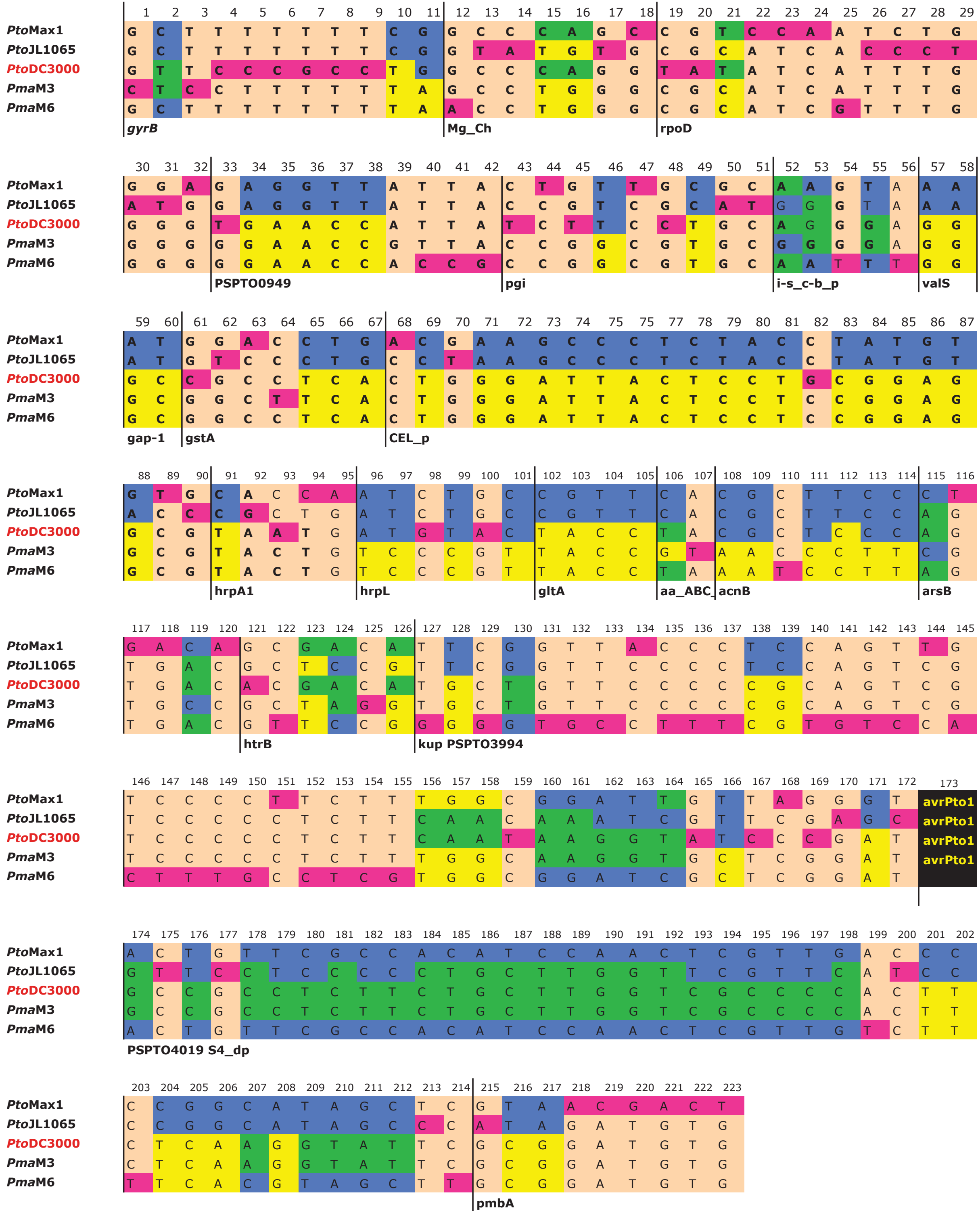
Figure S1. Virulence tests of representative *P. syringae* isolates on *A. thaliana*, cauliflower, and tomato.

A. thaliana ecotype 'Columbia' *rps2* plants, and cauliflower (*Brassica oleracea* var. *botrytis)* cultivar 'Early Snowball A*'* plants were sprayed with bacterial suspensions of *P. syringae* isolates. **A** and **C:** Photographs of representative leaves were taken one week later. Severe symptoms are indicated by '++'. Weak symptoms are indicated by '+'. Leaves with no symptoms are indicated by '-'. **B**, **D,** and **E**: Populations of representative *Pto* and *Pma* isolates were counted in four infected leaf areas/isolate three days after infection of *A. thaliana*, cauliflower, and tomato respectively.

Phylogenetic networks (Fig. S2).



DNA alignments (Fig. S3).



Primer sequences (Table S1).

Supplementary Table 1: Genes, genome locations, primers, length of products

PID	name		position in *Pto*DC3000 genome	forward primer	reverse primer	length of analyzed sequence
28867247	DNA gyrase subunit B	gyrB	4147..6564	TGCVTTCGTTGARTACCTGA	ACGGAAGAAGAAGGTSAGCA	701
na	Magnesium chelatase pseudogene	Mg_Ch	81732..82382	ACYTGCTCCAGATCCGCC	GCTGGAGGTTYTRCGTGAGCC	310
28867765	RNA polymerase sigma factor	rpoD	588846..590696	GAAGGCATCCGTGAAGTGAT[1]	GCCACGGTTGGTGTACTTCT	645
28868169	hypothetical protein	PSPTO0949	1026472..1028727	AYTGACCGCSATYGTCGAG	AATGCGCTGAAACGCGCAG	674
28868179	glucose-6-phosphate isomerase	pgi	1040763..1042427	GCGTACTACCGYAMYCCB TC	CCACATMGGRAARATRTTYT	569
28868192	iron-sulfur cluster-binding protein	PSPTO0972	1057534..1057866	AGCGAYRCCCAGATBCCC	YGCCRTGMGGGAARTTGAC	620
28868479	valyl-tRNA synthetase	valS	1389375..1392221	CCAGATACCAGTCGCAATACTG	TGGACCTTCTCCACGCTC	588
28868497	glyceraldehyde-3-phosphate dehydro	gap-1	1415258..1416259	CGTATCGCAATCAACGGTTT	GACTCTCCGTATCGCAATCA	600
28868573	glutathione S-transferase	gstA	1502901..1503527	GCGACCCGAAAASCAAYC	RTAYTCVCCGCCCTGCTC	506
28868577	hopAA1-1 effector plus up-stream re	hopAA1-1	1504948..1505427	ACTCCYTTGCTGGKTTGGAA	KGTCGARCTTYTCCGCCTTG	677
28868585	type III effector avrE1$_{PtoDC3000}$	avrE1	1514116..1519503	TTGTCGTTYTTCTGGGTCAR	CMGAYATGGGCTTTACCGAT	636
28868589	type III helper protein HrpA1	hrpA1	1524281..1524622	GCCTGGARCTSGAYRTSCCAC	CAGAVGCRCCRAARTTGTCRCC	753
28868612	RNA polymerase sigma factor HrpL	hrpL	1542838..1543392	GATCCGYAATCACTTCCG	TCCTGATAATTGCCRTCCA	205
28869396	citrate synthase	gltA	2414332..2415621	WYTRACCGGYACMGTBGGY	TGGGCTGATSGGYTTRATYT	507
28869961	amino acid ABC transporter	PSPTO2775	3094410..3095261	GGMGCRGACGGMAAGTTG	CGGAAAACCGATTGCMGG	589
28870907	aconitate hydratase	acnB	4239241..4241841	TGATGTTTGATGCCTTCCAC	TAAAACCCTTGGTGCTTTCG	650
28870953	arsenical pump membrane protein	arsB	4300065..4301345	TCAGYACGATRCCGACCTTG	ACATCGTKTCGGCGGAYTAC	1094
28871024	lipid A biosynthesis lauroyl acyltransf	htrB	4383469..4384401	ATACTGTTCCGGGCATTCTG	CTTCTACAGGCATCGCCTTC	526
28871138	potassium uptake protein	kup	4503214..4505109	CTGATCTTCTGGTCGCTGATC	TTCCAGGTGGTCATGAGGA	1943 [1]
28871137	hypothetical protein	PSPTO3992	4505431..4506519	AYACCACCATCATCATTTGC	CCGTACAGMAAMGGYACGAC	1943 [1]
28871145	type III effector avrPto1$_{PtoDC3000}$	avrPto1	4514766..4515260	GTCATKGCAGTGGTTATCRCAGA	GTYTGCAGGCTCCKAAAAACTG	various[2]
28871162	hypothetical protein	PSPTO4019	4530408..4531739	AAGGTTACGAAGTCGTTGCG	CTCGCGCCAGTTCTTGAT	720
28871164	S4 domain protein	PSPTO4021	4532376..4533086	GAGCTGGTCGGTTGCTCC	GGCGTGTTTCCTGATAAGTCG	1837
28871598	pmbA protein	pmbA	5024076..5025422	GCTTGCGRCCGGAATAVG	GCGCTGATGGCCAAAGAG	778

1 F primer starts within Kup and PCR product encompasses intergenic region
2 primers are locoated in the genes up- and downstream of the *avrPto1* gene. The length of the product is different in different strains.

Chapter 3

Genome Sequencing of *Pseudomonas syringae* strains and Identification of Genomic Differences

Title: A Draft Genome Sequence of *Pseudomonas syringae* pv. *tomato* T1 Reveals a Type III Effector Repertoire Significantly Divergent from That of *Pseudomonas syringae* pv. *tomato* DC3000

Authors and author affiliations:

Nalvo F. Almeida,[1, 7] Shuangchun Yan,[2] Magdalen Lindeberg,[3] David J. Studholme,[4] David J. Schneider,[5] Bradford Condon,[3] Haijie Liu,[2] Carlos J. Viana,[7] Andrew Warren,[1] Clive Evans,[1] Eric Kemen,[4] Dan MacLean,[4] Aurelie Angot,[3, 6] Gregory B. Martin,[3, 6] Jonathan D. Jones,[4] Alan Collmer,[3] Joao C. Setubal,[1, 8] and Boris A. Vinatzer[2]

[1]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, U.S.A.; [2]Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Latham Hall, Ag Quad Lane, Blacksburg, VA 24061, U.S.A.; [3]Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, U.S.A. [4]Sainsbury Laboratory, Norwich NR4 7UH, U.K.; [5]U. S. Department of Agriculture Agricultural Research Service, Ithaca, NY 14853, U.S.A.; [6]Boyce Thompson Institute for Plant Research, Ithaca, NY 14853, U.S.A.; [7]Department of Computing and Statistics, Federal University of Mato Grosso do Sul, Brazil; [8]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, U.S.A.

Submitted 11 July 2008. Accepted 10 September 2008.

Footnote:

Nalvo F. Almeida, Shuangchun Yan, and Magdalen Lindeberg contributed equally to this work.

Corresponding author: Boris A. Vinatzer; E-mail: vinatzer@vt.edu

Nucleotide sequence data has been deposited as a Whole Genome Shotgun (WGS) project in the DDBJ/EMBL/GenBank database under project accession ABSM00000000. The version described in this paper is the first version, ABSM01000000.

a. Abstract

Diverse gene products including phytotoxins, pathogen-associated molecular patterns, and type III secreted effectors influence interactions between *Pseudomonas syringae* strains and plants, with additional yet uncharacterized factors likely contributing as well. Of particular interest are those interactions governing pathogen-host specificity. Comparative genomics of closely related pathogens with different host specificity represents an excellent approach for identification of genes contributing to host-range determination. A draft genome sequence of *Pseudomonas syringae* pv. *tomato* T1, which is pathogenic on tomato but non-pathogenic on *Arabidopsis thaliana*, was obtained for this purpose and compared with the genome of the closely related *A. thaliana* and tomato model pathogen *P. syringae* pv. *tomato* DC3000. Although the overall genetic content of each of the two genomes appears to be highly similar, the repertoire of effectors was found to diverge significantly. Several *P. syringae* pv. *tomato* T1 effectors absent from strain DC3000 were confirmed to be translocated into plants, with the well-studied effector AvrRpt2 representing a likely candidate for host-range determination. However, the presence of *avrRpt2* was not found sufficient to explain *A. thaliana* resistance to *P. syringae* pv. *tomato* T1, suggesting that other effectors and possibly type III secretion system–independent factors also play a role in this interaction.

Additional keywords: effector-triggered immunity, nonhost resistance, pyrosequencing, Solexa sequencing.

b. Introduction

Many closely related plant pathogens, though adapted to fundamentally similar plant-associated niches, exhibit striking differences in the species and cultivars that they can infect. Host ranges of plant pathogens are governed by a complex network of interactions between plant defense mechanisms and pathogen gene products, with current models of plant defense emphasizing the roles of pathogen (or microbe) associated molecular pattern (PAMP)–triggered immunity (PTI) and effector-triggered immunity (ETI) (Jones and Dangl 2006).

Also referred to as basal resistance, PTI bears significant resemblance to the basal immunity exhibited by animals against various pathogens (Angot et al. 2007; Bhavsar et al. 2007; Zipfel 2008). PTI is typically triggered when the plant in question detects conserved elements of the microbial structure, such as flagellin (Felix et al. 1999), elongation factor Tu (Kunze et al. 2004), or lipopolysaccharides (Newman et al. 2002). Detection gives rise to a variety of defense-related phenotypes, including callose formation and upregulation of defense-related proteins, ultimately resulting in host- range restriction for a wide variety of microbes (Ham et al. 2007).

In contrast to the broad spectrum immunity conferred by recognition of the highly conserved PAMPs, ETI, also known as “gene-for-gene” resistance, involves detection of the more variable type III–secreted effector (T3E) proteins by plant resistance (R) gene products (Chisholm et al. 2006). Detection of effectors can occur directly or indirectly via the plant’s detection of alterations brought about by effector actions and frequently results in the induction of programmed cell death (PCD) (Dangl and Jones 2001). The fact that individual effectors are generally more dispensable than PAMPs coupled with heavy selective pressure on the pathogen and host to alternately evade and achieve detection can lead to extensive

variation in the repertoires of effectors and R genes among otherwise similar pathogens and hosts (Rohmer et al. 2004; Rose et al. 2007). Though ETI has been generally regarded as the basis for "host resistance" exhibited by individual plant cultivars and ecotypes to pathogen pathovars and races, evidence is accumulating that ETI can also play a critical role in "nonhost resistance" exhibited by all cultivars and ecotypes of a plant species against all strains of a pathogen. For example, HopQ1-1 was recently shown to be the determining factor in interfering with pathogenicity of *P. syringae* pv. tomato DC3000 on *Nicotiana benthamiana* (Wei et al. 2007), and recognition of the effectors AvrPto1 (AvrPto) and HopAB2 (AvrPtoB) by the corresponding *R* genes appears to exclude tomato from the host range of several *P. syringae* strains (Lin and Martin 2007).

For many nonhost plant-pathogen interactions, the ultimate outcome is shaped by the interplay of multiple defense pathways and gene products. For example, resistance of *Arabidopsis thaliana* to the fungal wheat pathogen *Blumeria graminis* appears to involve PTI and ETI. During a prehaustorial phase, formation of papillae and callose layers typical of PTI is observed. During the posthaustorial phase, an induction of the hypersensitive response (HR) and consequent PCD typical of ETI is observed (Lipka et al. 2005).

Among bacterial pathogens, the outcomes of nonhost plant interactions may similarly involve diverse pathways, with T3E proteins involved at many levels. In fact, in addition to limiting host range through ETI, effectors also have the capacity to extend host range by suppressing defense pathways, including those implicated in PTI and ETI. In DC3000, several effectors as well as the toxin coronatine confer the ability to suppress induction of the PAMP-induced *A. thaliana* NHO1 gene implicated in nonhost resistance (Li et al. 2005). *P. syringae* pv. *phaseolicola* 3121, which lacks coronatine and most of these effectors, is unable to suppress *NHO1* expression, likely contributing to its inability to cause disease on *A.*

thaliana. Further evidence for the complexity of *A. thaliana* resistance to strain 3121 was uncovered by Ham and colleagues (2007), who found that strain 3121 elicits several parallel defense pathways in *A. thaliana*, with one of them being suppressed by the DC3000 effector HopM1.

To date, the genomes of three *Pseudomonas syringae* strains have been sequenced (Buell et al. 2003; Feil et al. 2005; Joardar et al. 2005) with bioinformatic analyses leading to the identification of over 40 T3E (Lindeberg et al. 2008). Of these strains, the tomato isolate DC3000 has been most extensively studied as a model pathogen of *A. thaliana* and tomato. Multilocus sequence typing (MLST) (Maiden et al. 1998) has revealed that DC3000 belongs to a group of *P. syringae* isolates found on members of the *Solanaceae* and *Brassicaceae* families that is characterized by the ability to cause disease on both *A. thaliana* and tomato (Yan et al. 2008). Other isolates, though closely related to DC3000, have been found exclusively on tomato and cause disease only on tomato but not on *A. thaliana* (Yan et al. 2008). The isolate T1, whose inability to cause disease on *A. thaliana* was first described by Whalen and colleagues (Whalen et al. 1991), is a representative of this latter group of typical *P. syringae* pv. *tomato* strains.

To investigate the genetic basis of T1's inability to cause disease on *A. thaliana*, a deep-coverage draft genome sequence of T1 was obtained using two high-throughput DNA sequencing technologies, pyrosequencing (Margulies et al. 2005), and Solexa sequencing (Bentley 2006). The draft genome sequence was annotated and analyzed for the presence of conserved and potentially novel T3E, for the purpose of identifying genetic differences that account for the differing abilities of DC3000 and T1 to infect *A. thaliana*. Although the two genomes were found to align for most of their length, their effector repertoires were found to be highly divergent. As a first step in dissecting *A. thaliana* resistance response to T1, the role of ETI in this response was investigated.

c. Results

T1 is a nonpathogen of *A. thaliana* ecotypes and causes a HR.

Previously reported comparisons of disease symptoms and bacterial growth between *A. thaliana* and tomato strain DC3000, the typical *P. syringae* pv. *tomato* isolate T1, and other *P. syringae* isolates closely related to T1 revealed significant differences in host range. Specifically, the T1 population size on the *A. thaliana* ecotype 'Columbia' was shown to be approximately 4 log lower than that of DC3000 at 3 days post-infection and, in contrast to DC3000, T1 did not cause any disease symptoms (Yan et al. 2008). To evaluate whether this inability to multiply and to cause disease on *A. thaliana* is ecotype-specific or more widespread, comparisons of growth and symptom production were extended to three additional *A. thaliana* ecotypes and a type III secretion system (T3SS)-deficient derivative of DC3000 missing the *hrcC* gene (DC3000 CUCPB5112 [Penaloza-Vazquez et al. 2000]) was included as a control. As shown in Figure 1, T1 only grows to a population size equivalent to the T3SS-deficient DC3000 strain on all four *A. thaliana* ecotypes and does not cause any disease symptoms. Similar results were obtained with additional *P. syringae* pv. *tomato* isolates that are very closely related to T1, based on MLST (Yan et al. 2008), and additional *A. thaliana* ecotypes (data not shown). Therefore, *A. thaliana* resistance to T1-like isolates is not limited to a single ecotype-strain combination, suggesting that *A. thaliana* might be generally resistant to typical *P. syringae* pv. *tomato* strains.

While plant and bacterial factors contributing to *A. thaliana* resistance to T1 await characterization, infection of ecotype 'Columbia' with a high dose of T1 induces a HR, suggesting that ETI may contribute to the resistance exhibited by *A.*

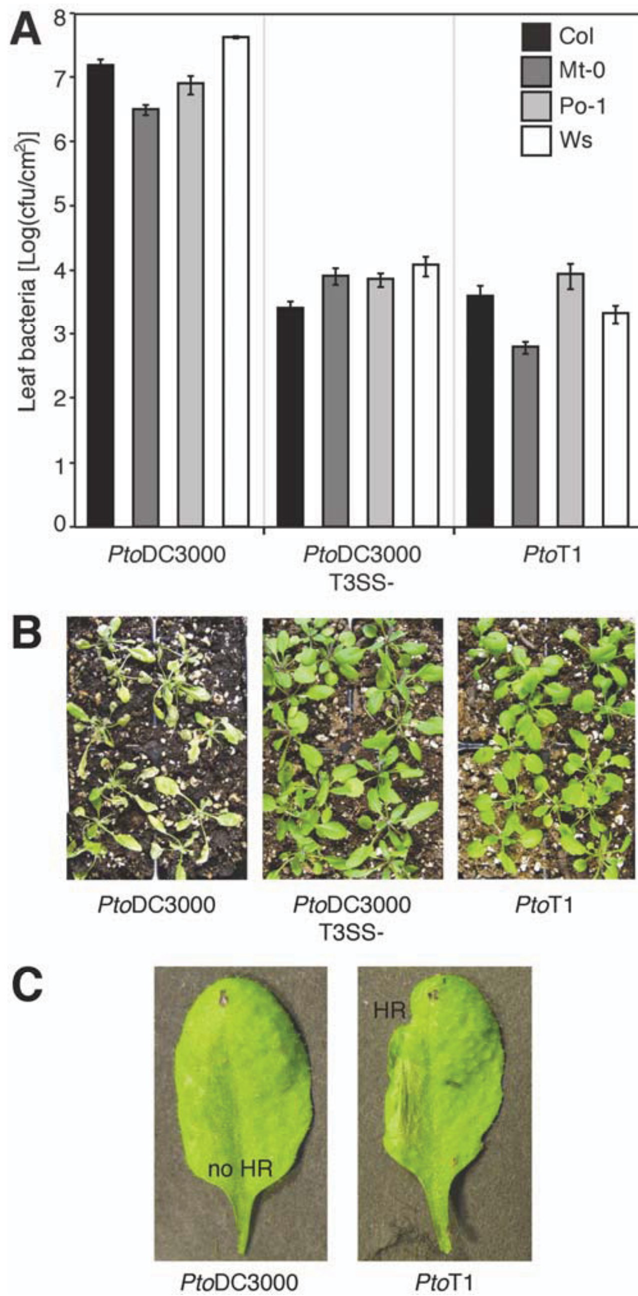


Fig. 1. Comparison of in planta growth and symptoms of the *Arabidopsis thaliana* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000, a type III secretion system–deficient mutant of DC3000 (DC3000 T3SS–), and the tomato pathogen *P. syringae* pv. *tomato* T1. **A,** Bacteria were sprayed onto *A. thaliana* ecotypes Columbia (Col-0), Mt-0, Po-1, and WS at an optical density at 600 nm (OD_{600}) of 0.1. At 3 days postinfection, CFU per square centimeter were determined. While population sizes of DC3000 were significantly larger than those of DC3000 T3SS– and T1 on all ecotypes ($P < 0.0001$), sizes of DC3000 T3SS– and T1 were not significantly different from each other on any of the ecotypes, based on Student's *t*-test ($P > 0.1$). **B,** *A. thaliana* Col-0 plants were photographed a week postinfection. **C,** Bacteria were infiltrated into leaves of *A. thaliana* Col-0 at an OD_{600} of 0.3, and pictures were taken after 16 h.

thaliana to the T1-like isolates (Fig. 1C). This is in agreement with the earlier observation (Whalen et al. 1991) that cosmids containing T1 genome fragments when expressed in DC3000 reduced virulence on *A. thaliana*.

Table 1. Comparison of the *Pseudomonas syringae* pv. *tomato* T1 draft genome sequences obtained by 454 pyrosequencing and Solexa sequencing

Sequence characteristics	454	Solexa
Number of reads	519,729	15,723,622
Mean length of reads (nt)	241	36
Total length of reads (nt)	125,281,22	566,050,392
Number of assembled contigs	134	3,825
Mean length of contigs (nt)	45,878	1,570
Length of longest contig (nt)	335,36	16,815
Total length of contigs (nt)	6,147,659	6,005,250
Calculated genome coverage	20.4	94

T1 draft genome sequence obtained using Solexa sequencing and 454 pyrosequencing and compared with the DC3000 genome.

To initiate identification of the genetic bases for the differing virulence of T1 and DC3000 on *A. thaliana*, a draft genome sequence of *P. syringae* pv. *tomato* T1 was obtained using 454 pyrosequencing (Margulies et al. 2005) and Solexa sequencing (Bentley 2006). Reads were assembled separately into draft genome sequences. Characteristics of the two drafts are compared in Table 1. Based on the total contig length of the two draft genomes, the T1 genome was sequenced to a 20-fold coverage using 454 pyrosequencing and to a 94-fold coverage with Solexa sequencing. It is obvious that, although the Solexa genome coverage was more than four times greater than that of the 454 genome sequence, the longer pyrosequencing reads allowed a much more extensive assembly. To compare the quality of the two draft genome sequences, the two genomes were aligned and their

average DNA identity was determined. The two drafts were found to be over 99.998% identical (1 difference every 100,000 bp) and are, thus, both of excellent quality. Moreover, all Solexa genome sequences were contained in the 454 genome sequence. Therefore, most of the subsequent analyses were conducted using the 454 draft genome sequence.

Given that the T1 454 draft genome sequence consists of 134 contigs, there are at least 134 gaps in the genome sequence. Most contig ends consist of repeated sequences, which is the reason why further assembly was not possible with current data. However, given the estimated genome coverage, we believe that the vast majority of the critical virulence genes, including effectors, are identifiable from the current draft.

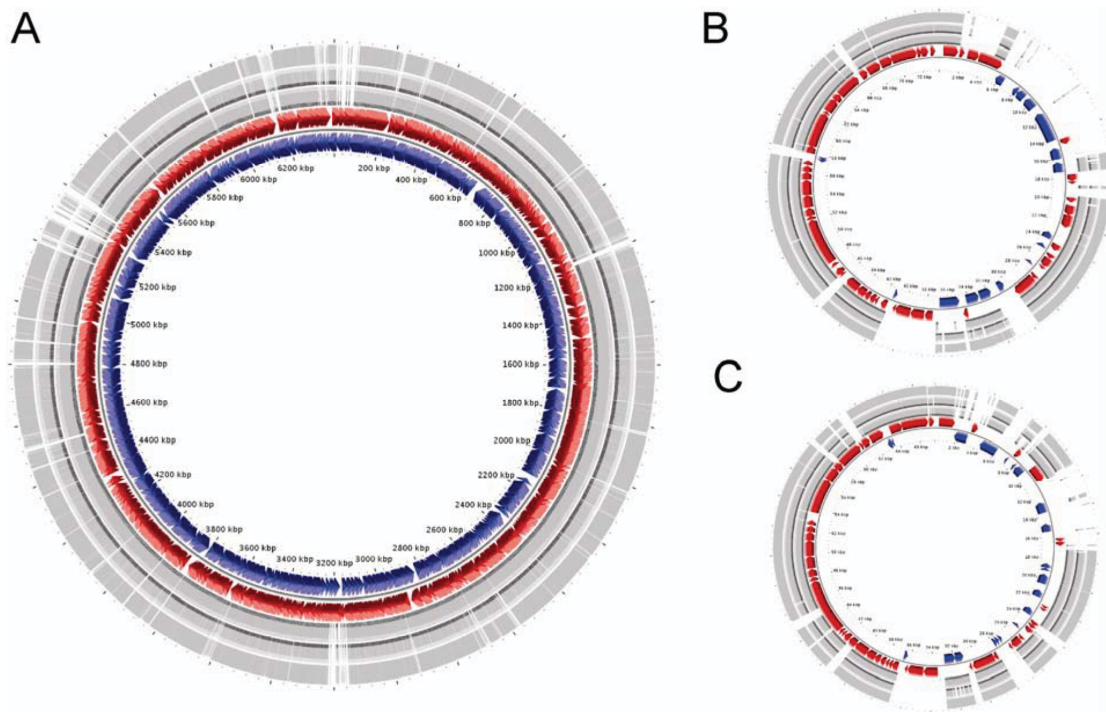


Fig. 2. Pairwise alignments between the *Pseudomonas syringae* pv. *tomato* T1 454 and Solexa draft genome sequences and the *P. syringae* pv. *tomato* DC3000 genome. **A,** Alignment of the T1 draft genome sequences with the DC3000 chromosome using the latter as reference. **B** and **C,** Alignment of the T1 draft genome sequences with the DC3000 plasmids pDC3000A and pDC3000B, respectively. Genes in the reference genome are in blue and red. Plotted Going outwards in each panel as follows: alignment with T1 Solexa contigs, alignment with T1 454 contigs, and single nucleotide polymorphism based on the Solexa consensus sequence. Closeups of local alignments corresponding to the type III–secreted effectors can be viewed on the PtoT1 genome website.

Alignment of the T1 genome with the three sequenced *P. syringae* genomes

To investigate differences between T1 and the other sequenced *P. syringae* strains, the T1 draft genome was aligned with the closed genomes of DC3000 (Buell et al. 2003), strain 1448A (Joardar et al. 2005), and *P. syringae* pv. *syringae* B728a (Feil et al. 2005).

The T1 genome was first compared with the DC3000 genome using the latter as reference. Figure 2 shows alignments of the DC3000 chromosome and its two plasmids pDC3000A and pDC3000B (panels A, B, and C, respectively) with both the 454 and the Solexa T1 draft genome sequences. It is evident from these figures that the DC3000 and the T1 genomes align over most of their length. Moreover, the alignments with the 454 and the Solexa T1 drafts are nearly identical, confirming the high quality of both genome sequences. As expected, alignment of the T1 draft genome sequence with the completely assembled B728a and 1448A genome sequences revealed significant divergence relative to the alignment of the two pathovar tomato strains. While 5,618,731 nt of DNA align between T1 and DC3000, only 4,547,158 nt of T1 align with the B728a genome and 4,689,063 nt align with the 1448A genome.

Comparison of protein complements between the four *P. syringae* genomes.

In order to identify the repertoire of putative virulence genes in the T1 genome and to reveal differences between the DC3000 and T1 genomes that underlie differences in host range, an annotation of the T1 draft genome was generated. The annotation can be accessed at the DDBJ/EMBL/GenBank database as part of WGS project ABSM00000000 (version ABSM01000000). A total of

5,733 putative genes were predicted in this preliminary annotation, of which 3,987 (69%) are similar to genes that encode proteins with known function in other bacterial genomes, 729 (13%) are similar to genes that encode proteins with unknown function in other genomes ("conserved hypothetical proteins"), and 1,047 (18%) putative genes are unique to T1 ("hypothetical proteins"). However, it should be noted that estimates of the numbers of repeated elements such as mobile genetic elements are less reliable owing to the unclosed nature of the genome.

The predicted proteins from T1 were compared with those of the other sequenced *P. syringae* genomes. A Venn diagram illustrating the number of predicted proteins shared between and unique to T1 and DC3000 relative to the B728a and 1448A protein repertoires is shown in Figure 3. In summary, 4,271 proteins are shared between all four genomes and thus represent conserved housekeeping gene products and virulence gene products most likely underlying fundamental adaptations to the plant-associated niche. A total of 471 proteins are only shared between DC3000 and T1 and are absent from B728a and 1448A and may include proteins with roles in pathogenicity on the shared host tomato. A total of 757 proteins are unique to T1. Of these, 511 are hypothetical proteins, 79 proteins can be classified as mobile genetic elements (transposons, insertion sequences, bacteriophage components, and other genes with predicted function in DNA replication), and 167 are proteins with other predicted functions. This latter category represents the most interesting class of genes, with potential involvement in host specificity through either elicitation of *A. thaliana* defenses or contribution to virulence on tomato. DC3000 contains a similar number of gene products (196) that are unique among the four strains and have some functional annotation, which may account for the host range extension of DC3000 to *A. thaliana*. The hypothetical proteins unique to T1 or DC3000 may be found to also encode host

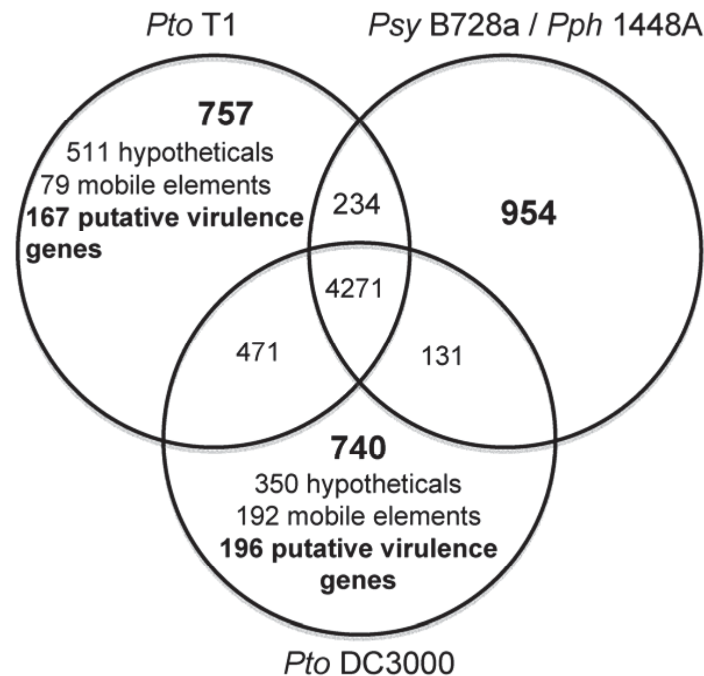


Fig. 3. Venn Diagram comparing the protein repertoire of *Pseudomonas syringae* pv. *tomato* (*Pto*) T1 with that of the completely sequenced genomes of *P. syringae* pv. *tomato* DC3000, *Pseudomonas syringae* pv. *syringae* (*Psy*) B728a, and *P. syringae* pv. *phaseolicola* (*Pph*) 1448A. Proteins from the B728a and 1448A genomes were combined. Numbers of predicted proteins for each section are indicated. An interactive version of this figure in which the different sections of the diagram are hyperlinked to corresponding lists of proteins are available on the Vinatzer lab's website.

specificity factors, if additional analyses are able to demonstrate which ones indeed code for expressed proteins.

A comparison of the predicted protein repertoire of T1 with all finished *Pseudomonas* genomes was also performed. Results can be queried in the supplemental data section at Virginia Tech's PtoT1 genome website.

Prediction of the T3SS effector repertoire of T1 and comparison with the DC3000 effector repertoire.

Comparison of major virulence factors identified in DC3000 (Lindeberg et al. 2008) with those encoded by T1 reveals that all major classes of virulence proteins appear to be present, with the exception of those encoding the toxin coronatine and a subset of T3E. To investigate the hypothesis that T3E are

primarily responsible for the differences in host range observed between T1 and DC3000, a thorough analysis of T3E in T1 was conducted. The list of all known P. syringae effectors maintained at the Pseudomonas-Plant Interaction website was compared against the T1 genome using BLAST analysis. A subset of the T1 effector genes lacking orthologs in DC3000 (*avrA*, *avrRpt2*, *hopAS1*, and *hopAE1*) were cloned and expressed as Cya fusion proteins in DC3000, and translocation was confirmed as previously described (Schechter et al. 2004). The putative T1 effector genes *hopW1* and *avrD* were not tested for translocation, since these two genes are probably not contributing to *A. thaliana* resistance to T1; the effector gene *hopW1* is present in *P. syringae* pv. *maculicola* ES4326, well known to cause disease on many *A. thaliana* ecotypes, including Col-0, and *avrD* was shown not to reduce bacterial growth of ES4326 in *A. thaliana* (Wanner et al. 1993).

To identify those effector genes likely to be expressed and to identify any novel effectors that might be present in the T1 genome, a list of predicted binding sites for the HrpL alternate sigma factor was generated for the T1 genome as previously described (Fouts et al. 2002; Lindeberg et al. 2006; Vencato et al. 2006), and the candidate HrpL-regulated genes were analyzed for the presence of conserved N-terminal motifs consistent with type III secretion. Among the candidates with *hrp* promoters and plausible translocation motifs were T1 orthologs of the D3000 exchangeable effector locus open reading frame (ORF)2 gene product and of the *P. syringae* pv. *syringae* B728a gene encoding Psy_0737, though neither was found to be translocated. Cya test results are shown in Table 2. A list of the T1 *hrp* regulon and putative effector genes is shown in Table 3, together with the verified DC3000 *hrp* regulon and effector repertoire. T1 effectors have been added to the effector database at the Pseudomonas-Plant Interaction website.

Comparison of the effector repertoires for T1 and DC3000 reveals surprising differences despite the otherwise high degree of similarity and shared tomato host. Of the effectors that are likely to be deployed, as determined by the presence of an *hrp* box and intact coding region, only 14 are shared between the two, while as many as 15 are present only in the DC3000 genome and 11 only in the T1 genome. With regard to some of the best-studied T3SS effectors, T1 alone deploys AvrRpt2 (with only two amino-acid differences when compared with AvrRpt2Pto JL1065, the allele originally cloned by R. Innes, [Innes et al. 1993]), while DC3000 alone produces AvrPto, HopU1, HopX1, and two HopAM1 paralogs. HopX1 is of particular interest, given that this gene is conserved in 1448a, B728a, and at least ten other *P. syringae* pathovars (Charity et al. 2003; Deng et al. 2003). HopM1 also appears to be deployed by DC300 alone, as the presence of a frameshift in the T1 ortholog reduces the likelihood of functionality. A homolog of HopAB2 (AvrPtoB) is encoded by both genomes but shows evidence of extensive divergence relative to flanking DNA, indicative of high selection pressure and of

Table 2. In planta adenylate cyclase activity in *Nicotiana benthamiana* of *Pseudomonas syringae* pv. *tomato* T1 effector candidate-Cya fusion proteins

Cya fusion protein	pmol cAMP/μg protein	
	DC3000 (T3SS+)[a]	5112 (T3SS–)
AvrRpt2 (PSPTOT1_2469)	371.3 ± 47.8	19.1 ± 6.7
AvrA (PSPTOT1_4933)	191.3 ± 70.8	0.6 ± 0.2
HopAE (PSPTOT1_0516)	217.2 ± 17.8	0.7 ± 0.2
HopAS1 (PSPTOT1_1672)[b]	136.3 ± 11.6	0.6 ± 0.2
PSPTOT1_0854	15.3 ± 5.4	16.1 ± 5.2
EELORF2 (PSPTOT1_4759)	11.3 ± 3.2	10.7 ± 4.0

[a] T3SS = type III secretion system.
[b] Only the first 1 kb of coding sequence was tested.

Table 3. Repertoires of type III secretion system (T3SS) effectors and T3SS-independent genes in *Pseudomonas syringae* pv. *tomato* T1 compared with *P. syringae* pv. *tomato* DC3000[a]

DC3000	Locus tag	hrp Box	T1	Locus tag	hrp box
T3E orthologs deployed by both DC3000 and T1					
avrE1	PSPTO_1377	Yes	avrE1	PSPTOT1_0672	Yes
avrPtoB (hopAB2)	PSPTO_3087	Yes	hopAB3	PSPTOT1_1462	Yes
hopC1	PSPTO_0589	Yes	hopC1-1	PSPTOT1_1097	Yes
hopD1	PSPTO_0876	Yes	hopD1	PSPTOT1_2827	Yes
hopF2	PSPTO_0502	Yes (operon)	hopF2	PSPTOT1_0639	Yes (operon)
hopH1	PSPTO_0588	Yes	hopH1	PSPTOT1_1413	Yes
hopI1	PSPTO_4776	Yes	hopI1	PSPTOT1_0569	Yes
hopO1-1	PSPTO_A0018	Yes	hopO1-1	PSPTOT1_4763	Yes (operon)
hopQ1-1	PSPTO_0877	Yes	hopQ1-1	PSPTOT1_2826	Yes
hopR1	PSPTO_0883	Yes	hopR1	PSPTOT1_5535	Yes
hopT1-1	PSPTO_A0019	Yes (operon)	hopT1-1	PSPTOT1_4762	Yes (operon)
hopY1	PSPTO_0061	Yes	hopY1	PSPTOT1_1266	Yes
hopAA1-1	PSPTO_1372	Yes	hopAA1	PSPTOT1_0677	Yes
hopAF1	PSPTO_1568	Yes	hopAF1	PSPTOT1_4398	Yes
T3E orthologs deployed by DC3000 alone					
hopA1	PSPTO_5354	Yes (operon)	hopA1′	PSPTOT1_2530	Yes (operon)
hopB1	PSPTO_1406	Yes (operon)	hopB1′	PSPTOT1_0643	Yes (operon)
hopM1	PSPTO_1375	Yes (operon)	hopM1′	PSPTOT1_0674	Yes (operon)
avrPto	PSPTO_4001	Yes	Absent		
hopE1	PSPTO_4331	Yes	Absent		
hopG1	PSPTO_4727	Yes	Absent		
hopK1	PSPTO_0044	Yes	Absent		
hopN1	PSPTO_1370	Yes	Absent		
hopU1	PSPTO_0501	Yes	Absent		
hopV1	PSPTO_4720	Yes	Absent		
hopX1	PSPTO_A0012	Yes	Absent		
hopAA1-2	PSPTO_4718	Yes	Absent		
hopAM1-1	PSPTO_1022	Yes	Absent		
hopAM1-2	PSPTO_A0005	Yes	Absent		
hopAO1	PSPTO_4722	Yes	Absent		
T3E orthologs deployed by T1 alone					
hopS1′	PSPTO_4597	Yes (operon)	hopS1	PSPTOT1_0510	Yes (operon)
hopPtoAG:: ISPssy	PSPTO_0901	Yes	hopAG1	PSPTOT1_5517	Yes
hopAH1	PSPTO_0905	No	hopAH1	PSPTOT1_5516	Yes (operon)
hopAI1	PSPTO_0906	No	hopAI1	PSPTOT1_5515	Yes (operon)
hopAS1′	PSPTO_0474	Yes	hopAS1	PSPTOT1_1672	Yes
Absent			hopAE1	PSPTOT1_0516	Yes

Continued on following page

[a] Reasons for lack of deployment include: absence of the structural gene; truncation, indicated by a single quotation; truncation by an insertion sequence, indicated by a double colon and the insertion sequence name; absence of a functional hrp promoter. Light gray background indicates that either one or both expression and translocation have not been fully evaluated; dark gray background indicates that either one or both expression and translocation are absent or unlikely; and no background shading indicates orthologs are likely functional.

sufficient difference to warrant inclusion in a different subgroup (HopAB3) relative to the ortholog present in DC3000 (HopAB2).

Because of the draft status of the T1 genome sequence, there is uncertainty regarding a small number of effectors. The effector *hopD1* had a frameshift in the first 454 pyrosequencing assembly. However, upon resequencing the gene from a polymerase chain reaction (PCR) product, it became apparent that two full-length copies of the gene with one alternate nucleotide in position 927 were wrongly assembled into one contig and both nucleotides were integrated into the sequence. Only one copy of the gene is currently annotated, since we do not know the location of the second copy. Homologs of *hopC1* also appear to be present in two copies in T1; however, in this instance the two copies were assembled correctly in separate contigs. A frameshift error present in one of the two copies was confirmed after resequencing the gene from a PCR product. The 412-bp-long 454 contig 00076 has similarity with the 3′ end of the *hopAW1*$_{\mathrm{Pph\ 1448A}}$ gene. The 5′ end is not present in either the 454 or the Solexa draft genome sequences, suggesting that the copy of hopAW1 in T1 is most likely a pseudogene.

The presence of *avrRpt2* in the T1 genome is not sufficient to explain the inability of T1 to cause disease in *A. thaliana*.

The observed HR caused by T1 on *A. thaliana* ‘Columbia’ and its inability to replicate and to cause disease on the other tested *A. thaliana* ecotypes suggests a “gene for gene” interaction between one of the T1-specific effectors and an *A. thaliana* resistance gene. The presence of the *avrRpt2* effector gene in the T1 genome makes the well-known interaction between *avrRpt2* and the cognate *A. thaliana* resistance gene RPS2 (Mindrinos et al. 1994) the prime suspect. To evaluate the role of AvrRpt2$_{\mathrm{Pto\ T1}}$ in resistance, DC3000, DC3000 expressing

AvrRpt2$_{Pto\ JL1065}$, and T1 were compared for growth and presence of disease symptoms between *A. thaliana* ‘Columbia and the *rps2* mutant line rps2-101c of *A. thaliana* ‘Columbia (Mindrinos et al. 1994). As shown in Figure 4A, T1 population densities remained at a low level similar to those of the T3SSS-deficient DC3000 mutant (Fig. 1), even in the absence of the *avrRpt2-RPS2* interaction. However, no HR could be detected on *A. thaliana* ‘Columbia’ *rps2-101c* (Fig. 4B). These observations suggest that the *avrRpt2-RPS2* interaction is the source of the HR elicited by T1 on *A. thaliana* ‘Columbia’ but does not fully explain resistance of *A. thaliana* to T1.

Can gene for gene interactions other than *avrRpt2-RPS2* explain the inability of T1 to cause disease in *A. thaliana*?

Since gene for gene interactions are not always accompanied by a HR (Gassmann 2005), the absence of a HR during T1 infection of the *A. thaliana* ‘Columbia’ rps2 mutant line does not exclude the possibility that ETI induced by T1 effectors other than AvrRpt2 is the determining factor in the inability of T1 to cause disease in *A. thaliana*. This possibility can be evaluated using *A. thaliana* lines mutagenized in the genes *RAR1* or *SGT1b*, previously shown to be essential for ETI in *A. thaliana* (Azevedo et al. 2006). Specifically, *SGT1b* was shown to be necessary for various *Hyaloperonospora parasitica–A. thaliana* gene for gene interactions and a mutation in *RAR1* was shown to reduce *A. thaliana* resistance to the *P. syringae* effectors AvrRpt2, AvrRpm1, and AvrRps4 (Austin et al. 2002). The *SGT1b* homolog *SGT1a* was not required for resistance triggered by these *P. syringae* effectors (Azevedo et al. 2006), nor could it complement a mutation in *SGT1b* during *Hyaloperonospora parasitica–A. thaliana* gene-for-gene interactions (Austin et al. 2002), except when overexpressed (Azevedo et al. 2006).

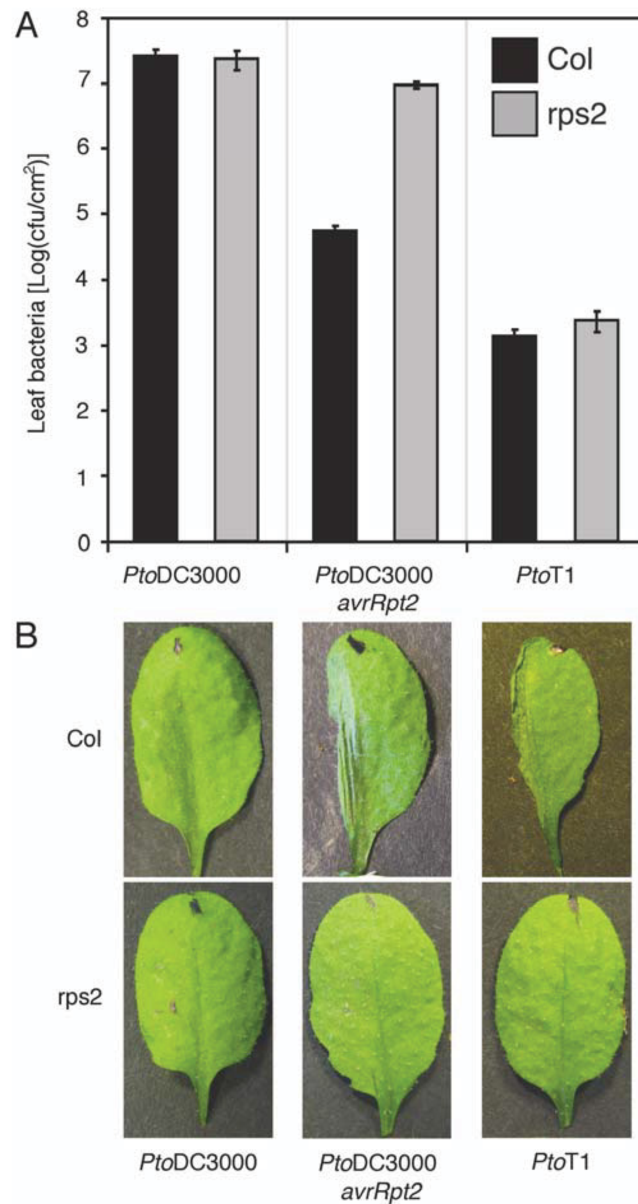


Fig. 4. Comparison of in planta growth of *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000, DC3000 expressing AvrRpt2, and T1 on *Arabidopsis thaliana* ecotype Col-0 and Col-0 *rps2-101c* mutated in the *RPS2* resistance gene. **A,** Plants were inoculated by spraying bacteria at an optical density at 600 nm (OD_{600}) of 0.1, and CFU per square centimeter was determined 3 days later for six leaf disks per isolate and plant. While DC3000 expressing *avrRpt2* grew statistically significantly more in the Col-0 *rps2-101c* plants than in the Col-0 plants ($P < 0.01$), growth of T1 was not statistically significantly different on the Col-0 *rps2-101c* plants comapred with Col-0 plants, based on Student's *t*-test ($P > 0.1$). **B,** Bacteria were infiltrated into leaves of *A. thaliana* Col-0 and Col-0 *rps2-101c* at an OD_{600} of 0.3, and pictures were taken 16 h later.

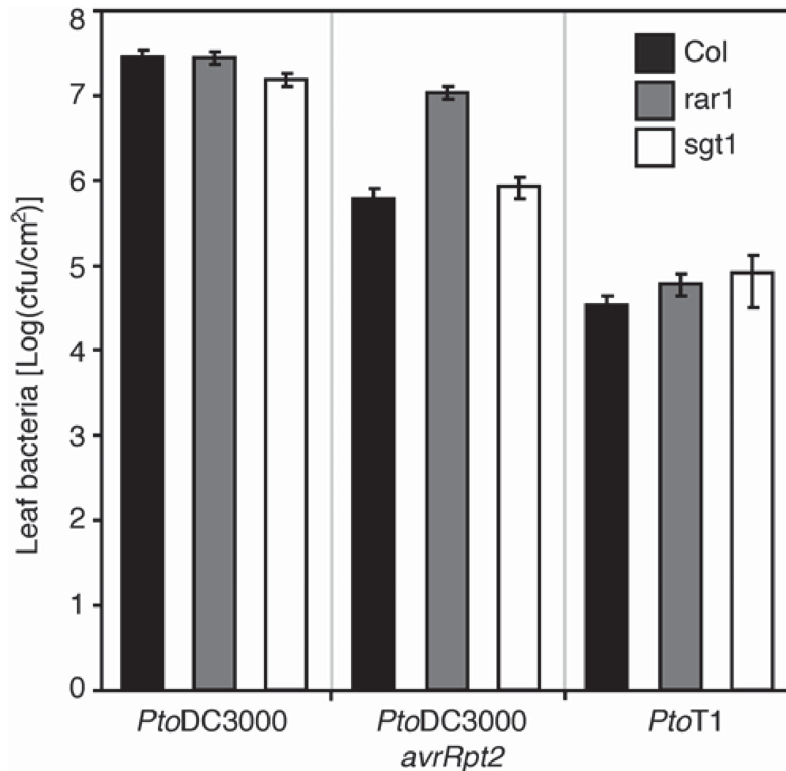


Fig. 5. Comparison of in planta growth of *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000, DC3000 expressing AvrRpt2, and *Pseudomonas syringae* pv. *tomato* T1 on wild-type *Arabidopsis thaliana* ecotype Col-0 and *A. thaliana* Col-0 mutated in the genes *SGT1* and *RAR1*, respectively. Plants were inoculated, and CFU were counted 3 days postinfection. While DC3000 expressing *avrRpt2* grew statistically significantly more in Col-0 plants mutated in *RAR1* than in Col-0 plants ($P < 0.0001$), growth of T1 was not statistically significantly different in plants mutated in *RAR1* or *SCT1* as compared with Col-0 plants, based on Student's *t*-test ($P > 0.1$).

Therefore, we reasoned that if ETI is induced during the T1–*A. thaliana* interaction by effectors other than AvrRpt2, T1 growth on *SGT1b* or *RAR1* mutant lines of A. *thaliana* might be increased. However, as shown in Figure 5, while growth of DC3000 expressing *avrRpt2* was increased 10-fold in the RAR1 mutant, T1 growth was not significantly different in either the SGT1b or the RAR1 mutant as compared with wild-type *A. thaliana* 'Columbia'. This was also true for the *P. syringae* pv. *tomato* isolate Max13 (Yan et al. 2008), from which *avrRpt2* is naturally missing, and for another *P. syringae* pv. *tomato* isolate with a disruption

of *avrRpt2* (data not shown). Both isolates are identical to T1 based on our published MLST analysis (Yan et al. 2008). Therefore, while immunity triggered by AvrRpt2 and possibly other effectors contributes to *A. thaliana* resistance to T1 and similar *P. syringae* pv. *tomato* isolates, other mechanisms independent of ETI can be expected to play critical roles also.

Table 4. *Pseudomonas syringae* pv. *tomato* isolates used in this study

Isolate name	Description	Reference
DC3000	*Arabidopsis thaliana* and tomato pathogen; rifampicin-resistant derivative of strainNCPPB 1106, isolated from tomato on the island of Guernsey (U.K.) by R. A. Lelliott.	Buell et al. 2003; Cuppels 1986; Yan et al. 2008
DC3000 CUCPB5112	Type III secretion system–deficient DC3000, because of deletion of the *hrcC* gene	Penaloza-Vazquez et al. 2000
DC3000 pLH12	DC3000 expressing AvrRpt2$_{\text{Pto JL1065}}$ from its native promoter on plasmid pLAFR3	Whalen et al. 1991
T1	Isolated from tomato in Canada by G. Bonn (Agriculture & Agri-Food Canada)	Whalen et al. 1991; Yan et al. 2008

d. Discussions

Although T1 and DC3000 are both considered to be members of *P. syringae* pv. *tomato*, we have previously shown that they belong to two closely related but clearly distinct phylogenetic groupings with distinct host ranges (Yan et al. 2008). While T1-like strains cause disease only on tomato and are found on tomato with bacterial speck disease all over the world, DC3000-like strains cause disease on *A. thaliana*, cauliflower, and tomato and have been isolated mainly from family *Brassicaceae* and wild *Solanaceae* plants. Here, we show that resistance of *A. thaliana* to T1-like strains occurs in several ecotypes and that T1-like strains only reach population densities similar to a T3SS-deficient mutant of DC3000. These findings suggest that T1-like strains share a fundamental difference relative to DC3000 that prevents them from being pathogens of *A. thaliana*.

To date, the basis of nonhost resistance exhibited by *A. thaliana* toward *P. syringae* strains has been best studied using strains of *P. syringae* pv. *phaseolicola* (de Torres et al. 2006; Ham et al. 2007; Lu et al. 2001). However, the large number of differences between the *P. syringae* pv. *phaseolicola* strains and DC3000

greatly complicates identification of pathogen–host interaction factors using genome comparison. In contrast, the T1 genome exhibits a high level of similarity to DC3000 with the majority of differences consisting of short hypothetical genes distributed throughout otherwise colinear regions. This suggests that many of these hypothetical genes may not be true coding sequences.

A surprising outcome of this study is that, while the two *P. syringae* pv. *tomato* genomes exhibit a high level of overall similarity, their repertoires of T3SS effectors are strikingly diverse. This indicates that T3SS effectors may play an important role in determining the difference in host specificity between the two strains, particularly with regard to *A. thaliana*. Effectors present in T1 but absent in DC3000 could lead to gene for gene resistance of *A. thaliana* to T1, while effectors present in DC3000 but missing in T1 could account for the inability of T1 to suppress ETI or PTI. Although T1 also lacks biosynthetic genes for the phytotoxin coronatine, this toxin has no demonstrated role in host range determination.

Of the 11 effectors present in the T1 genome but not in DC3000, the most notable is the well-characterized AvrRpt2, indirectly recognized in *A. thaliana* by the resistance gene product RPS2 upon degradation of RIN4 (Axtell and Staskawicz 2003; Kim et al. 2005; Mackey et al. 2003). However, it has been shown here that the *avrRpt2-RPS2* gene for gene interaction is not the sole mechanism of *A. thaliana* resistance to T1, since T1 growth on *A. thaliana* mutated in the RPS2 gene is not increased compared with *A. thaliana* wild-type plants and symptoms are not increased. Moreover, additional putative gene for gene interactions are also insufficient to explain resistance of *A. thaliana* to T1, since the genes RAR1 and SGT1, known to contribute to gene for gene resistance, are not necessary for this resistance. Therefore, we hypothesize that *A. thaliana* resistance to T1 is based on multiple parallel defense pathways, as previously observed for the *P. syringae* pv. *phaseolicola–A. thaliana* interaction (Ham et al. 2007). To

identify the pathways involved, future work will be directed towards expression of DC3000 effectors in T1 and evaluation of their phenotypes in various single- and multiple-defense mutants of *A. thaliana*.

The two DC3000 effectors that probably have the greatest potential of conferring increased virulence of T1 on *A. thaliana* are *avrPto1*$_{\text{Pto DC3000}}$ and *hopM1*$_{\text{Pto DC3000}}$, which have both been shown to be important virulence factors during DC3000 infection of *A. thaliana* (Hauck et al. 2003; Nomura et al. 2006). However, the *P. syringae* pv. *tomato* isolate JL1065, which is closely related to the T1-like strains (although phylogenetically distinct [Yan et al. 2008]) expresses AvrPto1 as well as HopM1 (data not shown) but is still unable to cause disease on *A. thaliana* even in the absence of the *avrRpt2–RPS2* gene for gene interactions (Yan et al. 2008). Moreover, *hopM1* and *avrE1* have redundant function in the DC3000–*A. thaliana* interaction (DebRoy et al. 2004) and *avrE1* is present in the T1 genome (DNA identity, 95%). This makes it unlikely that simply expressing *hopM1*$_{\text{Pto DC3000}}$ in T1 could render T1 pathogenic on *A. thaliana*. Nonetheless, we performed this experiment and, as expected, could not observe any increase in bacterial growth or formation of disease symptoms in *A. thaliana* 'Columbia' (data not shown).

Based on the results obtained so far, we predict that identification of gene products capable of conferring increased growth or symptomology of T1, or both, in *A. thaliana* will require the simultaneous mutation of several *A. thaliana* defense pathways in combination with expression of one or more DC3000 effectors in T1. It is also possible that the inability of T1 to grow and cause disease on *A. thaliana* is due to yet unknown incompatibility mechanisms that are independent of characterized defense pathways and effectors, for example, gene for gene resistance independent of *RAR1* and *SGT1b* or nutritional requirements of T1 not met by the *A. thaliana* apoplast. The identified gene differences between T1 and

DC3000 and allelic differences between shared genes represent an ideal unbiased starting point for experimental characterization of these mechanisms.

In conclusion, this study demonstrates that the comparison of a high quality *P. syringae* draft genome sequence with a completely assembled genome sequence of a closely related strain represents a powerful approach for identification of differences in gene content between strains. In particular, draft genome sequencing is an effective and rapid means of identifying the complete effector repertoire of individual *P. syringae* strains when used in combination with BLAST analysis and experimental testing of effector candidates identified by proximity to *hrp* box motifs. Experimental dissection of the roles of strain-specific effectors and other candidate host specificity genes identified through genome comparison is expected to reveal important insights into the specific interactions that govern differences in host range observed for the two strains compared here. Additional draft genome sequences currently being generated for other *P. syringae* strains by our group and by other researchers can be expected to take *P. syringae* genomics to new levels, accelerating the discovery of yet unknown aspects of molecular *P. syringae*–plant interactions.

e. Materials and Methods

Bacterial strains and growth media.

Bacterial strains are listed in Table 4. *P. syringae* isolates were grown at 28°C on King's broth (King et al. 1954) agar plates supplemented with the following antibiotics at the following concentrations when necessary to maintain plasmids: kanamycin (50 μg/ml), tetracycline (15 μg/ml), and gentamycin (10 μg/ml).

Plant growth and infections.

For bacterial growth in planta and disease assays, *A. thaliana* plants were grown with a 16-h day and temperature was kept at 23°C during the day and 19°C during the night. Plants were infected at an age of 3 weeks, and bacterial population sizes were determined 3 days after infection as previously described (Yan et al. 2008). Differences between population sizes were statistically analyzed in the JMP 7 program (SAS Institute, Cary, NC, U.S.A.) using Student's t-test.

Genome sequencing and assembly.

For 454 sequencing, T1 (G. Martin lab glycerol stock #3876) was grown overnight in 1 ml of liquid media. Genomic DNA was extracted using a Wizard DNA purification kit (product #A1120; Promega, Madison, WI, U.S.A.) and by following the manufacturer's protocol for isolation of genomic DNA from Gram-negative bacteria. The optical densities at 260 and 280 nm were determined using a NanoDrop ND-1000 spectrophotometer and were found to be 2.08. A total of 60 µg of DNA was precipitated in ethanol and 3 M NaAC. An aliquot of 10 µg of DNA was then used for single-stranded template DNA library construction and a titration run was carried out on the GS FLX sequencer (Roche, Basel, Switzerland) in accordance with the manufacturer's protocol. It was determined that the ratio of one DNA molecule per bead was the optimum. Subsequently, a bulk emulsion PCR was carried out using this ratio, to provide the beads to run a full LR 70 plate. A total of one 7-h run was performed. The quality filtered reads were then assembled into contigs using the Newbler assembler.

For Solexa sequencing, DNA was prepared from bacteria grown in L medium using the Puregene Genomic DNA purification kit (Gentra Systems, Inc., Minneapolis) according to manufacturer's instructions. Libraries for single-read

Illumina sequencing were prepared from 5 μg of DNA using a Genomic DNA sample prep kit (Illumina, Inc., Cambridge). DNA was fragmented by nebulization for 6 min at a pressure of 32 psi. For end-repair and phosphorylation, sheared DNA was purified using the QIAquick nucleotide removal kit (Qiagen, Crawley, U.K.). The end-repaired DNA was A-tailed and adaptors were ligated according to manufacturer's instructions. Size fractionation and purification of ligation products was performed using a 5% polyacrylamide gel run in tris-borate-EDTA at 180 V for 120 min. Gel slices were cut containing DNA in the 130- to 150-bp range. DNA was than extracted using 0.3 M sodium acetate and 2 mM EDTA (pH 8.0), followed by ethanol precipitation. Using 18 PCR cycles with primer 1.1 and 2.1 (Illumina, Inc.), 5′ adaptor extension and enrichment of the library was performed. The library was finally purified using a QIAquick PCR purification kit and adjusted to a concentration of 10 nM in 0.1% Tween. Flow cells were prepared according to manufacturer's instructions using cluster generation kits and a cluster station. Sequencing reactions were performed on a 1G genome analyzer (Illumina, Inc.). A total of 5 pM of the library were used to achieve approximately 25,000 to 30,000 clusters per tile. Sequence reads were assembled using Velvet 0.5 (Zerbino and Birney 2008).

Annotation.

The T1 draft genome was annotated by first blasting all open reading frames longer than 33 amino acids against all sequenced genomes in genera *Pseudomonas*, *Ralstonia*, and *Xanthomonas* available at National Center for Biotechnology Information (NCBI) in September, 2007; and all T3SS effector genes from the P. syringae effector database at the Pseudomonas-Plant Interaction website, using the GRC program (A. Warren and J. C. Setubal, unpublished). All ORF with no hits

were then blasted against the nonredundant database at NCBI and were analyzed using InterProScan (Quevillon et al. 2005).

Comparative genomics.

The T1 454 and Solexa draft genome sequences were aligned against the published sequence of DC3000 (Buell et al. 2003), using blastn from the NCBI BLAST software suite (Altschul et al. 1997) using default parameters except -e 1e-10 -b 1 -v 1. Circular genome views of the alignments were generated using CGView (Stothard and Wishart 2005).

Protein repertoires were compared using pairwise blast (Altschul et al. 1997), 3GC (Telles et al. 2005) (used for the Venn diagrams shown in Figure 3), and OrthoMCL (Li et al. 2003), the latter of which clusters proteins in orthologous groups based on the MCL algorithm (Enright et al. 2002).

Effector prediction, cloning, *P. syringae* transformation, and testing for translocation.

T3SS effectors and other genes under the control of the HrpL sigma factor were predicted as described by Vencato and associates (2006). Effectors were cloned using the Gateway (Invitrogen, Carlsbad, CA, U.S.A.) cloning system into pENTR/SD/D-TOPO (primers available upon request), were recombined into the P. syringae vectors pBBRPnpt-GW-Cya and expressed in DC3000, and were tested for translocation as previously described (Schechter et al. 2004).

f. Acknowledgements

Work in the Vinatzer lab and pyrosequencing of T1 was funded by the Virginia Tech Institute for Biomedical and Public Health Sciences and the National Science Foundation (award number 0746501). N. F. Almeida was funded by Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq), fellowship 200447/2007-6. Research at Cornell University and the Boyce Thompson Institute was supported by the National Science Foundation (award number DBI-0605059).

g. Literature Cited

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389-402.

Angot, A., Vergunst, A., Genin, S., and Peeters, N. 2007. Exploitation of eukaryotic ubiquitin signaling pathways by effectors translocated by bacterial type III and type IV secretion systems. PLoS Pathog. 3:e3. Published online.

Austin, M. J., Muskett, P., Kahn, K., Feys, B. J., Jones, J. D., and Parker, J. E. 2002. Regulatory role of *SGT1* in early *R* gene-mediated plant defenses. Science 295:2077-80.

Axtell, M. J., and Staskawicz, B. J. 2003. Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. Cell 112:369-77.

Azevedo, C., Betsuyaku, S., Peart, J., Takahashi, A., Noel, L., Sadanandom, A., Casais, C., Parker, J., and Shirasu, K. 2006. Role of SGT1 in resistance

protein accumulation in plant immunity. EMBO (Eur. Mol. Biol. Organ.) J. 25:2007-2016.

Bentley, D. R. 2006. Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 16:545-552.

Bhavsar, A. P., Guttman, J. A., and Finlay, B. B. 2007. Manipulation of host-cell pathways by bacterial pathogens. Nature 449:827-834.

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M. J., Haft, D. H., Nelson, W. C., Davidsen, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S. E., Feldblyum, T. V., D'Ascenzo, M., Deng, W. L., Ramos, A. R., Alfano, J. R., Cartinhour, S., Chatterjee, A. K., Delaney, T. P., Lazarowitz, S. G., Martin, G. B., Schneider, D. J., Tang, X., Bender, C. L., White, O., Fraser, C. M., and Collmer, A. 2003. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186.

Charity, J. C., Pak, K., Delwiche, C. F., and Hutcheson, S. W. 2003. Novel exchangeable effector loci associated with the *Pseudomonas hrp* pathogenicity island: Evidence for integron-like assembly from transposed gene cassettes. Mol. Plant-Microbe Interact. 16:495-507.

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell 124:803-814.

Cuppels, D.A. 1986. Generation and characterization of Tn5 insertion mutations in *Pseudomonas syringae* pv. *tomato*. Appl. Environ. Microbiol. 51:323-327.

Dangl, J. L., and Jones, J. D. 2001. Plant pathogens and integrated defence responses to infection. Nature 411:826-833.

DebRoy, S., Thilmony, R., Kwack, Y. B., Nomura, K., and He, S. Y. 2004. A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. Proc. Natl. Acad. Sci. U.S.A. 101:9927-9932.

de Torres, M., Mansfield, J. W., Grabov, N., Brown, I. R., Ammouneh, H., Tsiamis, G., Forsyth, A., Robatzek, S., Grant, M., and Boch, J. 2006. *Pseudomonas syringae* effector AvrPtoB suppresses basal defence in Arabidopsis. Plant J. 47:368-382.

Deng, W. L., Rehm, A. H., Charkowski, A. O., Rojas, C. M., and Collmer, A. 2003. *Pseudomonas syringae* exchangeable effector loci: Sequence diversity in representative pathovars and virulence function in *P. syringae* pv. *syringae* B728a. J. Bacteriol. 185:2592-2602.

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30:1575-1584.

Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J. E., Lapidus, A., Detter, J. C., Land, M., Richardson, P. M., Kyrpides, N. C., Ivanova, N., and Lindow, S. E. 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 102:11064-11069. Felix, G., Duran, J. D., Volko, S., and Boller, T. 1999. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. 18:265-276. Fouts, D. E., Abramovitch, R. B., Alfano, J. R., Baldo, A. M., Buell, C. R.,

Cartinhour, S., Chatterjee, A. K., D'Ascenzo, M., Gwinn, M. L., Lazarowitz, S. G., Lin, N. C., Martin, G. B., Rehm, A. H., Schneider, D. J., van Dijk, K., Tang, X., and Collmer, A. 2002. Genomewide identification of *Pseudomonas syringae* pv. *tomato* DC3000 promoters controlled by the HrpL alternative sigma factor. Proc. Natl. Acad. Sci. U.S.A. 99:2275-2280.

Gassmann, W. 2005. Natural variation in the Arabidopsis response to the avirulence gene *hopPsyA* uncouples the hypersensitive response from disease resistance. Mol. Plant-Microbe Interact. 18:1054-1060.

Ham, J. H., Kim, M. G., Lee, S. Y., and Mackey, D. 2007. Layered basal defenses underlie non-host resistance of Arabidopsis to *Pseudomonas syringae* pv. *phaseolicola*. Plant J. 51:604-616.

Hauck, P., Thilmony, R., and He, S. Y. 2003. A *Pseudomonas syringae* type III effector suppresses cell wall-based extracellular defense in susceptible Arabidopsis plants. Proc. Natl. Acad. Sci. U.S.A. 100:8577- 8582.

Innes, R. W., Bent, A. F., Kunkel, B. N., Bisgrove, S. R., and Staskawicz, B. J. 1993. Molecular analysis of avirulence gene *avrRpt2* and identification of a putative regulatory sequence common to all known *Pseudomonas syringae* avirulence genes. J. Bacteriol. 175:4859-4869.

Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., Daugherty, S. C., Deboy, R., Durkin, A. S., Giglio, M. G., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S., Crabtree, J., Creasy, T., Davidsen, T., Haft, D. H., Zafar, N., Zhou, L., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C. M., Chatterjee, A. K., Cartinhour, S., Schneider, D. J., Mansfield, J., Collmer, A., and Buell, C. R. 2005. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. J. Bacteriol. 187:6488-6498.

Jones, J. D., and Dangl, J. L. 2006. The plant immune system. Nature 444:323-329.

Kim, M. G., da Cunha, L., McFall, A. J., Belkhadir, Y., DebRoy, S., Dangl, J. L., and Mackey, D. 2005. Two *Pseudomonas syringae* type III effectors inhibit RIN4-regulated basal defense in Arabidopsis. Cell 121:749- 759.

King, E.O., Ward, M. K., and Raney, D. E. 1954. Two simple media for the demonstration of pyocyanin and fluorescin. J. Lab. Clin. Med. 44:201-207.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. 2004. The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell 16:3496-3507.

Li, L., Stoeckert, C. J., Jr., and Roos, D. S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178- 189.

Li, X., Lin, H., Zhang, W., Zou, Y., Zhang, J., Tang, X., and Zhou, J. M. 2005. Flagellin induces innate immunity in nonhost interactions that is suppressed by *Pseudomonas syringae* effectors. Proc. Natl. Acad. Sci. U.S.A. 102:12990-12995.

Lin, N. C., and Martin, G. B. 2007. Pto- and Prf-mediated recognition of AvrPto and AvrPtoB restricts the ability of diverse *Pseudomonas syringae* pathovars to infect tomato. Mol. Plant-Microbe Interact. 20:806- 815.

Lindeberg, M., Myers, C. R., Collmer, A. and Schneider, D. J. 2008. Road-map to new virulence determinants in *Pseudomonas syringae*: Insights from comparative genomics and genome organization. Mol. Plant-Microbe Interact. 21:685-700.

Lindeberg, M., Cartinhour, S., Myers, C. R., Schechter, L. M., Schneider, D. J., and Collmer, A. 2006. Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains. Mol. Plant-Microbe Interact. 19:1151-1158.

Lipka V., Dittgen, J., Bednarek, P., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., Landtag, J., Brandt, W., Rosahl, S., Scheel, D., Llorente, F., Molina, A., Parker, J., Sommerville, S., and Schulze-Lefert, P. 2005. Pre- and postinvasion defenses both contribute to nonhost resistance in Arabidopsis. Science 310:1180-1183.

Lu, M., Tang, X., and Zhou, J. M. 2001. Arabidopsis *NHO1* is required for general resistance against *Pseudomonas* bacteria. Plant Cell 13:437- 447.

Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. 2003. Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell 112:379-389.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M., and Spratt, B. G. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. 95:3140- 3145.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Mindrinos, M., Katagiri, F., Yu, G. L., and Ausubel, F. M. 1994. The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089-1099.

Newman, M. A., von Roepenack-Lahaye, E., Parr, A., Daniels, M. J., and Dow, J. M. 2002. Prior exposure to lipopolysaccharide potentiates expression of plant defenses in response to bacteria. Plant J. 29:487-495.

Nomura, K., Debroy, S., Lee, Y. H., Pumplin, N., Jones, J., and He, S. Y. 2006. A bacterial virulence protein suppresses host innate immunity to cause plant disease. Science 313:220-223.

Penaloza-Vazquez, A., Preston, G. M., Collmer, A., and Bender, C. L. 2000. Regulatory interactions between the Hrp type III protein secretion system and coronatine biosynthesis in *Pseudomonas syringae* pv. *tomato* DC3000. Microbiology 146 (Pt 10):2447-2456.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. 2005. InterProScan: Protein domains identifier. Nucleic Acids Res. 33:W116-20. Published online.

Rohmer, L., Guttman, D. S., and Dangl, J. L. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. Genetics 167:1341-1360.

Rose, L. E., Michelmore, R. W., and Langley, C. H. 2007. Natural variation in the *Pto* disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of Pto. Genetics 175:1307-1319.

Schechter, L. M., Roberts, K. A., Jamir, Y., Alfano, J. R., and Collmer, A. 2004. *Pseudomonas syringae* type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. J. Bacteriol. 186:543-555.

Stothard, P., and Wishart, D. S. 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21:537-539.

Telles, G. P., Brigido, M. M., Almeida, N. F., Viana, C. J., Anjos, D. A. S., and Walter, M. E. M. T 2005. A Method for Comparing Three Genomes. In: Advances in Bioinformatics and Computational Biology, J. C. Setubal and S. Verjovsky-Almeida, eds. Springer, Heidelberg, Germany.

Vencato, M., Tian, F., Alfano, J. R., Buell, C. R., Cartinhour, S., DeClerck, G. A., Guttman, D. S., Stavrinides, J., Joardar, V., Lindeberg, M., Bronstein, P. A., Mansfield, J. W., Myers, C. R., Collmer, A., and Schneider, D. J. 2006. Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A. Mol. Plant-Microbe Interact. 19:1193-1206.

Wanner, L. A., Mittal, S., and Davis, K. R. 1993. Recognition of the avirulence gene *avrB* from *Pseudomonas syringae* pv. *glycinea* by *Arabidopsis thaliana*. Mol. Plant-Microbe Interact. 6:582-591.

Wei, C. F., Kvitko, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N. C., Martin, G. B., Huang, H. C., and Collmer, A. 2007. A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. Plant J. 51:32-46.

Whalen, M. C., Innes, R. W., Bent, A. F., and Staskawicz, B. J. 1991. Identification of *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. Plant Cell 3:49-59.

Yan, S., Liu, H., Mohr, T. J., Jenrette, J., Chiodini, R., Zaccardelli, M., Setubal, J. C., and Vinatzer, B. A. 2008. Role of recombination in the evolution of the

model plant pathogen *Pseudomonas syringae* pv. tomato DC3000, a very atypical tomato strain. Appl. Environ. Microbiol. 74:3171-181.

Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-9.

Zipfel, C. 2008. Pattern-recognition receptors in plant innate immunity. Curr. Opin. Immunol. 20:10-16.

AUTHOR-RECOMMENDED INTERNET RESOURCES

PtoT1 genome website: genome.ppws.vt.edu/pseudomonas Pseudomonas-Plant Interaction website: pseudomonas-syringae.org Newbler assembler: www.454.com

Chapter 4

Investigation of the Basis for Non-host Resistance of *A. thaliana* to Pto T1 and the Roles of Type III Effectors in Host Range Determination with HopM1 as Case Study

Title: Investigation of the Basis for Non-host Resistance of *A. thaliana* to *Pseudomonas syringae* pv. tomato T1 and the Roles of Type III Effectors in Host Range Determination with HopM1 as Case Study

Authors: Shuangchun Yan, Haijie Liu, Rongman Cai, Magdalen Lindeberg, Boris A. Vinatzer

a. Abstract

P. syringae pv. *tomato* (Pto) DC3000 causes disease in *A. thaliana* and tomato while a close relative, Pto T1, causes disease only in tomato. We find that *A. thaliana* resistance to T1 is an example of multi-layered non-host resistance that includes – but goes beyond – effector-triggered immunity (ETI). Based on the transcriptional analysis of *A. thaliana* defense genes and DC3000/T1 co-infections, the inability of T1 to infect *A. thaliana* appears to be primarily due to its inability to suppress pathogen associated molecular patterns (PAMPs)-triggered immunity (PTI). The genomes of nine strains that are closely related to DC3000 and T1 and that either cause, or do not cause, disease in Arabidopsis were sequenced. Conserved differences in type III secreted (T3S) effector repertoires between Arabidopsis pathogens and non-pathogens were identified to determine the molecular basis of host range differences. Unexpectedly, expression of DC3000 effectors conserved among *A. thaliana* pathogens and absent from most non-*A. thaliana* pathogens did not significantly increase growth of T1. This was also true when these effectors were expressed in a T1 strain in which the only two effectors with ETI activity were deleted. Experiments to determine if these effectors can at least suppress individual *A. thaliana* PTI defense pathways is under way. Some

DC3000 effectors even reduce T1 growth in *A. thaliana*. In particular, *hopM1*$_{DC3000}$ not only reduces growth of T1 and induces cell death in *A. thaliana* but also reduces growth of T1 in tomato, increases the number of specks during tomato infection, and induces cell death when transiently expressed in tomato. When *hopM1*$_{DC3000}$ is expressed in *Pseudomonas fluorescens* expressing also a *P. syringae* type III secretion system, growth of *P. fluorescens* in *A. thaliana* is dramatically increased and symptoms are induced. Therefore, *hopM1*$_{DC3000}$ consistently induces cell death on different plant species when secreted from different pathogens but this cell death may either contribute to virulence or interfere with it depending on the particular pathogen–host interaction. Genetic background effects of the pathogen may also explain why other DC3000 effectors besides *hopM1*$_{DC3000}$ that have known virulence effects in DC3000 did not increase virulence of T1 in *A. thaliana* when ectopically expressed in this strain.

b. Introduction

Pseudomonas syringae represents a bacterial species complex that includes plant pathogens that cause disease on over 50 plant species (Sarkar, Gordon et al. 2006; Cunnac, Lindeberg et al. 2009) The species complex is divided into pathovars based on host range (Gardan, Shafik et al. 1999; Collmer, Badel et al. 2000). Members of some pathovars have been intensively studied because they cause disease on the model organism *Arabidopsis thaliana,* for example, *P. syringae pv.* tomato (Pto) DC3000 and *P. syringae* pv. *maculicola*) (Pma) ES4326 (Whalen, Innes et al. 1991; Buell, Joardar et al. 2003). Since the *P. syringae* species complex contains so many different pathovars with different host ranges. *P. syringae* also represents an ideal model group of pathogens to study host range determination.

Plant innate immunity, such as PAMP-trigger immunity (PTI) exemplified by flg22-FLS2 interaction (Gomez-Gomez and Boller 2000), and effector-triggered immunity (ETI) (Jones and Dangl 2006) are believed to be playing major roles in host range determination. Both PTI and ETI result in a variety of defense responses, such as callose deposition, release of reactive oxygen species, and accumulation of defense response proteins, such as PR1 (Asai, Tena et al. 2002; Göhre and Robatzek 2008). For *P. syringae*, the type III secretion system (T3SS) and effectors translocated through the T3SS into plant cells are key to host compatibility because they are required for overcoming these plant defenses (Lindeberg, Cunnac et al. 2009). However these effectors can also betray the pathogen and elicit strong effector-triggered immunity (ETI) in plants, which often results in incompatibility. ETI is highly cultivar and strain specific, and is often the sole determinant of host-resistance at the pathogen race – host cultivar level (Lindeberg, Cunnac et al. 2009).

In some cases, ETI is the sole mechanism of host specificity at the pathovars – plant species level. For example, deleting a single effector, *hopQ1*, from the tomato and *A. thaliana* pathogen DC3000 allows this pathogen to extend its host range to *Nicotiana benthamiana* (Wei, Kvitko et al. 2007). In other cases, non-host resistance at the pathovars – host species level is believed to be multi-genic. For example, the bean pathogen *P. syringae* pv. *phaseolicola* (Pph) NSP3121 can only grow to high levels in the *A. thaliana* double mutant pmr4/pad4 (severely compromised in its innate immune response) after expressing ectopically the DC3000 effector *hopM1* (Ham, Kim et al. 2007).

PTI is also involved in host range determination. Introduction of pattern recognition receptors (PRRs) from *A. thaliana* to tomato and *N. benthamiana* has, for example, significantly reduced pathogen growth (Lacombe, Rougon-Cardoso et al. 2010). And different alleles of *fliC* present in different strains of the plant pathogen *Xanthomonas* trigger different degree of immunity in Arabidopsis (Sun, Dunning et al. 2006). The bottom line is, pathogens have to be able to overcome several layers of plant defenses in order to be successful.

Although three *P. syringae* strains have been completely sequenced and annotated (Buell, Joardar et al. 2003; Feil, Feil et al. 2005; Joardar, Lindeberg et al. 2005), the differences in gene content and in allelic differences between these strains are so large, and the survey of strains with different host range so limited, that we cannot easily correlate specific genomic differences with actual host range differences. We recently characterized a group of closely related *P. syringae* strains that yet exhibit highly diverse host ranges (Yan, Liu, et al. 2008; Cai, Yan et al., in preparation). These strains are thus suitable material for a comparative genomics study to understand host range determination, where genomic differences between strains with different host range are so limited that the roles of these differences can be validated experimentally. We found significantly

divergent effector repertoires between the draft genome sequence of strain Pto T1, which does not cause disease in *A. thaliana*, and genome sequence of DC3000, a very closely related strain. This suggests that the T3E repertoire is evolving at much faster pace than the rest of the genome (Almeida, Yan et al. 2009). This divergence presumably occurred during a short period of time, therefore a more in depth comparison of DC3000 and T1 should provide more insights into how host range is determined and evolved. Additionally, tomato strains similar to DC3000 based on multi-locus sequence typing are extremely rare, while T1-like strains have been continuously isolated from the field (Yan, Liu, et al. 2008; Cai, Lewis, et al. submitted). This suggests that although DC3000 was isolated from tomato, it is less adapted to tomato than T1-like typical tomato strains. Therefore we consider DC3000 and T1 ideal for a comparative case study of a recent host range divergence event.

We previously showed that ETI and hypersensitive reaction (HR) triggered by the T1 effector AvrRpt2 alone does not explain why T1 cannot cause disease in *A. thaliana*. We also showed that resistance of *A. thaliana* to T1 was likely multi-layered non-host resistance since T1 did neither cause disease in four different *A. thaliana* ecotypes nor in different mutants. Here we extend the analysis of the *A. thaliana* responses to T1 beyond the AvrRpt2–RPS2 interaction and we compare additional draft genome sequences to identify conserved genomic differences between strains that either cause disease or do not cause disease in *A. thaliana*. We investigate whether adding effectors to T1 that are present only in *A. thaliana* pathogens can increase growth of T1 in *A. thaliana* or whether they can suppress PAMP/effector-triggered defense responses. Then we focus on the investigation of effector gene *hopM1* and its ambiguous roles in suppressing host defenses and symptom formation.

c. Results

Table 1. Strains and plasmid used in this study.

Species	Pathovar	Strain Name	Strain #	Source of Strain	Plasmid/Note	Reference
P. syringae	tomato	Max4	#11			Yan et al. 2008
P. syringae	tomato	LNPV17.41	#20	M. Zaccardelli		Yan et al. 2008
P. syringae	tomato	K40	#45			Yan et al. 2008
P. syringae	tomato	JL1065	#22	R. Jackson		Yan et al. 2008
P. syringae	tomato	NCPPB1108	#447	D. Arnold		Yan et al. 2008
P. syringae	tomato	T1	#156/#611(EV)	T. Denny	WT; EV=pME6010	Whalen et al. 1991; Yan et al. 2008
P. syringae	antirrhini	126	#224	D. Cuppels		Yan et al. 2008
P. syringae	macolicola	F1	#247	J. Damicone		Yan et al. 2008
P. syringae	macolicola	M3	#134	J. Greenberg		Yan et al. 2008
P. syringae	macolicola	M6	#136	J. Greenberg		Yan et al. 2008
P. syringae	tomato	DC3000	#1/#650(EV tet)/#1095(Kan)		WT; EV=pME6010; random Kan^R insert	Buell et al. 2003; Cuppels 1986; Yan et al. 2008
P. syringae	tomato	DC3000 ΔhopM1	#1581		without *hopM1*	this study
P. syringae	tomato	DC3000 ΔCEL		D. Mackey	without CEL region	Alfano, et al. 2000
P. syringae	tomato	DC3000 DB29	#1023	B. Kunkel	does not produce COR	Uppalapati, Ishiga et al. 2007
P. syringae	tomato	DC3000 ΔCELΔCOR	#956	D. Mackey	without CEL region and does not produce COR	
P. syringae	tomato	DC3000 ΔhrcC	#953	D. Mackey	without *hrcC*, no functional T3SS	
P. syringae	tomato	DC3000 ΔhopM1	#1595		complemented with pME6010-ORF43	this study
P. syringae	tomato	T1’	#1032/#1068(EV)		without *avrRpt2*; EV=pME6010	this study
P. syringae	tomato	T1*	#1408/#1551(EV)		without *avrRpt2* and *hopAS1*;	Kee at al.;this study

				EV=pME6010	
P. fluorescens	Pf0-1 EtHAn	#1351/#1477(EV)		WT; EV=pME6010	Thomas, et al. 2009; this study
P. fluorescens	Pf0-1 EtHAn	#1436		pME6010-ORF43	this study
E. coli				pME6010	Heeb, et al. 2000
E. coli		#837	D. Mackey	pORF43	Badel et al. 2003
E. coli		#1434	A. Collmer	pCPP3283	Badel et al. 2006
E. coli		#1406		pME6010-ORF43	this study
E. coli		#1552		pHopM1del	this study
E. coli		#1283	J. Greenberg	pBAV150-5: GFP fusion	Vinatzer, et al. 2006
E. coli		#1284	J. Greenberg	pBAV154-24; HA fusion	Vinatzer, et al. 2006
E. coli		#86	J. Greenberg	pBAV224	this study
E. coli		#460	J. Greenberg	pBAV226	Vinatzer, et al. 2006
E. coli		#849		*hopAO1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#858		*hopE1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#851		*avrPto1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#1093		*hopB1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#1132		*hopG1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#554		*hopAM1-1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#1131		*hopN1*$_{\text{Pto DC3000}}$ in pBAV226	this study
E. coli		#1094		*hopA1*$_{\text{Pto DC3000}}$ in pBAV226; cannot be transformed into any T1 or T1 mutants	this study
A. tumefaciens	C58C1	#1281	J. Greenberg		
A. tumefaciens	C58C1	#1545	plasmid from ABRC	ER-rk; mCherry marker for ER	Nelson, et al. 2007
A. tumefaciens	C58C1	#1546	plasmid from ABRC	G-rk; mCherry marker for Golgi	Nelson, et al. 2007

				apparatus	
A. tumefaciens	C58C1	#1548	plasmid from ABRC	px-rk; mCherry marker for peroxisome	Nelson, et al. 2007
A. tumefaciens	C58C1	#1300		*hopM1*$_{\text{Pto DC3000}}$ in pBAV150-5 (plasmid #1293)	this study
A. tumefaciens	C58C1	#1310		*hopM1*$_{\text{Pto JL1065}}$ in pBAV150-5 (plasmid #1298)	this study
A. tumefaciens	C58C1	#1311		*hopM1*$_{\text{Pto PT21}}$ in pBAV150-5 (plasmid #1302)	this study
A. tumefaciens	C58C1	#1312		*hopM1*$_{\text{Pto T1}}$ in pBAV150-5 (plasmid #1303)	this study
A. tumefaciens	C58C1	#1313		*hopM1*$_{\text{Pto ICPPB1108}}$ in pBAV150-5 (plasmid #1304)	this study
A. tumefaciens	C58C1	#1356		*hopM1*$_{\text{Pto T1}}$ in pBAV154-24 (plasmid #1328)	this study

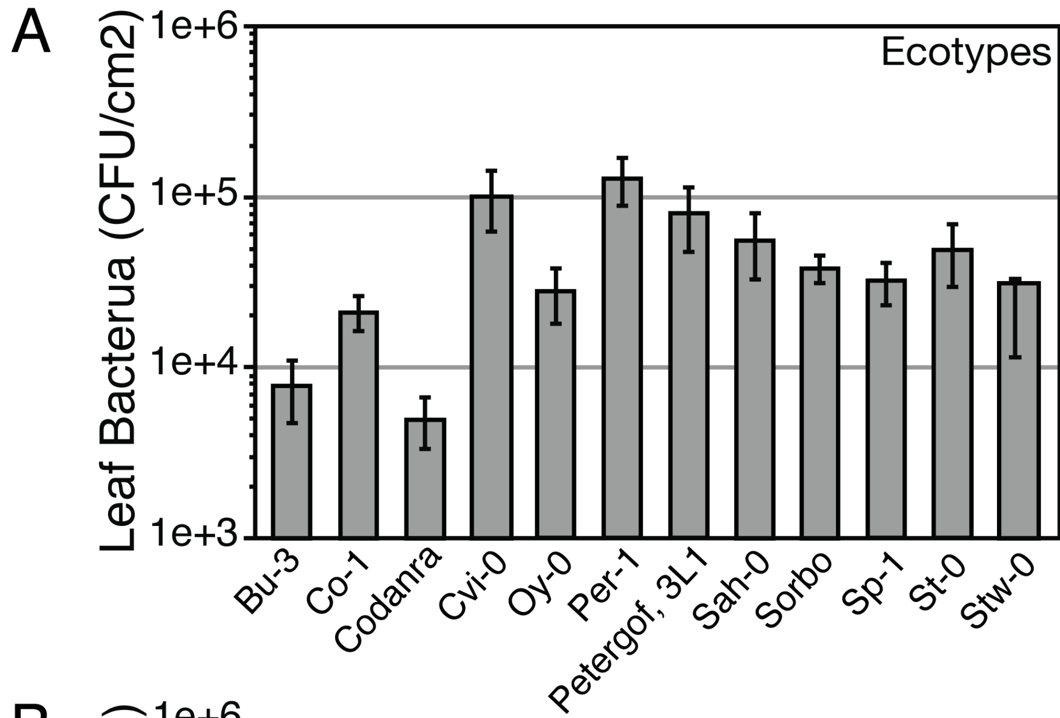
A. thaliana Resistance to Pto T1 is a kind of multi-layered non-host resistance

We previously showed that the recognition of *avrRpt2* by the *RPS2* resistance gene in *A. thaliana* ecotype Col-0 alone cannot explain T1's inability to grow in *A. thaliana* (Almeida, Yan et al. 2009). To unravel *A. thaliana* defenses to T1 beyond the *avrRpt2–RPS2* interaction, we disrupted *avrRpt2* in T1. The resulting strain is called T1' from here on. We confirmed that T1' does not cause an HR in ecotype Col-0 and that T1' expressing *avrRpt2* ectopically from a plasmid complemented T1' by re-establishing its ability to cause an HR in ecotype Col-0 (data not shown).

Since T1 and DC3000 are very closely related, the *A. thaliana* resistance to T1 may be ecotype–strain specific instead of non-host resistance. Therefore, we

tested T1''s ability to grow on twelve different *A. thaliana* ecotypes. T1' was incapable of growing to more than 10^5 colony forming units (CFUs) per square centimeter three days post inoculation in any of the twelve *A. thaliana* ecotypes tested (Fig 1A). This is over 100-fold less than the population density DC3000 reaches during the same period of time. Neither did T1' cause disease symptoms in any of the twelve ecotypes (data not shown). Therefore, resistance to T1 in the absence of the *avrRpt2–RPS2* interaction is not limited to ecotype Col-0 and *A. thaliana* resistance to T1 should be considered an example of non-host resistance.

To determine whether a single *A. thaliana* defense mechanism is responsible for *avrRpt2/RPS2*-independent resistance to T1, we tested T1' in *A. thaliana* mutants defective in different defense pathways: the salicylic acid signaling-defective mutants *sid2*, *pad4* and *npr1-1,* the PTI-defective mutants *fls2* and *efr1*, and the ETI-defective mutants *RAR1* and *SGT1*, and the *nho1* mutant defective in non-host resistance (Kang, Li et al. 2003). T1' was only able to grow significantly better in the *sid2* mutant (Fig 1B), which is severely defective in salicylic acid signaling (Wildermuth, Dewdney et al. 2001). However, T1' is still unable to cause disease symptoms in the *sid2* mutant. We conclude that resistance of *A. thaliana* to T1' is multi-layered and partially dependent on SA.


A
Ecotypes
Leaf Bacterua (CFU/cm2)
1e+6
1e+5
1e+4
1e+3
Bu-3
Co-1
Codanra
Cvi-0
Oy-0
Per-1
Petergof, 3L1
Sah-0
Sorbo
Sp-1
St-0
Stw-0

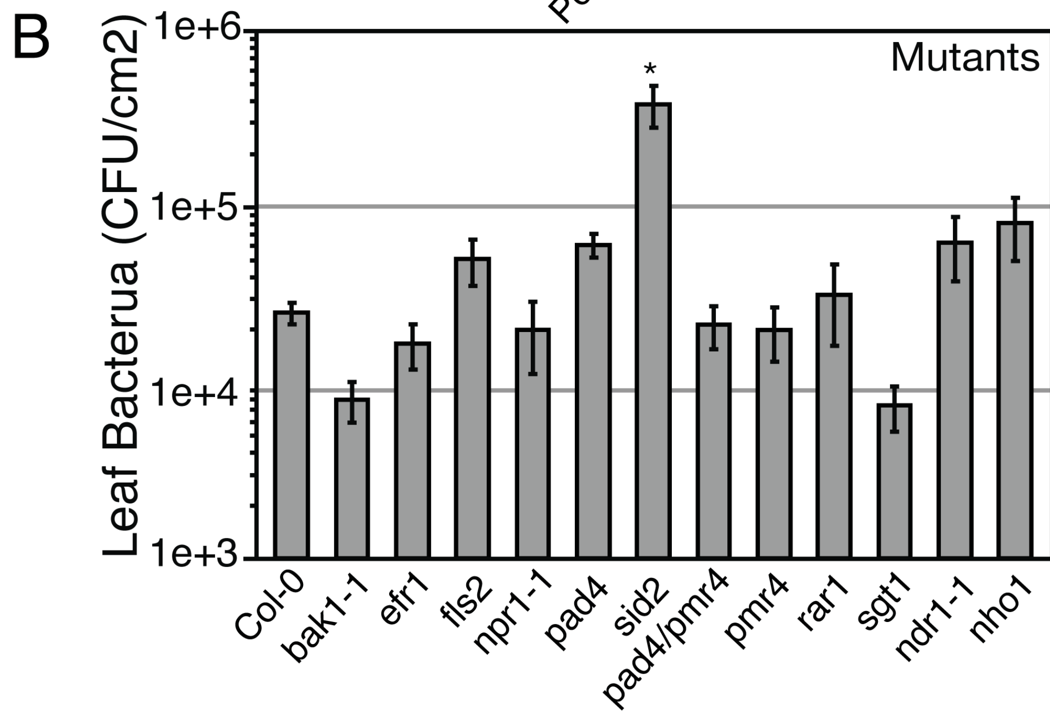

B
Mutants
Leaf Bacterua (CFU/cm2)
1e+6
1e+5
1e+4
1e+3
*
Col-0
bak1-1
efr1
fls2
npr1-1
pad4
sid2
pad4/pmr4
pmr4
rar1
sgt1
ndr1-1
nho1

Fig 1. *A. thaliana* resistance to T1 is not ecotype–pathovar specific and the non-host resistance is multi-genic.
Plants were spray infected with bacterium suspension at concentration of OD_{600}=0.1. Bacterial growth was determined three days post infection. (A) T1' does not cause disease on 12 randomly selected *A. thaliana* ecotypes and (B) T1' growth in *A. thaliana* defense mutants. T1' growth in *sid2* mutant is significantly higher than in Col-0 ($p<0.0001$) by student's t-test.

Comparisons of Arabidopsis defense responses to T1 and DC3000

To start finding out what else besides the *avrRpt2–RPS2* interaction does not allow T1 to grow and cause disease in *A. thaliana*, we compared the *A. thaliana* response to T1 with that to DC3000 (Fig. 2). First of all, we monitored T1 and DC3000 growth in *A. thaliana* Col-0 over a period of 8 days (Fig 2A). When spraying *A. thaliana* with bacteria at a concentration of OD_{600} of 0.1, T1 is able to persist in *A. thaliana* at a population density of around $1{\sim}3\times10^4$ CFU/cm^2, then slightly declines to about 1×10^3CFU/cm^2 on day 8, while DC3000 reaches its maximum population density of about 1×10^7CFU/cm^2 after 4 days of infection and then declines to 1×10^6CFU/cm^2 on day 8, at which time point *A. thaliana* leaves are very sick and almost dry.

Since callose deposition in plant cell walls is a generally accepted readout of PTI, we compared the strength of PTI by comparing callose deposition after leaf infiltrations with DC3000 and T1'. Instead of comparing callose deposition only 12-18 hours after infiltration of bacteria at an OD_{600} of 0.1 or higher as is commonly done (Nomura, Debroy et al. 2006; Ham, Kim et al. 2007), we decided to monitor callose deposition through 4 days, using two different inoculum concentrations, OD_{600} of 0.01 and 0.1. We confirmed earlier reports that DC3000 is able to suppress callose deposition when infiltrated at OD_{600} of 0.1 (Fig 2B) and 0.2 (data not shown). However, we noticed that callose deposition is not well

suppressed by DC3000 after infiltration at OD_{600} of 0.01, nor was DC3000 able to sustain suppression of callose deposition throughout infection (Fig 2B). It is important to compare callose deposition in response to DC3000 and T1', and we observed that T1' is able to suppress callose deposition when inoculated at bacterial concentration of OD_{600} of 0.1, albeit less effectively than DC3000 but more efficiently than DC3000 ΔCEL (Alfano, Charkowski et al. 2000), a strain that has been shown to have a reduced ability to suppress callose (Debroy, Thilmony et al. 2004). DC3000 Δcoronatine mutant (Uppalapati, Ishiga et al. 2007) does not produce phytotoxin coronatine, but still produces T3SS, shows similar suppression capability to DC3000 WT. Importantly, T1 elicits more callose depositions than T1', possibly due to AvrRpt2-induced ETI (data not shown), however when infiltrated at OD_{600} of 0.1, T1 elicits HR. Thus only T1' is shown here. T1', DC3000 and DC3000 mutants were all incapable of complete suppression of callose deposition at day 2 (Fig 2B) and beyond (data not shown). Since we observed both "big" and "small" callose depositions, and their morphologies are not always oval or "dots" (Supplementary Fig 1), we did not quantify the number of callose depositions.

We hypothesized that T1' might be able to suppress callose deposition but still triggers an otherwise stronger defense response than DC3000. Therefore, we used quantitative real-time PCR to measure the expression of several defense genes eight hours after inoculation of *A. thaliana* with T1, T1' and DC3000 (Fig. 2C). All three strains induced significant defense gene expression, except for *PR1*, whose transcription was suppressed by DC3000 10-fold compared to control, and for *ACD6*, which was not as significantly induced or repressed by any strain. For all genes that were tested, DC3000 induced a lower level of expression compared to both T1 and T1'. T1 induced stronger expression of *pad4* and *PR1* compared to T1', suggesting the involvement of these genes in ETI induced by *AvrRpt2*. T1 and

T1' induced similar levels of *FRK1*, *WRKY30* and *WRKY22* in most replicates (but not in the replicate shown in Fig 2C), suggesting these genes are not involved in AvrRpt2 elicited ETI. The dramatically lower expression level induced by DC3000 for *FRK1*, a gene induced early during PTI (Asai, Tena et al. 2002), suggests that DC3000 is either able to suppress PTI more efficiently than T1 or that T1 PAMPs trigger a stronger Arabidopsis immune response than DC3000 PAMPs.

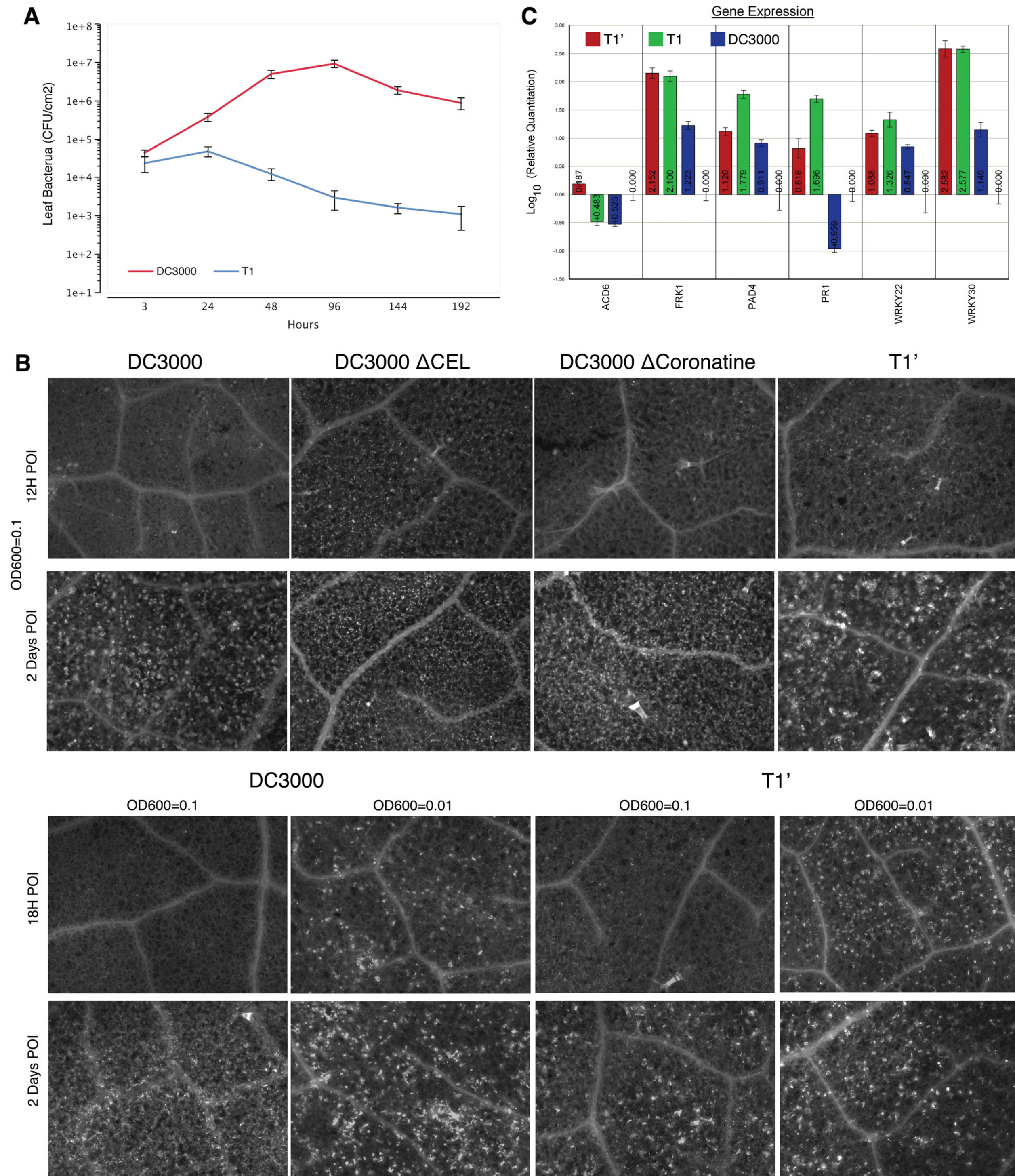


Fig 2. Comparison of T1 and DC3000 after inoculation in *A. thaliana*.

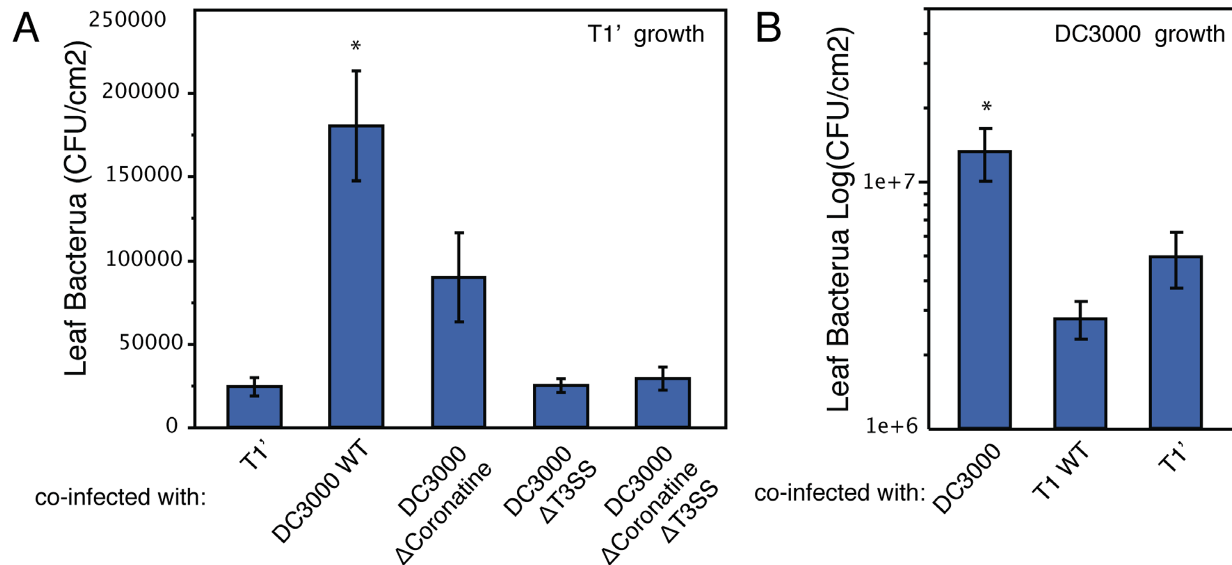
(A) Comparison of T1 and DC3000 growth in *A. thaliana* over time. T1 and DC3000 were sprayed onto *A. thaliana* Col-0 (bacterial concentration at OD_{600}=0.1). Similar

results were observed in two more replicates. (B) Comparisons of callose deposition elicited by 1) T1', DC3000 WT, DC3000 ΔCEL and DC3000 ΔCoronatine (bacterial concentration at OD_{600}=0.1) measured at time points 12H and 2 days POI; 2) T1' and DC3000 WT (bacterial concentration at OD_{600}=0.1 and OD_{600}=0.01) measured at time points of 18H and 2 days POI. (C) Comparison of relative expression level of defense genes induced by T1, T1' and DC3000 (bacterial concentration at OD_{600}=0.1) at 8 hours POI. Similar results were observed in another biological replicates and two technical replicates.

Co-infection of DC3000 with T1 slightly lowers DC3000 grows and dramatically increases T1 growth

Co-infection of T1' with DC3000 WT, a DC3000 ΔT3SS mutant, a DC3000 Δcoronatine mutant (DB29) (Uppalapati, Ishiga et al. 2007), and a DC3000 ΔT3SS Δcoronatine double mutant revealed that DC3000 WT can dramatically increase T1' growth. This increase appears to be mainly due to T3Es since the DC3000 ΔT3SS mutant and the DC3000 ΔT3SS Δcoronatine double mutant have very similar effects on the growth of T1' (Fig 3A). This result suggests that the lack of DC3000-specific T3Es may be the main reason why T1 cannot cause disease in *A. thaliana*. However, since growth of DC3000 was slightly decreased in the co-infection with T1', T1' may trigger some defenses that even DC3000 T3Es cannot completely suppress (Fig 3B). Moreover, we often observed that T1 ΔT3SS mutant reduced growth of co-infected DC3000 WT slightly more than a DC3000 ΔT3SS mutant or DC3000 ΔT3SS Δcoronatine double mutant. Although the reduction is small and statistically insignificant, it is overall reproducible (data not shown). This suggests that T1 is not only missing effectors to suppress PTI but that it is also eliciting a stronger PTI response than DC3000.

Fig 3. Co-infection of T1 and DC3000 revealed that T1 lacks effectors necessary for suppressing plant defenses and causing disease, and may also



elicit stronger PTI.

(A) T1' growth when co-infected with DC3000 is statistically significantly higher than all the others ($P<0.0001$) based on Tuckey-Kramer HSD. (B) Growth of DC3000 (marked with tetracycline resistance gene) co-infected with DC3000 (marked with kanamycin resistance gene) is statistically significantly higher than when co-infected with either T1 or T1' ($P<0.05$) based on student's t-test. These experiments were each repeated over 3 times with overall similar results.

Genome sequencing of *P. syringae* strains closely related to both T1 and DC3000 revealed conserved effector repertoire differences

We previously identified several DC3000 and T1 relatives that either cause disease in *A. thaliana*, such as DC3000, or cannot cause disease in *A. thaliana*,

such as T1 (Yan, Liu et al. 2008). We hypothesized that the *A. thaliana* pathogens may all share effectors necessary to suppress *A. thaliana* defenses and that these effectors are absent from the non-*A. thaliana* pathogens. On the other hand, the non-*A. thaliana* pathogens may share effectors that trigger *A. thaliana* defenses and are missing from all *A. thaliana* pathogens. By comparing the genomes of multiple *A. thaliana* pathogens and non-pathogens, it should be possible to identify these conserved differences in effector repertoires between *A. thaliana* pathogens and non-pathogens, and then test these effector differences experimentally for their roles in either suppressing or triggering *A. thaliana* defenses.

Since in our previous publication (Yan, Liu et al. 2008) we had mainly evaluated host range of strains based on symptoms and we expanded our strain collection since then, we extended multi-locus sequence typing (Maiden 2006; Yan, Liu et al. 2008) to additional T1 and DC3000 relatives and built a new phylogenetic tree (Fig 4A). We then picked one strain from each sequence type (ST) to test their growth when sprayed on two *A. thaliana* ecotypes, Col-0 and Mt-0 (Fig 4B). We decided to sequence the genomes of six strains that do not cause disease in *A. thaliana*: four additional Pto strains that have the same ST as T1 (NCPPB1108, LNPV17.41, Max4, and K40), one Pto strain with a different ST (JL1065), and the *P. syringae* pv. *antirrhini* (*Pan*) 126 strain. We also sequenced the genomes of three strains that cause disease in *A. thaliana*: *P. syringae* pv. *maculicola* (*Pma*) M3, M6, and F1. These strains are highlighted in Figure 4 and listed in Table 2. We sequenced these strains using Illumina sequencing technology. After assembling the draft genomes (see Table for description of assemblies), we compared the predicted T3E repertoires of these strains (supplementary table 1) (see materials and methods for genome assembly and T3E gene prediction). Although each draft genome is composed of a few hundred

contigs, most contigs end with repetitive sequences. Therefore, we are confident that we identified most, if not all, effectors in each strain.

We identified 9 effector genes that are present in all strains pathogenic in *A. thaliana* and absent from all strains non-pathogenic in *A. thaliana* (Table 3A). We identified four effector genes present in all strains non-pathogenic in *A. thaliana* but absent from all strains pathogenic in *A. thaliana* (Table 3B).

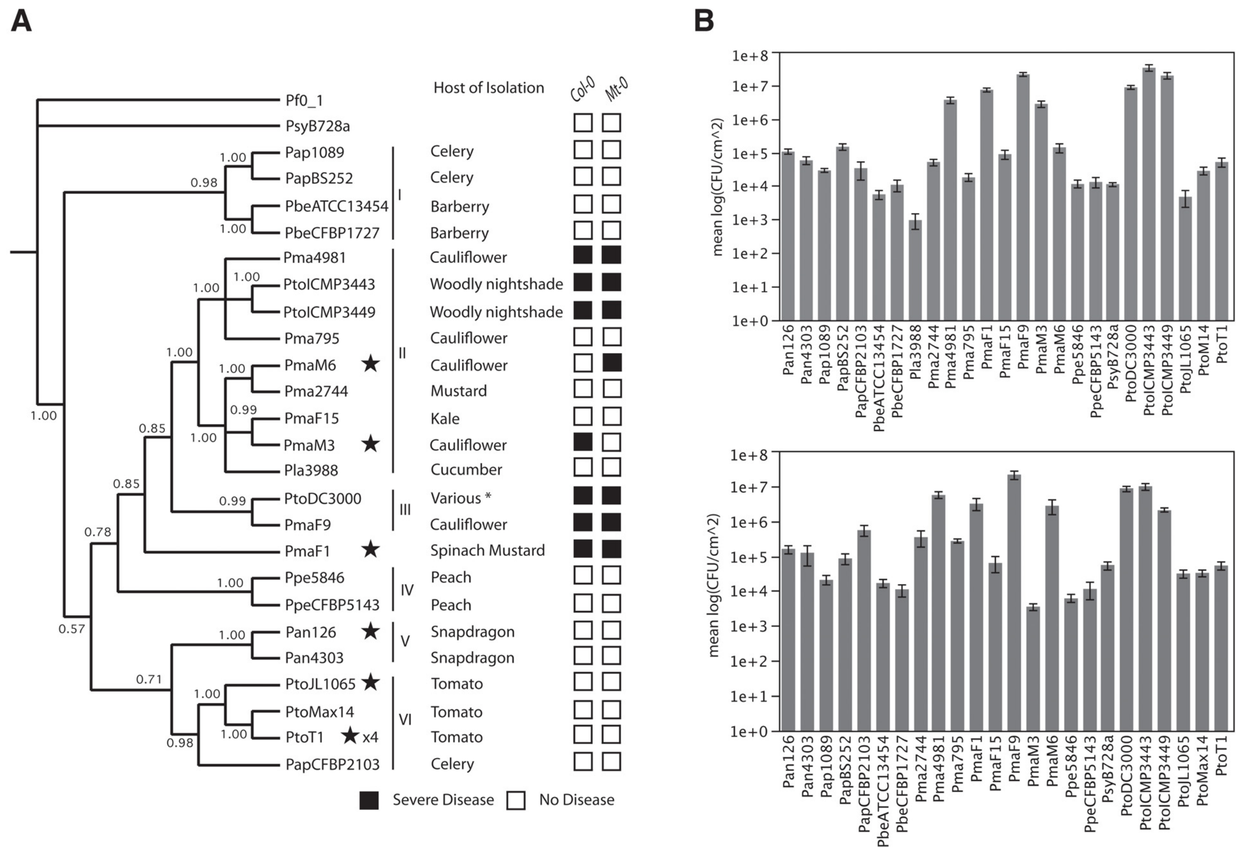


Fig. 4. Phylogenetic tree of expanded *P. syringae* strain collection and the abilities of representative strains of each ST to cause disease in *A. thaliana*.

(A) Phylogenetic tree built with Bayesian inference. (B) The abilities of strains representative of each sequence type (ST) to cause disease in *A. thaliana* ecotypes Col-0 and Mt-0. Black stars mark the strains that were selected for genome sequencing.

4 strains from the Pto T1 ST were selected for genome sequencing. Disease symptoms correlated with bacterial growth well. Symptoms were only observed when growth was higher than about 1×10^6 CFU/cm^2.

Strain	Coverage (X)	N50	Max length of contig	Genome size	Number of contigs	Scaffolding[b]
Pto JL1065	70	65097	149551	6338641	346	Yes
Pan 126	85	110884	265136	6178611	145	Yes
Pma F1	86	54025	143721	6301349	323	Yes
Pma M3	49	31589	96642	6316889	497	Yes
Pma M6	43	39142	131769	6014638	338	Yes
Pto LNPV17.41	75[a]	62385	239369	6157021	350	No
Pto K40	32[a]	25354	104626	6254280	582	No
Pto NCPPB1108	43[a]	46775	153603	6182607	304	No
Pto Max4	28[a]	11264	53242	6209056	1176	No

Table 2. Characteristics of genome assemblies. [a]: effective coverage calculated based on the number of reads that were actually used in the assembling process. All other coverage data are based on total reads generated by the sequencing machine. [b]: Scaffolding: "yes" means Velvet used pair-end information from Illumina raw data to connect contigs, and there are nucleotides labelled as "N" in the draft genome sequences due to the scaffolding process; "no" means pair-end information was not used, and every single read in the raw data was treated as independent by Velvet, and there is no nucleotide labelled as "N" in the draft genome sequences.

A

effectors	DC3000	F1	M3	126	JL1065	Max4	K40	LNPV17.41	1108	T1
hopA1	Yes	Yes	Yes	Yes	No	No	No	No	No	No
hopB1	Yes	Yes	Yes	No	No	No	No	No	No	No
hopM1	Yes	Yes	Yes	Yes	Trunc.	Trunc.	Trunc.	Trunc.	Trunc.	Trunc.
avrPto1	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No
hopE1	Yes	Yes	Yes	No	No	No	No	No	No	No
hopG1	Yes	Yes	Yes	No	No	No	No	No	No	No
hopK1	Yes	No	No	No	No	No	No	No	No	No
hopN1	Yes	Yes	Yes	No	No	No	No	No	No	No
hopU1	Yes	No	Yes	No	No	No	No	No	No	No
hopX1	Yes	No	No	No	No	No	No	No	No	No
hopAA1	Yes	Yes	Yes^	Yes	Yes	Yes^	Yes	Yes	Yes	No
hopAM1	Yes	Yes	Yes	No	Yes	No	No	No	No	No
hoAO1	Yes	Yes	Yes	No	No	No	No	No	No	No

B

effectors	DC3000	F1	M3	126	JL1065	Max4	K40	LNPV17.41	1108	T1
hopS1	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
hopAG1	Yes[PT]	Yes^	Yes	Yes	Yes	Yes[PT]	Yes	No	Yes	Yes
hopAH1	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
hopAI1	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
hopAS1	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes[SD]	Yes
hopAE1	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
hopW1	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
avrA1	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
avrRpt2	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes
avrD1	No	Yes	Yes	No	Yes	Yes	Yes^	Yes	No	Yes

Table 3. List of conserved effectors among strains that are (A) pathogenic in *A. thaliana*; (B) typical tomato strains and non-pathogenic in A. thaliana. Effectors in bold and underlined are the ones likely important for host range differences. PT: premature stop. SD: slightly different. ^: broken by contig breaks. Color coding: orange corresponds to strong ability to cause disease in *A. thaliana*; yellow corresponds to moderate ability to cause disease in *A. thaliana*; green corresponds to inability to cause disease in *A. thaliana*.

HopAS1 in T1 elicits ETI

The effector *hopAS1*, one of the four effectors absent from all *A.thaliana* pathogens but present in all non-*A. thaliana* pathogens, was found to trigger ETI in *A. thaliana* (Kee et al, in preparation) while no ETI activity was found for the other three effectors. A Δ*hopAS1* Δ*avrRpt2* double mutant (T1*) was thus constructed from T1' to obtain a strain that does not elicit any ETI in *A. thaliana* (Kee et al, in preparation). Therefore, we used T1* instead of T1' in most of the subsequent analyses. T1* grows to slightly higher population density than T1' in *A. thaliana* but still does not cause any disease symptoms confirming our initial hypothesis that *A. thaliana* resistance to T1 goes beyond ETI (Kee et al, in preparation).

DC3000 effectors do not significantly increase T1' or T1* growth in *A. thaliana*

We added effector genes present in all *A. thaliana* pathogens from Table 3A to T1, T1' and T1*. Due to technical difficulties, some genes were not transformed into these strains. Then we tested whether individually expressing these effectors in T1, T1'or T1* would increase or decrease bacterial growth in *A. thaliana* and/or tomato. We mainly tested T1' and T1* strains in *A. thaliana* and T1 strains in tomato (Fig 5). Unexpectedly, we did not observe significant growth increase in *A.*

thaliana of any T1' or T1* strains expressing effectors typical of *A. thaliana* pathogens (Fig 5A and 5B), nor did we observe disease symptoms (data not shown). However, when the same effectors were expressed in T1, growth of the resulting strains in tomato (Fig 5C) and disease symptoms were reduced (Fig 5D). We also tried co-infection to combine all T1* strains expressing individual DC3000 effectors, but did not see growth or disease symptoms comparable to DC3000 even when each T1* strain was inoculated at the concentration of OD_{600} of 0.02. However, we did observe chlorosis.

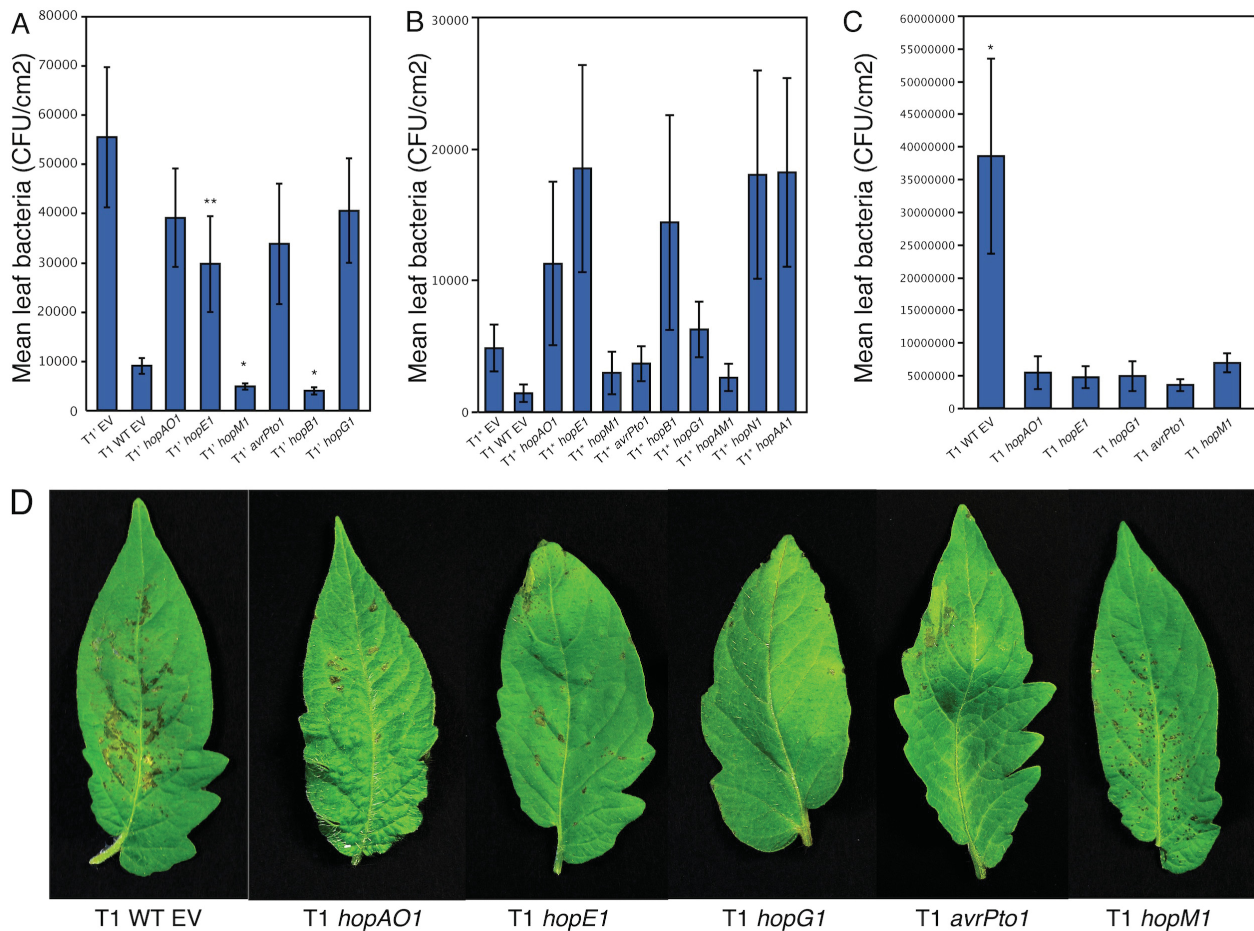


Fig 5. Growth of T1 WT, T1' and T1* with individual DC3000 effectors in plants.

(A) T1' with DC3000 effector genes infected in *A. thaliana* by infiltration at a concentration of OD600 of 0.0001. EV: empty vector. * & **: Statistically smaller compared to growth of T1' EV. *: $P \leq 0.0001$. **: $P < 0.05$. Student's t-test was used. Data from over 8 experiments are combined to cover all the strains, and to increase power, however also resulting in larger standard error. T1' EV are included in all the 8 experiments and in every case has the highest growth.

(B) T1* with DC3000 effectors infected in *A. thaliana* by spray (bacterial concentration at OD_{600}=0.1). The mean growths of different strains are not statistically significant, although more replications are needed to increase the power of this experiment.

(C) Growth of T1 WT EV and T1 WT with individual DC3000 effector genes in tomato cultivar Chico III. T1 *hopAM1-1* was not included in the chart due to inconsistent data, while the other T1 strains with DC3000 effector genes grew consistently less than T1 WT EV in tomato in four replications. Tomato was spray infected (bacterial concentration at OD_{600}=0.1). Growth data were sampled between 3-5 days POI when symptoms are well developed. The mean growths of different strains are statistically significantly different based on one-way ANOVA ($P=0.0017$). T1 WT EV growth was statistically significantly higher than the rest of the strains based on Tukey-Kramer HSD (for each pair $P<0.01$ except for T1 hopM1, $P=0.0121$).

(D) Symptom reduction caused by DC3000 effectors in tomato cultivar Chico III. Representative leaves with moderate disease symptoms judged within each strain are shown.

The presence of *hopM1* is actively selected against in typical Pto strains

We found that *hopM1* is either disrupted or has a deletion in all *Pto* strains besides DC3000: a 180 bp in frame internal deletion is present in Pto JL1065, an A-to-T mutation at 463 bp results in a premature stop at codon 155 in Pto T1 and in other 80 similar strains (including sequenced strains Max4, K40 and LNPV17.41), a 1bp internal deletion results in a frame shift and premature stop at codon 212 in Pto NCPPB1108, and a 5bp internal deletion results in a premature stop at codon 394 in four other Pto strains. See Figure 6A for a schematic view of these different *hopM1* alleles. These independent mutations strongly suggest a strong selection pressure for the loss of *hopM1* in Pto strains (Cai et al. with Yan as co-author, in review).

Truncated HopM1 alleles lost their ability to cause cell death in tomato and *N. benthamiana*

It was previously shown that HopM1$_{\text{Pto DC3000}}$ causes severe leaf necrosis in transgenic *A. thaliana* after expression is induced (Nomura, Melotto et al. 2005) and causes necrosis and chlorosis in a number of other plant species, including several tomato cultivars and wild tomato relatives when transiently expressed (Wroblewski, Caldwell et al. 2009). Also HopM1$_{\text{Psy B728a}}$ causes cell death when transiently expressed in tobacco (Vinatzer, Teitzel et al. 2006). We speculate that the truncated *hopM1* alleles and the *hopM1*$_{\text{Pto JL1065}}$ allele with the internal in frame deletion may have lost the ability to induce necrosis on tomato. To test this, we transiently expressed all identified *hopM1* alleles in both tomato and *N. benthamiana.* We observed that, as expected, only HopM1$_{\text{Pto DC3000}}$ is capable of causing cell death in the two plant species (Fig 6B and 6C). Also when delivered by *Pseudomonas fluorescens* Pf0-1 EtHAn (Thomas, Thireault et al. 2009), cell death was only elicited by HopM1$_{\text{Pto DC3000}}$ (data not shown). The ability of an

effector to induce cell death when stably or transiently expressed *in planta* may either indicate that it triggers a defense response accompanied by an HR or that the effector causes cell death as part of its virulence function and that the observed cell death may contribute to symptom formation during infection. Since most Pto strains isolated from diseased tomatoes in the field have *hopM1* alleles that lost the ability to trigger cell death, we hypothesize that the cell death triggered by *hopM1*$_{\text{Pto DC3000}}$ might in fact be a results of a defense response and that Pto strains are thus under selection pressure to lose functional full-length *hopM1*.

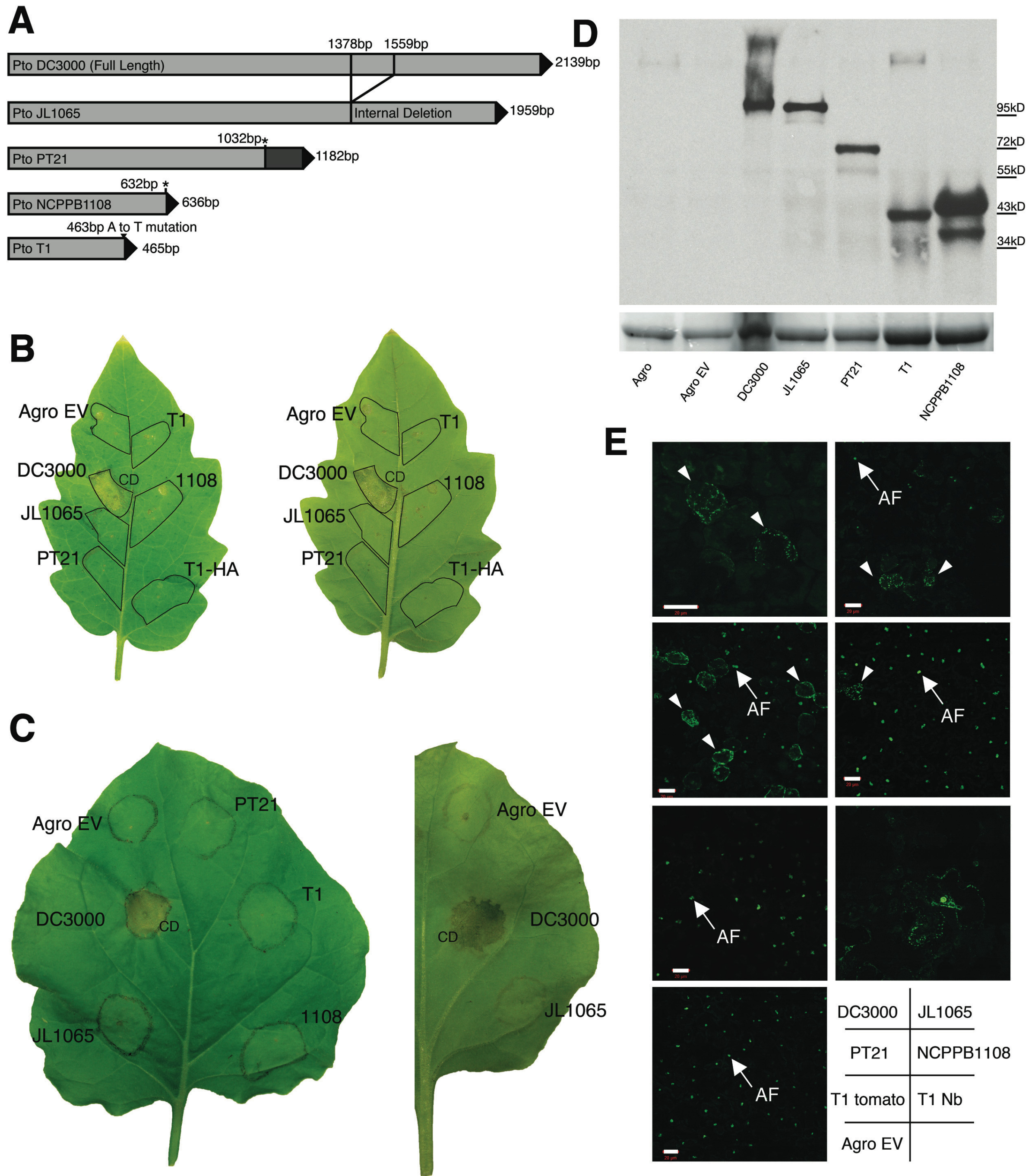


Fig 6. Different HopM1 alleles have different abilities to cause cell death but the same cellular localization.

(A) Schematic view of different *hopM1* alleles. Different abilities of HopM1 alleles to elicit cell death when transiently expressed in tomato (B) and *N. benthamiana* (C). The front and back views of the same leaf are shown. (D) Western blot of proteins extracted

from *N. benthamiana* leaf infiltrated with Agrobacterium for transient expression of *hopM1* alleles. All the transient expression vectors were in Agrobacterium strain C58C1. HopM1 alleles were fused with GFP. HopM1::GFP alleles were probed by anti-GFP antibody. Bromophenol blue staining of RuBisCO was shown as loading control below the western blot. (E) Transient expression of *hopM1*::*GFP* alleles in tomato and *N. benthamiana* (only *hopM1*$_{Pto\ T1}$::*GFP*) were detected by confocal microscopy. White arrow heads point to transformed cells with GFP activities. AF, autofluorescence in tomato. White bar equal to 20μm.

The inability of truncated HopM1 alleles to cause cell death is not due to different cellular localization

We wondered whether the inability of truncated HopM1 alleles to cause cell death is due to a change of cellular localization. It was previously shown by cellular fractionation that HopM1$_{Pto\ DC3000}$ localizes to vesicles, which were hypothesized to be involved in the delivery of toxins and cell wall components to the extracellular spaces as part of the PTI response. HopM1 is believed to interfere with this cellular trafficking (Nomura, Debroy et al. 2006). We cloned and fused the *hopM1* alleles described above to GFP and transiently expressed these genes in *N. benthamiana* and tomato, and determined the cellular localization of the GFP fusion proteins by confocal microscopy (Fig 7A and 6E). We observed many fluorescent "little bubbles" in transformed cells (Fig 7A; Fig 6E, white arrow heads). There are no difference between different HopM1::GFP alleles, except that we were able to see HopM1$_{Pto\ T1}$::GFP in these "little bubble" structures only in *N. benthamiana* but not in tomato. This suggests that the HopM1 N-terminus is sufficient to determine the cellular localization of this effector protein. We then picked HopM1$_{Pto\ JL1065}$::GFP for co-localization with mCherry labeled organelle markers (Nelson, Cai et al. 2007). The results clearly show that HopM1 does not localize either to the Golgi apparatus or the ER (Fig 7B). Since HopM1 proteins

are pre-made before being injected into plant cells, it is in fact unlikely they would go through the ER and the Golgi apparatus to enter vesicles. Since we found that HopM1$_{Pto\ JL1065}$::GFP does not co-localize with the peroxisome either, our data support the conclusion made by Nomura and co-workers that HopM1 localizes to vesicles (Fig 7) (Nomura, Debroy et al. 2006).

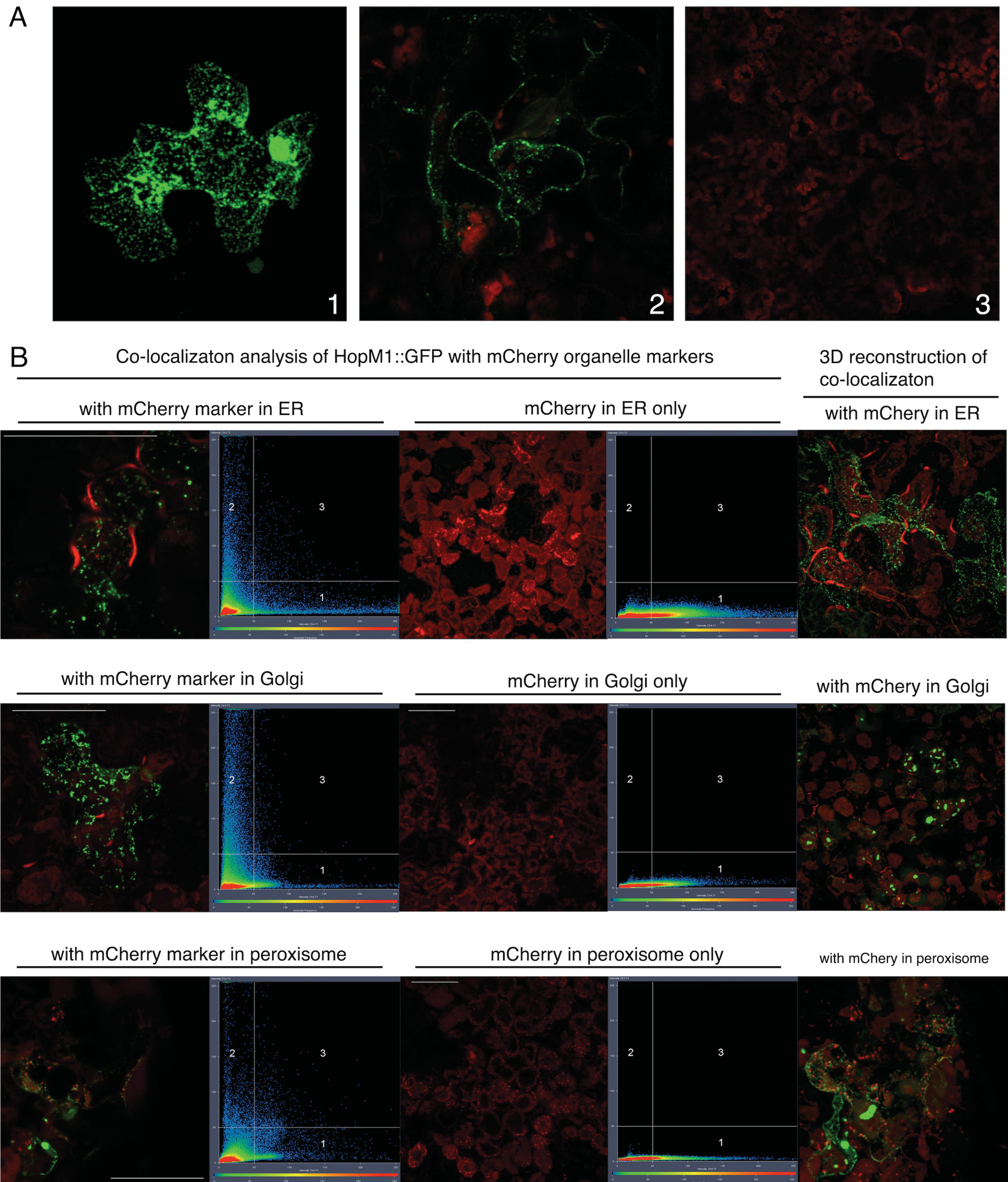


Fig 7. Cellular localization of HopM1.

(A) Cellular localization of HopM1::GFP in *N. benthamiana*. 1: 3D reconstruction of a *N. benthamiana* cell that expresses *hopM1*$_{\text{Pto JL1065}}$::GFP. 2: *N. benthamiana* cells that expresses *hopM1*$_{\text{Pto DC3000}}$::*GFP*. 3: Negative control. (B) Co-localization data showing

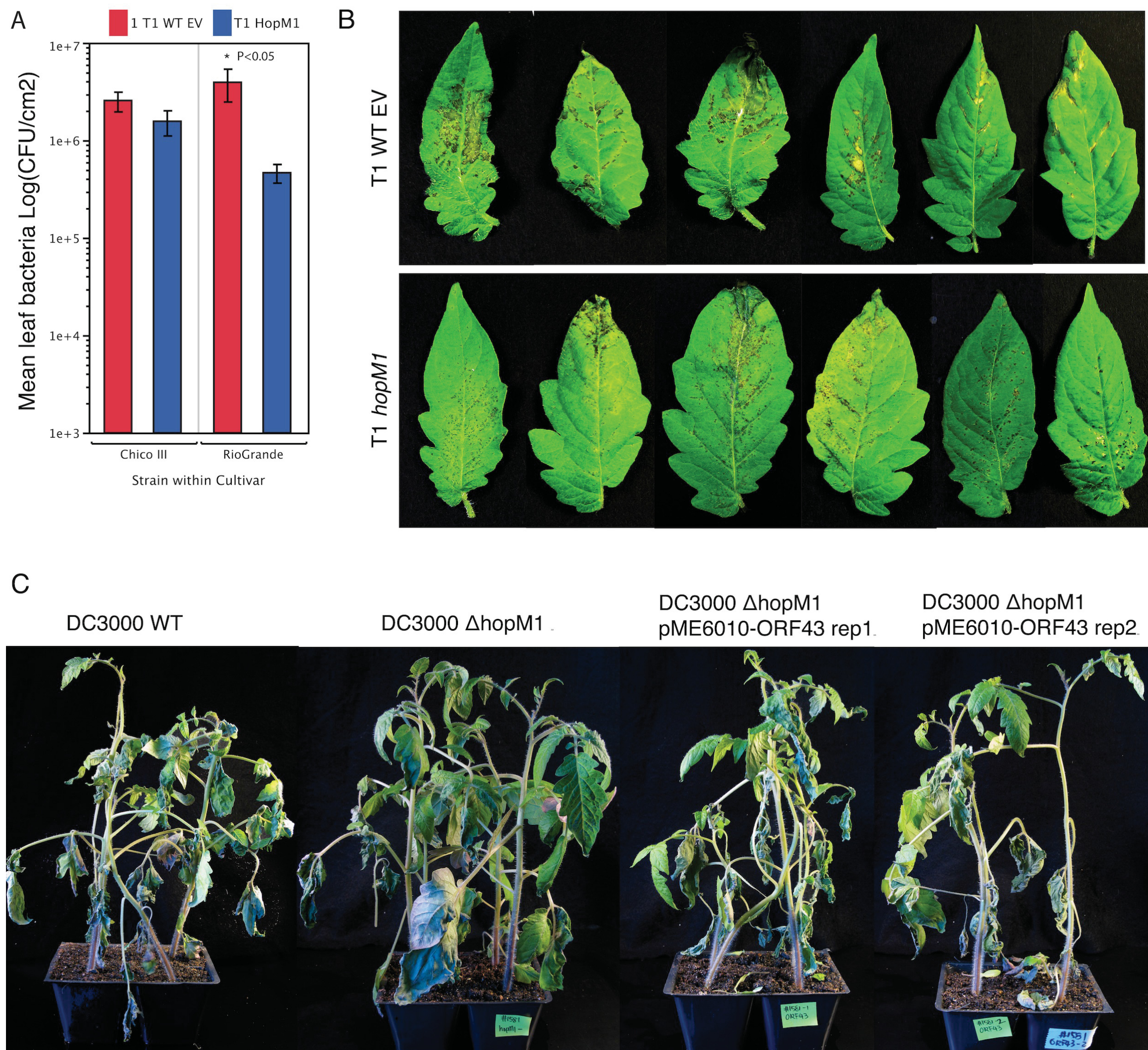
no co-localization with ER, Golgi or Peroxisome. The chart showing pixel co-localization has Y-axis corresponding to GFP channel intensity, X-axis corresponding to Red channel intensity. A large number of pixels in area 3 suggest indicate co-localization. White bar equals to 50µm. The last picture of each row is 3D reconstruction of z-stack showing minimal co-localization.

HopM1 reduces T1 growth in tomato, reduces bacterial speck size, and causes wilting of tomato

In order to further investigate the effects of HopM1$_{\text{Pto DC3000}}$ in tomato, and determine why there might be selection for loss of full-length *hopM1* in Pto strains, we expressed full-length *hopM1*$_{\text{PtoDC3000}}$ in T1. We found that HopM1$_{\text{Pto DC3000}}$ significantly reduced T1 growth in two tomato cultivars, Chico III and Rio Grande (Fig 8A), increased the number of symptomatic leaves, and also increased the number of often-smaller bacterial specks per leaf (Fig 8B). These results suggest that the increased number of bacterial specks may be the result of a localized HR, which limits bacterial growth and the expansion of specks. We also observed that when DC3000 WT is sprayed at high dose on tomato plants the entire plant wilts within 2 days while a *hopM1* deletion mutant of DC3000 causes a more moderate response (Fig 8C). No wilting is observed with T1 although T1 grows to higher population densities than DC3000 on the same tomato cultivars (data not shown). This observation strengthens the hypothesis that *hopM1* causes an HR-like response in tomato (Fig 6B), and that *hopM1* has an avirulence activity in tomato.

Figure 8. HopM1 effects in tomato when delivered by either T1 or DC3000.

(A) HopM1 reduces T1 growth in two tomato cultivars. Although in this particular dataset the reduction is not statistically significant for Chico III, in other replications it is often statistically significant, such as shown in Fig 5. (B) HopM1 delivered by T1



increase the number but reduce the size of bacterial specks in tomato cultivar Chico III. (C) DC3000 Δ*hopM1* has significantly reduced ability to cause wilting in tomato when sprayed at bacterial concentration of OD_{600}=0.2. The ability to cause wilting is complemented by pME6010-ORF43 containing both *hopM1* and chaperon *shcM*. Two complemented strains showed the same results. Plants were covered in plastic bags for a day after infection. Pictures were taken 3 days after infection.

HopM1 alone allow *Pseudomonas fluorescens* to grow to pathogen-level population density in both *N. benthamiana* and *A. thaliana*

As a further means to study the defense triggering and/or symptom inducing virulence activity of *hopM1* in *N. benthamiana* and *A. thaliana*, we delivered HopM1 from *P. fluoresens* Pf0-1 EtHAn. We observed that the cell death area in *N. benthamiana* was not as dry as the cell death reaction we had seen when *hopM1* was expressed after Agrobacterium-mediated transient expression. In *A. thaliana* inoculation with *P. fluoresens* Pf0-1 EtHAn expressing *hopM1* caused chlorosis and necrosis (Fig 9B). As with the Agrobacterium-mediated expression assays these symptoms may have been due to a *hopM1*-triggered plant defense response or an effect of *hopM1* virulence action. Intriguingly, we found that growth of Pf0-1 expressing *hopM1* in *A. thaliana* and *N. benthamiana* was almost on par with pathogenic *P. syringae* strains, while Pf0-1 EtHAn not expressing *hopM1* was 2-3 log lower. However, a high initial inoculum was necessary to obtain this result. These data suggest that either *hopM1* triggers a defense response accompanied by cell death and that Pf0-1 benefits from the dead cells and – since is is a detritivore – uses them as source of nutrition, or that *hopM1* suppresses plant defenses so efficiently to allow Pf0-1 to grow. Plant defense gene expression assays should help determine which of the two hypotheses is correct.

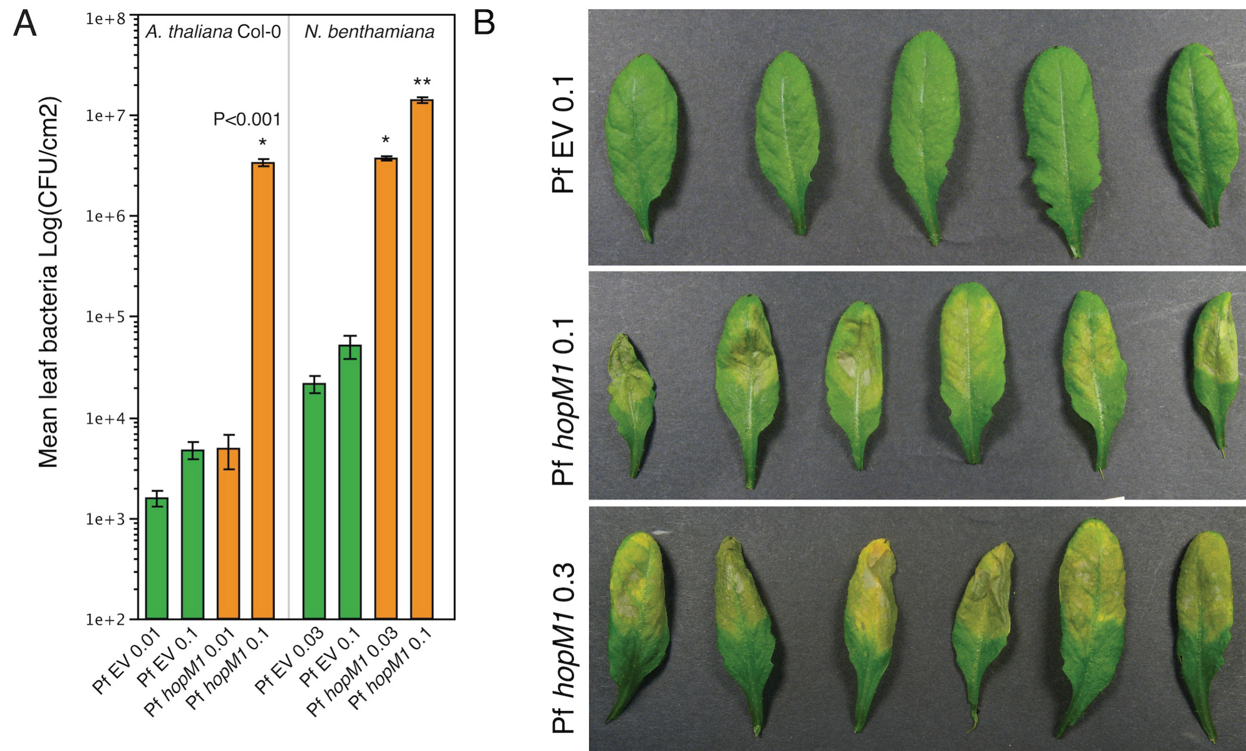


Figure 9. Growth of *P. fluorescens* Pf0-1 EtHAn with HopM1 and symptoms it causes in *A. thaliana* and *N. benthamiana.*

(A) Growth measured 3 days POI by infiltration at various concentration of OD_{600} ranging from 0.01 to 0.3 as indicated in the figure. For comparisons within *A. thaliana*, Pf0-1 EtHAn with hopM1 infiltrated at OD_{600} of 0.1 grew statistically significantly higher than the others (P<0.0001 by one-way ANOVA). For comparisons within *N. benthamiana*, ** means statistically significantly higher than all the other strains (P<0.0001); * means statistically significantly higher than all the other strains except the one marked with ** (P<0.0001); statistical test are done by student's t-test.

(B) Symptoms of *A. thaliana* leaves 3 days POI infiltrated with Pf0-1 EtHAn with empty vector (EV) or pME6010-ORF43. Bacterial concentrations of OD_{600} are indicated in the figure, ranging from 0.03 to 0.3. *A. thaliana* leaves infiltrated with Pf0-1 EtHAn EV at concentration of OD_{600}=0.01 was not shown because they look at least as healthy as the first row of leaves shown above.

d. Discussions

Non-host resistance is difficult to dissect genetically because of its usually multi-genic basis. On the other hand, exactly for the same reason, i.e. its multi-genic basis, non-host resistance is difficult to overcome by a pathogen. It is thus very attractive to plant breeders and plant scientists whose goal is to breed crops with durable resistance.

Since the multi-genic basis of non-host resistance makes it difficult to identify the individual genes at its basis, we decided to approach non-host resistance with the comparison of two closely related pathogen strains that can either cause disease in *A. thaliana*, i.e., strain DC3000 or not, i.e., strain T1. Identifying the differences between the two strains and experimentally characterizing the role of these differences in the strains' interaction with *A. thaliana* should then allow us to explore *A. thaliana* non-host resistance against T1.

However, first of all we needed to confirm that the resistance of *A. thaliana* to T1 was in fact multi-genic non-host resistance. We had previously shown that *avrRpt2-RPS2* ETI was not the only basis of *A. thaliana* resistance to T1 (Almeida, Yan et al. 2009). Here we show that *A. thaliana* resistance to T1' (T1 with a disruption of *avrRpt2*) is still multi-genic non-host resistance since it is active in all tested ecotypes and all tested defense mutants. Even deleting *hopAS1*, the only other T1 effector gene found to elicit ETI in *A. thaliana*, from T1' (creating T1*) did not abolish resistance (Kee et al, in preparation) suggesting that *A. thaliana* resistance to T1 goes beyond ETI. Since we found T1 and T1' (T1* still needs to be tested) to induce a higher level of defense genes in *A. thaliana* than DC3000 and we found that DC3000, but not a T3SS-deficient DC3000 mutant, could increase growth of T1' when co-infiltrated in *A. thaliana* leaves, we hypothesized

that DC3000 secretes effectors into *A. thaliana* that can efficiently interfere with *A. thaliana* defenses while T1 cannot. Moreover, a slight decrease in DC3000 growth when co-inoculated with T1' and a more conspicuous reduction of DC3000 growth when co-inoculated with a T3SS-deficient T1 strain than when co-inoculated with a T3SS-deficient DC3000 strain, suggested that T1 might also encode PAMP(s) that trigger a stronger immune response than PAMPs encoded by DC3000. T1 and DC3000 have exactly the same peptide sequence in the known PAMPs flg22, elf18, and csp (data not shown). However, many more PAMPs may exist and could have allelic differences between T1 and DC3000. Also, allelic differences in the T3SS itself and in T3SS helper proteins between DC3000 and T1 could contribute to the different strength of the immune response triggered by the two strains. We found, for example, that the protein sequence of the helper protein HrpZ in DC3000 and T1 differs in one amino acid (data not shown). HrpZ1 has long been known to trigger a plant immune response (Wei, Laby et al. 1992). Therefore, HrpZ of T1 could trigger a stronger immune response than HrpZ of DC3000.

To further characterize the *A. thaliana* response to T1', we compared *A. thaliana* callose deposition after inoculation with T1' and DC3000. Although T1' was able to suppress callose deposition when infiltrated at a concentration of OD_{600} of 0.1, it is clear that the suppression is only transient. The fact that all the strains we tested, including DC3000, were incapable of efficient suppression of callose deposition by day 2 post inoculation, suggests that callose deposition is ineffective against bacterial pathogens, at least when the pathogen has already taken control of the plant cell. Although some effectors, such as those encoded in the CEL, clearly have roles in suppression of callose deposition (Debroy, Thilmony et al. 2004; Nomura, Debroy et al. 2006; Ham, Majerczak et al. 2009), theses effectors are either incapable of maintaining such suppression, or plant cells have alternative pathways that will eventually activate this defense mechanism through converging

pathways. It is, however, noteworthy that DC3000 populations at day 2 are over 100 times higher than T1' populations. Although Figure 2B showed similar amount of callose deposition induced by both T1' and DC3000 at 2 days post-inoculation, it is thus still likely that DC3000 is more efficient at suppressing callose deposition.

It is also worth noting that the excessive amount of callose deposition at later time points during infection may be favorable to pathogen growth, since the whole process may use a lot of energy and resources which could have been used for more effective defense mechanisms. Consequently, pathogens may even actively induce callose deposition at a later infection stage. Regardless, our results highlight that the often-used callose assay at a very early time point after infiltration of pathogens at high dose only gives a limited view of callose-deposition dynamics and need to be interpreted with caution.

We did not quantify the number of callose deposits because 1) we observed various morphologies of callose deposition (Supplementary Fig 1), thus simply counting the dots would have been misleading; 2) at 2 days post-infection, callose depositions are technically difficult to quantify even using sophisticated software; 3) even without quantification, the pictures shown in Figure 2 are a good representation of the qualitative differences observed between strains, time points, and inoculum concentrations.

To identify the DC3000 effectors that can interfere with the defenses that T1 triggers, we then set out to identify those DC3000 effectors that are conserved in other strains pathogenic in *A. thaliana* closely related to DC3000 and T1, but missing from relatives unable to cause disease in *A. thaliana*. Sequencing the genomes of eight such strains (three strains that cause disease in *A. thaliana* and six strains that do not cause disease in *A. thaliana*), we identified 9 effectors presumably important for specialization in *A. thaliana*. We tested the effects of

these genes when they are ectopically expressed in T1. Individually, none of these effectors allowed T1, T1' or T1* to grow significantly better in *A. thaliana*, and they slightly reduced growth in tomato, although unfortunately we were unable to transform every effectors into all these strains. We expected the reduction in growth on tomato since T1 is a typical tomato pathogen and presumably has an effector repertoire fine-tuned for tomato. DC3000 instead causes disease also on *Brassicaceae* and thus presumably has an effector repertoire that is sub-optimal for tomato. Hence we expected some DC3000 effectors to trigger defenses in tomato and to reduce growth.

The difficult question to answer is why none of the DC3000 effectors increased growth of T1* in *A. thaliana*. Although we have not yet confirmed expression of these effectors in T1, all genes have been used in Agrobacterium-mediated transient expression assays and were confirmed to be cloned correctly. We expressed effectors from the npt2 promoter from the same vector that was previously used and that was found to express effectors properly (Vinatzer, Teitzel et al. 2006). However, it is very well possible that T1 down-regulates DC3000 effectors post-translationally. It has been found that the DC3000 effector AvrPtoB is expressed to only very low protein levels in T1 although the mRNA is expressed as expected (Lin, Abramovitch et al. 2006). We are thus planning a comparative proteomics study between T1 and DC3000 to investigate the possibility that effectors in T1 accumulate to different levels compared to DC3000.

Another explanation for the inability of DC3000 effectors to increase growth of T1 may be that T1 does not provide an adequate genetic background. Maybe, these effectors need other DC3000 effector(s) to function properly or to suppress defenses they trigger. Alternatively, they may only function in a strain that is already more virulent in *A. thaliana* than T1. T1 virulence may be under a

virulence threshold that is needed for these effectors to have additive virulence effects.

It is also difficult to explain why these DC3000 effectors all actually reduced T1' growth in *A. thaliana*. We speculated above that they might need other DC3000 effector(s) to suppress defenses they trigger. And these variable effectors, or exchangeable effectors exclusive to strains sharing the same host range, may act on top of conserved or core effectors to allow compatibility with specific host. Their functions in plant cells may be more delicate and specific. Additionally, adding one of these exchangeable effectors to T1 could disturb the delicate fine-tuned interactions between the existing effectors. However, except *hopM1* in pME6010-ORF43, all other effector genes were not expressed under their native promoters. This could be a problem since these ectopically expressed effectors are constitutively expressed. We also need to test whether adding an effector gene that T1 already has in the same vector would also cause reduction in growth. Therefore, we currently do not have enough evidence to confidently conclude the reasons behind the reduction, but the relatively small differences in effector repertoire between strains that share the same host support the notion that these strains only fine-tuned their effector repertoire during evolution, as acquisition of a cluster of additional effectors was not observed. The bottom line is, these effectors individually do not allow T1 to cause disease in *A. thaliana*.

Due to the limitation of current tools and because T1 is highly recalcitrant to transformation, we were unable to add more than one DC3000 effector to T1 at the same time. Nonetheless, we did try to mix individual T1* strains expressing DC3000 effectors and infect *A. thaliana* by both spraying and infiltration. Preliminary results are promising. We observed reproducible slight increase of growth for mixed T1* strains with individual DC3000 effector genes compared to T1* alone (data not shown), by both infiltration and spray infections. However, the

increase was too small to be statistically significant. The reason for this could be that this type of complementation by co-infection is not efficient. However, because of functional redundancy of effectors (Kvitko, Park et al. 2009), it may be possible to reduce the number of T1* strains in the mix. In any case, combining multiple effector genes into one strain would be ideal. These results also partially address the concern of reduced growth by every individual effector tested.

In regard to the discussion of the importance genetic background, the effector gene *hopM1* is particularly fascinating because of its ambiguous virulence and avirulence activities in several plant species. If HopM1-induced cell death is indeed the results of ETI as our data suggest, then the deletion that is present in *hopM1*$_{\text{Pto JL1065}}$ is possibly abolishing recognition. It would be interesting to test whether *hopM1*$_{\text{Pto JL1065}}$ still has a virulence function and whether it can complement the DC3000 ΔCEL mutant when expressed in *A. thaliana*. We are already in the process of making transgenic *A. thaliana* expressing *hopM1*$_{\text{Pto JL1065}}$. Preliminary results showed that transient expression of *hopM1*$_{\text{Pto JL1065}}$ did not cause any visible symptoms, in contrast with full-length *hopM1* (Nomura, Debroy et al. 2006).

Because full-length *hopM1* increases the number of bacterial specks on tomato leaves in our lab condition, it is possible that selection pressure in nurseries played a role in *hopM1* truncations found in most Pto strains. Pto strains carrying a full-length *hopM1* may cause more visible symptoms because of the increased number of specks in nurseries and thus have a smaller chance of getting distributed with the infected seedlings to tomato fields than Pto strains with disrupted *hopM1* genes that cause a smaller number of specks. Unfortunately, this hypothesis is not as well grounded since *P. syringae* life style and transmission of *P. syringae* are still not well known.

The increased number of specks caused by HopM1 when delivered from T1 during tomato infection was also accompanied by reduced speck size and slightly reduced growth. We hypothesized that this may be the result of HopM1-elicited ETI and localized HR. However, this may also be the result of the dominant negative effect of truncated HopM1 (N-terminal 200-300 amino acids) abolishing the virulence activity of full-length HopM1 observed by Nomura et al (Nomura, Debroy et al. 2006) when truncated HopM1 was expressed *in planta*. It is possible that truncated HopM1 expressed naturally from the T1 chromosome interferes with the virulence activity of the ectopically expressed full-length HopM1 by competing for the binding site of the HopM1 host target. The increased number of unbound full-length HopM1 protein in the plant cells may thus lead to an increased number of full-length HopM1 proteins that interact with a cognate resistance protein. Accordingly, to confirm that HopM1 elicits ETI in tomato, we will need to delete the truncated *hopM1* alleles from Pto strains. Unfortunately, because T1 is very recalcitrant to transformation, the attempts to delete the truncated *hopM1* gene from the T1 chromosome have not been successful by the time of writing of this manuscript.

We found that *hopM1* expressed in T1' also slightly reduced bacterial growth in Arabidopsis. In the literature, *hopM1* also slightly reduced bacterial growth when expressed in Pph NSP3121 during infection of *A. thaliana* (Ham, Kim et al. 2007). The reduction was small and statistically insignificant and thus was not discussed. We repeated the experiment with Pph expressing *hopM1* and repeatedly saw a reduction of bacterial growth when spray-inoculated (data not shown). We also confirmed that Pph NSP3121 expressing *hopM1* can indeed grow to high population in a *A. thaliana* pmr4/pad4 double mutant similar to DC3000, albeit only when infiltrated (data not shown). This suggests that HopM1 triggers defenses when delivered into *A. thaliana* by Pph NSP3121 similarly to when

delivered by T1, but that these defenses are abolished in the pmr4/pad4 mutant so that HopM1 can perform its virulence activity. This would be compatible with our hypothesis that *hopM1* needs another effector/effectors to perform its virulence activity in *A. thaliana.* However, also these experiments need to be repeated with strains that do not express any truncated version of HopM1 so that it becomes possible to determine if truncated MopM1 proteins interfere with the virulence activity of full-length HopM1 protein when delivered together into host plants. Our preliminary results from competition assay by transiently co-expressing both full-length and truncated *hopM1* in *N. benthamiana* suggest such inhibition may be very rare, if it happens at all, since we found that large amounts of transiently expressed truncated HopM1 cannot suppress the cell death triggered by full-length HopM1.

If the HopM1-triggered cell death in *A. thaliana*, tomato, and *N. benthamiana* is due to ETI, it can be expected to be dependent on genes known to be necessary for ETI. *SGT1* and *RAR1* are two genes that have been found to be necessary for ETI (Azevedo, Sadanandom et al. 2002; Takahashi, Casais et al. 2003; Azevedo, Betsuyaku et al. 2006). However, *SGT1* has also been found to contribute to symptoms caused by Pto in tomato and *A. thaliana* (Wang, Uppalapati et al. 2010). This confirms that the main difference between cell death during the HR and symptom formation may be timing and that it may not be straightforward to determine if cell death caused by HopM1 is due to a plant defense response triggered by HopM1 or if HopM1 has a role in symptom formation favoring the pathogen. It was previously shown that $HopM1_{Psy\ B728a}$-elicited cell death is attenuated in tobacco transiently silenced in *SGT1* (Vinatzer, Teitzel et al. 2006). However, when we silenced *SGT1* in *N. benthamiana*, we did not observe reduction of cell death caused by $HopM1_{Pto\ DC3000}$ (data not shown). When we transiently expressed $HopM1_{Pto\ DC3000}$ in *SGT1* and *RAR1* mutants of *A.*

thaliana, or when we delivered HopM1$_{Pto\ DC3000}$ to these mutants by Pf0-1 EtHAn, we observed inconsistent enhancement of chlorosis and necrosis symptoms in *rar1* mutant (data not shown). Thus *RAR1* may interfere with HopM1-induced cell death but our data are not strong enough to make this conclusion yet. *SGT1*- and *RAR1*-dependence of HopM1-triggered cell death will need to be investigated further.

The most surprising effect of HopM1 was seen when it was delivered from *P. fluoresens*. *P. fluorescens* secreting HopM1 through an ectopically expressed *P. syringae* T3SS reached a population density in *A. thaliana* almost as high as that of a pathogen. This high population density may have been possible either due to suppression of host defenses by HopM1 or because HopM1 elicited cell death as part of ETI and *P. fluorescens* used the dead plant tissue as source of nutrition. Future work will need to use Real-Time PCR to test whether defense genes were up- or down-regulated by HopM1.

The possibly dual function of effectors in virulence and avirulence, makes effectors, such as HopM1, difficult to study. With the development of new technologies and computer science to generate, manipulate, and analyze large scale data, such as new approaches in transcriptomics and proteomics, the future lies in systems biology approaches to better dissect host–microbe interactions in order to understand the complex interactions between the large number of different effector combinations and the host defense and metabolism networks. When we added HopM1 to DC3000 Δ*hopM1* and T1, we had to ignore most of the background effects because it is currently technically impossible to track the effects and interference from the background. However, using transcriptomics, proteomics, and metabolomics we expect this to change, and it should become possible to study effectors in many different contexts.

Conclusions

In conclusion, we established a system to study how host range is determined using the closely related strains T1 and DC3000. Adding individual effector genes from DC3000 to T1 did not significantly increase bacterial growth. Therefore, genetic background of a strain may be important for newly introduced effectors to perform virulence functions, and multiple effectors may be required to synergistically allow host jumps. This would be good news for durability of non-host resistance but further complicates the genetic dissection of non-host resistance. What we can conclude at this point is that *A. thaliana* resistance goes beyond ETI, is multi-genic, and that T1 is unable to overcome *A. thaliana* resistance, most likely because it is missing several DC3000 effector genes that need to act synergistically. Moreover, some of our data suggest that PAMPs of T1 may trigger a stronger PTI response than PAMPs of DC3000.

Investigating differences in effector repertoires between T1 and DC3000, HopM1 turned out to be a particularly intriguing effector with partial virulence and avirulence effects depending on the particular host–pathogen interaction in which it is expressed. Additional studies of *A. thaliana* gene expression in response to HopM1 delivered from different pathogens and HopM1 expression *in planta* in different mutants will be necessary to unravel its mystery.

e. Materials and Methods

Plant growth conditions and bacterial infections

A. thaliana plants were grown in Sunshine Mix 1 (Sun Gro) in a Conviron (Controlled Environments Ltd) growth chamber under the condition of a 12-h day,

23°C during the day and 19°C during the night. Tomato plants were grown in 1/2 Pro-Mix (Premier Horticulture) and 1/2 Metro-Mix 360 (Sun Gro) under a 16-h day on growth racks at room temperature. All plants were infected when about 4-5 weeks old.

P. syringae strains used for infection were prepared as previously described (Yan, Liu et al. 2008), except the growing temperature was 28℃. We used a Preval sprayer for spray inoculation. Plants were covered to maintain humidity for 24 hours after spray inoculation. Bacterial growth was measured 3 days post inoculation unless otherwise noted. Infiltration inoculations using blunt-end syringe were carried out for confirmation of spray infection results, or instead of spray infection as needed, and for callose deposition assays. Plants were covered to maintain humidity for 24 hours post inoculation.

Callose deposition

Bacterial suspensions were prepared and then inoculated by infiltration as described above. After different periods of time, *A. thaliana* leaves were harvested, fixed, cleared and stained as described (Adam and Somerville 1996) with minor modifications. Briefly, leaves were incubated in lactic alcohol for 4 hours, washed in 100% ethanol overnight, then washed in deionized water. Leaves were then stained using 0.01% aniline blue solution in 150mM K_2HPO_4 buffer for 4 hours before observed under fluorescent microscope (Zeiss Imager M1 with AxioCam MRm). Five random but representative areas of each leaf were taken picture for calculating the amount of callose deposition.

Construction of plasmids and genetic manipulation of bacterial strains

P. syringae and *P. fluoresens* strains were transformed by either tri-parental mating, electroporation as previously described (Mohr, Liu et al. 2008). Agrobacterium strains were transformed by tri-parental mating. To disrupt *avrRpt2* in T1, pBAV224 (Table 1; this work) was transformed into T1. Colonies that have this plasmid integrated into the genome, which results in disruption of *avrRpt2*, was selected by kanamycin and verified by PCR.

To delete *hopM1* and chaperon *shcM* in DC3000 unmarked, pCPP3283 (Collmer, Badel et al. 2000) was digested by *Sac*I and *Apa*I to remove fragments used for *avrE1* deletion. A 1.2kb fragment upstream of *shcM* was amplified by PCR using primers 5'-AAAAGGATCCATTTCAGTTCCTTGCGTTGG-3' and 5'-AAAAGGGCCCACCTTTGGCACTGCTGTTCT-3' with *BamH*I and *Apa*I restriction site added, respectively (restriction site underlined). Another 1.7 kb fragment downstream of hopM1 was amplified by PCR using primers 5'-AAAGAGCTCGCGAACATCACCGTTAACCT-3' and 5'-AAAGGATCCGCTTAAGCAGCCGACAAAAG-3' with *Sac*I and *BamH*I restriction enzyme added, respectively (restriction site underlined). These two PCR fragments were ligated with digested pCPP3282 to create pHopM1del. pHopM1del was transformed into DC3000, T1 and T1*. Transformants were selected on KB gentamicin plates. Δ*hopM1* mutants were selected for the loss of the suicide vector and PCR verified for the loss of *hopM1*.

Because many strains we used are naturally resistant to ampicillin, we subcloned ORF43 into shuttle vector pME6010 (Heeb, Itoh et al. 2000) which carries tetracycline resistance gene to create a pME6010-ORF43. Briefly, pME6010 was digested by *EcoR*I and *Hind*III, and then ligated with ORF43

fragment, which was digested from pORF43 (Debroy, Thilmony et al. 2004) using the same restriction enzymes.

DC3000 effector genes in pDONR207 were obtained from collaborators (see Table 1). These genes were subcloned into pBAV226 using Invitrogen LR Clonase (Invitrogen) and then transformed into T1 and other strains for expression.

Genome Sequencing and Draft Genome Assembly

Genomic DNA of *P. syringae* strains were extracted using Gentra Puregene Yeast/Bacteria kit (Qiagen) following manufacturer's protocol and sent to sequencing centers at University of Toronto Centre for the Analysis of Genome Evolution and Function (CAGEF) and University of North Carolina. Illumina sequencing (Bentley 2006) with pair-end read was used to sequence the genomes. Raw data was assembled using Velvet v. 0.7.x (Zerbino and Birney 2008) to obtain draft genome sequences. Genomic DNA of strain Max4 was only using single read protocol as previously described (Almeida, Yan et al. 2009). For NCBI WGS submission, no scaffolding was performed.

Effector Repertoire Prediction

Pseudomolecules were created from the draft genome sequences by concatenating contigs in the oder from largest to smallest with TIGR linker sequence "nnnnnttaattaattaannnnn" delimiting contig boundaries. Effectors were identified in the pseudomolecules using a combination of automated annotation generated by RAST (http://rast.nmpdr.org/), alignment of pseudomolecules with the DC3000 sequence visualized using the Artemis Comparison Tool, HrpL binding sites predicted as previously described (Ferreira, Myers et al. 2006) and

PSI-BLAST of confirmed effector sequences against the pseudomolecule sequences. To build a refined effector tables (Table 3), effector sequences from either DC3000 and T1 was blasted (NCBI-BLAST) against draft genome sequences to find out the presence of absence, and the exactly location.

HopM1 cloning, transient expression

The open reading frames including the ribosome binding site but not the stop codon of *hopM1* alleles were amplified by PCR from genomic DNA of Pto strains DC3000, JL1065, T1, NCPPB1108, and PT21 with the specific primer pairs and with nested primers to add sequences for Gateway^{TM} 18 (Invitrogen) cloning using the protocol described previously (Vinatzer, Teitzel et al. 2006). The five PCR products were then cloned into the entry vector pDNOR207 (Invitrogen) using the Gateway^{TM} BP cloning kit (Invitrogen). Recombined plasmids were confirmed by sequencing and cloned into the destination vector pBAV150 (Vinatzer, Teitzel et al. 2006) using the Gateway^{TM} LR cloning kit (Invitrogen). *hopM1*-containing pBAV150 were mated from *Escherichia coli* into *Agrobacterium tumefaciens* C58C1 and used in transient assays of tomato leaves (at a concentration of OD_{600} of 0.04) and in *N. benthamiana* leaves (at a concentration of OD_{600} of 0.4) using the same protocol as described previously for *N. benthamiana* (Vinatzer, Teitzel et al. 2006). *hopM1* alleles were transiently expressed in *A. thaliana* as described by Wroblewski and colleagues (Wroblewski, Tomczak et al. 2005).

Western blot

Three leaf discs of $3.14cm^2$ were taken from each *N. benthamiana* plant infiltrated with Agrobacterium with *hopM1*::*GFP* alleles, and grounded up

thoroughly in liquid N_2. 150ul of protein extraction buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol , 0.004% bromophenol blue, 0.125 M tris-HCl) was added to the sample tubes, and mixed well by using pestle in ice. The samples were boiled at 100°C for 10 minutes, and briefly spinned down. 15μl of each sample and 10μl of pre-stained protein ladder (PageRuler, Fermentas) were loaded on the 10% Mini-Protean TGX gel (Bio-Rad), and then electrophoresed in Tris-glycine buffer at 100V for about 1.5 hours. The protein samples were transferred on the 0.45μm nitrocellulose membrane (Bio-Rad) at 50V overnight at 4°C. The membrane was washed twice in PBST buffer (137mM NaCl, 2.7mM KCl, 10mM sodium phosphate dibasic, 2mM potassium phosphate monobasic, 0.1% Tween-20, pH 7.4), and blot in 20ml of blot buffer (5% Non-Fat-Dry-Milk in PBST) by shaking at room temperature about 1 hour. The membrane was hybridized with the first antibody of rabbit anti GFP (Invitrogen) in 1:5000 dilution in the blot buffer by shaking at room temperature for about 1.5 hours, and hybridized with the second antibody of goat anti rabbit HRP (Invitrogen) in 1:5000 dilution in PBST buffer by shaking at room temperature for about 1 hour. After each hybridization, membrane was washed in PBST buffer for 4 times. The membrane was developed in the mixture of 1:1 reagent of I & II (ECL, Pierce) at room temperature for 5 min. The membrane was exposed under Molecular Imager Gel Doc XR System (Bio-Rad) for 5 min.

Confocal Microscopy and co-localization analysis

Confocal microscopy was done using Zeiss Observer Z1 microscope (Zeiss). Objective used was a 40X Water immersion C-Apochromat with a 1.2 NA. GFP was excited with Argon laser 488nm line with emission filter of band pass 505-550nm. mCherry and chloroplast was excited with HE Neon 543nm laser with long

pass emission filter of 560nm. Slices for all channels were all 1μm thick except for some z-stack captures. 3D reconstruction of z-stack and graphical co-localization analysis carried about in ZEN 2009 software (Zeiss).

PCR and Real-time PCR

Regular PCR was performed as previously described (Yan, Liu et al. 2008), except that Bioline ImmoMix™ or ImmoMix™ Red Taq master mix (Bioline) were used. Arabidopsis was grown as described above. Bacterial suspension in $MgSO_4$ of OD_{600} equal to 0.01 was infiltrated into *A. thaliana* leaves using blunt end syringe. 8 hours after infiltration, leaves were harvested and flash frozen in liquid nitrogen. Total RNA was extracted using Qiagen RNAeasy plant kit (Qiagen) following manufacturer protocol except that the tissue homogenization of frozen leaves is done using paint shaker. RNA quality was checked by Nanodrop® spectrophotometer ND-1000 (Thermo Scientific) and agarose gel electrophoresis. Genomic DNA contaminations were removed using Turbo DNA-*free*™ kit (Ambion), and checked by end-point PCR. cDNA was synthesized using Promega GoScript™ reverse transcription system kit (Promega) following manufacturer protocol. Genomic DNA contamination of cDNA is also checked by end-point PCR. Real-time PCR was performed using Promega GoTaq® qPCR Master kit (Promega) in Applied Biosystems 7500 Real-Time PCR System. Disassociation of end-point PCR products was checked to ensure single peak. Specificity of real-time PCR primers was also check by end-point PCR. Real-time PCR primers were designed to span exon whenever possible using Integrated DNA Technology's web tool (http://www.idtdna.com/Scitools/Applications/RealTimePCR/). See Table 4 for list of quantitative Real-Time PCR primers used in this study.

Table 4. Real-Time PCR primers.

Real-Time PCR Primers for *A. thaliana* genes	Forward sequence	Reverse sequence	Note
WRKY30	GAAGTCAATGCCAAGGTG	CGTGCATCTATAGTATCCTCT	
WRKY22	GCTTTAAACTCCGATGTCTG	GATCTATTTCGCTCCACTTG	
FRK1	GACTATTTGGCAGGTAAAAGG	TGTTGGCTTCACATCTCTG	
PR1	TGGTCACTACACTCAAGTTG	GGCTTCTCGTTCACATAATTC	
PAD4	CCGATGAACCTCTACCTATG	AATTCCAATTCCAATCCTTCC	
ACD6	CGTGAAGTGAGGAAGTTTAAC	CCTACCATCTTCATCTTGCTC	
ACTIN2	CCTTTAACTCTCCCGCTATG	AGAAACCCTCGTAGATTGG	Endogenous control
At1g13320	CTTTGTAGAATTGGTGCTCAG	CACTTGGGTATGCAATATGG	

f. Author attributions

B. A. Vinatzer and S. Yan conceived and designed the experiments. S. Yan carried out the experiments. H. Liu cloned the *hopM1* alleles, and performed the western blots. M. Lindeberg performed the initial effector repertoire prediction. R. Cai expanded the strain collection's MLST data on top of (Yan, Liu et al. 2008) and built phylogenetic trees. S. Yan and B.A. Vinatzer wrote the manuscript.

g. References

Adam, L. and S. C. Somerville (1996). "Genetic characterization of five powdery mildew disease resistance loci in *Arabidopsis thaliana*." Plant J **9**(3): 341-356.

Alfano, J. R., A. O. Charkowski, et al. (2000). "The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants." Proc Natl Acad Sci U S A **97**(9): 4856-4861.

Almeida, N. F., S. Yan, et al. (2009). "A draft genome sequence of *Pseudomonas syringae* pv. *tomato* T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. *tomato* DC3000." Mol Plant Microbe Interact **22**(1): 52-62.

Asai, T., G. Tena, et al. (2002). "MAP kinase signalling cascade in Arabidopsis innate immunity." Nature **415**(6875): 977-983.

Azevedo, C., S. Betsuyaku, et al. (2006). "Role of SGT1 in resistance protein accumulation in plant immunity." EMBO J **25**(9): 2007-2016.

Azevedo, C., A. Sadanandom, et al. (2002). "The *RAR1* interactor *SGT1*, an essential component of R gene-triggered disease resistance." Science **295**(5562): 2073-2076.

Bentley, D. R. (2006). "Whole-genome re-sequencing." Curr Opin Genet Dev **16**(6): 545-552.

Buell, C. R., V. Joardar, et al. (2003). "The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000." Proc Natl Acad Sci U S A **100**(18): 10181-10186.

Collmer, A., J. L. Badel, et al. (2000). "*Pseudomonas syringae* Hrp type III secretion system and effector proteins." Proc Natl Acad Sci U S A **97**(16): 8770-8777.

Cunnac, S., M. Lindeberg, et al. (2009). "*Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions." Curr Opin Microbiol **12**(1): 53-60.

Debroy, S., R. Thilmony, et al. (2004). "A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants." Proc Natl Acad Sci USA **101**(26): 9927-9932.

Feil, H., W. S. Feil, et al. (2005). "Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. tomato DC3000." Proc Natl Acad Sci U S A **102**(31): 11064-11069.

Ferreira, A. O., C. R. Myers, et al. (2006). "Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae* pv. *tomato* DC3000, allows de novo reconstruction of the Hrp cis clement, and identifies novel coregulated genes." Mol Plant Microbe Interact **19**(11): 1167-1179.

Gardan, L., H. Shafik, et al. (1999). "DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. *nov*. and *Pseudomonas cannabina* sp. *nov*. (ex Sutic and Dowson 1959)." Int J Syst Bacteriol **49 Pt 2**: 469-478.

Göhre, V. and S. Robatzek (2008). "Breaking the barriers: microbial effector molecules subvert plant immunity." Annu Rev Phytopathol **46**: 189-215.

Gomez-Gomez, L. and T. Boller (2000). "FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis." Mol Cell **5**(6): 1003-1011.

Ham, J. H., M. G. Kim, et al. (2007). "Layered basal defenses underlie non-host resistance of Arabidopsis to *Pseudomonas syringae* pv. *phaseolicola*." Plant J **51**(4): 604-616.

Ham, J. H., D. R. Majerczak, et al. (2009). "Multiple activities of the plant pathogen type III effector proteins WtsE and AvrE require WxxxE motifs." Mol Plant Microbe Interact **22**(6): 703-712.

Heeb, S., Y. Itoh, et al. (2000). "Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria." Mol Plant Microbe Interact **13**(2): 232-237.

Joardar, V., M. Lindeberg, et al. (2005). "Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among

pathovars in genes involved in virulence and transposition." J Bacteriol **187**(18): 6488-6498.

Jones, J. D. and J. L. Dangl (2006). "The plant immune system." Nature **444**(7117): 323-329.

Kang, L., J. Li, et al. (2003). "Interplay of the Arabidopsis nonhost resistance gene *NHO1* with bacterial virulence." Proc Natl Acad Sci USA **100**(6): 3519-3524.

Kunkeaw, S., S. Tan, et al. (2010). "Molecular and evolutionary analyses of Pseudomonas syringae pv. tomato race 1." Mol Plant Microbe Interact **23**(4): 415-424.

Kvitko, B. H., D. H. Park, et al. (2009). "Deletions in the repertoire of *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion effector genes reveal functional overlap among effectors." PLoS Pathog **5**(4): e1000388.

Lacombe, S., A. Rougon-Cardoso, et al. (2010). "Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance." Nature biotechnology.

Lin, N.-C., R. B. Abramovitch, et al. (2006). "Diverse AvrPtoB homologs from several *Pseudomonas syringae* pathovars elicit Pto-dependent resistance and have similar virulence activities." Appl Environ Microbiol **72**(1): 702-712.

Lindeberg, M., S. Cunnac, et al. (2009). "The evolution of *Pseudomonas syringae* host specificity and type III effector repertoires." Mol Plant Pathol **10**(6): 767-775.

Maiden, M. C. (2006). "Multilocus Sequence Typing of Bacteria." Annu Rev Microbiol.

Martin, G. B., S. H. Brommonschenkel, et al. (1993). "Map-based cloning of a protein kinase gene conferring disease resistance in tomato." Science **262**(5138): 1432-1436.

Mohr, T. J., H. Liu, et al. (2008). "Naturally occurring nonpathogenic isolates of the plant pathogen Pseudomonas syringae lack a type III secretion system and effector gene orthologues." Journal of bacteriology **190**(8): 2858-2870.

Nelson, B. K., X. Cai, et al. (2007). "A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants." The Plant journal : for cell and molecular biology **51**(6): 1126-1136.

Nomura, K., S. Debroy, et al. (2006). "A bacterial virulence protein suppresses host innate immunity to cause plant disease." Science **313**(5784): 220-223.

Nomura, K., M. Melotto, et al. (2005). "Suppression of host defense in compatible plant-*Pseudomonas syringae* interactions." Curr Opin Plant Biol **8**(4): 361-368.

Sarkar, S. F., J. S. Gordon, et al. (2006). "Comparative genomics of host-specific virulence in *Pseudomonas syringae*." Genetics **174**(2): 1041-1056.

Sun, W., F. M. Dunning, et al. (2006). "Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses." Plant Cell **18**(3): 764-779.

Takahashi, A., C. Casais, et al. (2003). "HSP90 interacts with *RAR1* and *SGT1* and is essential for RPS2-mediated disease resistance in Arabidopsis." Proc Natl Acad Sci USA **100**(20): 11777-11782.

Thomas, W. J., C. A. Thireault, et al. (2009). "Recombineering and stable integration of the *Pseudomonas syringae* pv. *syringae* 61 *hrp/hrc* cluster into the genome of the soil bacterium *Pseudomonas fluorescens* Pf0-1." The Plant journal : for cell and molecular biology **60**(5): 919-928.

Uppalapati, S. R., Y. Ishiga, et al. (2007). "The phytotoxin coronatine contributes to pathogen fitness and is required for suppression of salicylic acid accumulation in tomato inoculated with *Pseudomonas syringae* pv. *tomato* DC3000." Mol Plant Microbe Interact **20**(8): 955-965.

Vinatzer, B. A., G. M. Teitzel, et al. (2006). "The type III effector repertoire of *Pseudomonas syringae* pv. *syringae* B728a and its role in survival and disease on host and non-host plants." Mol Microbiol **62**(1): 26-44.

Vinatzer, B. A., G. M. Teitzel, et al. (2006). "The type III effector repertoire of *Pseudomonas syringae* pv. *syringae* B728a and its role in survival and disease on host and non-host plants." Mol Microbiol **62**(1): 26-44.

Wang, K., S. R. Uppalapati, et al. (2010). "*SGT1* positively regulates the process of plant cell death during both compatible and incompatible plant-pathogen interactions." Mol Plant Pathol **11**(5): 597-611.

Wei, C. F., B. H. Kvitko, et al. (2007). "A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*." Plant J **51**(1): 32-46.

Wei, Z. M., R. J. Laby, et al. (1992). "Harpin, elicitor of the hypersensitive response produced by the plant pathogen *Erwinia amylovora*." Science **257**(5066): 85-88.

Whalen, M. C., R. W. Innes, et al. (1991). "Identification of *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean." Plant Cell **3**(1): 49-59.

Wildermuth, M. C., J. Dewdney, et al. (2001). "Isochorismate synthase is required to synthesize salicylic acid for plant defence." Nature **414**(6863): 562-565.

Wroblewski, T., K. S. Caldwell, et al. (2009). "Comparative large-scale analysis of interactions between several crop species and the effector repertoires from

multiple pathovars of *Pseudomonas* and *Ralstonia*." Plant Physiol **150**(4): 1733-1749.

Wroblewski, T., A. Tomczak, et al. (2005). "Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis." Plant Biotechnol J **3**(2): 259-273.

Yan, S., H. Liu, et al. (2008). "Role of recombination in the evolution of the model plant pathogen Pseudomonas syringae pv. tomato DC3000, a very atypical tomato strain." Appl Environ Microbiol **74**(10): 3171-3181.

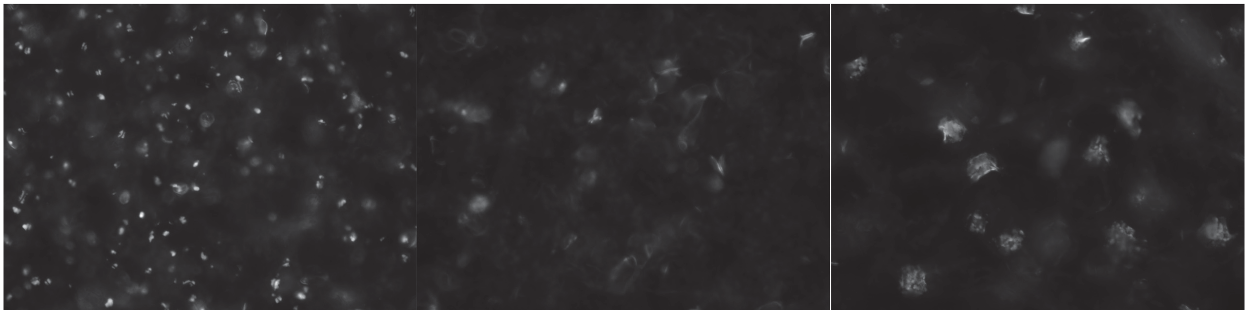
Zerbino, D. R. and E. Birney (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." Genome Res **18**(5): 821-829.

Supplementary Figures

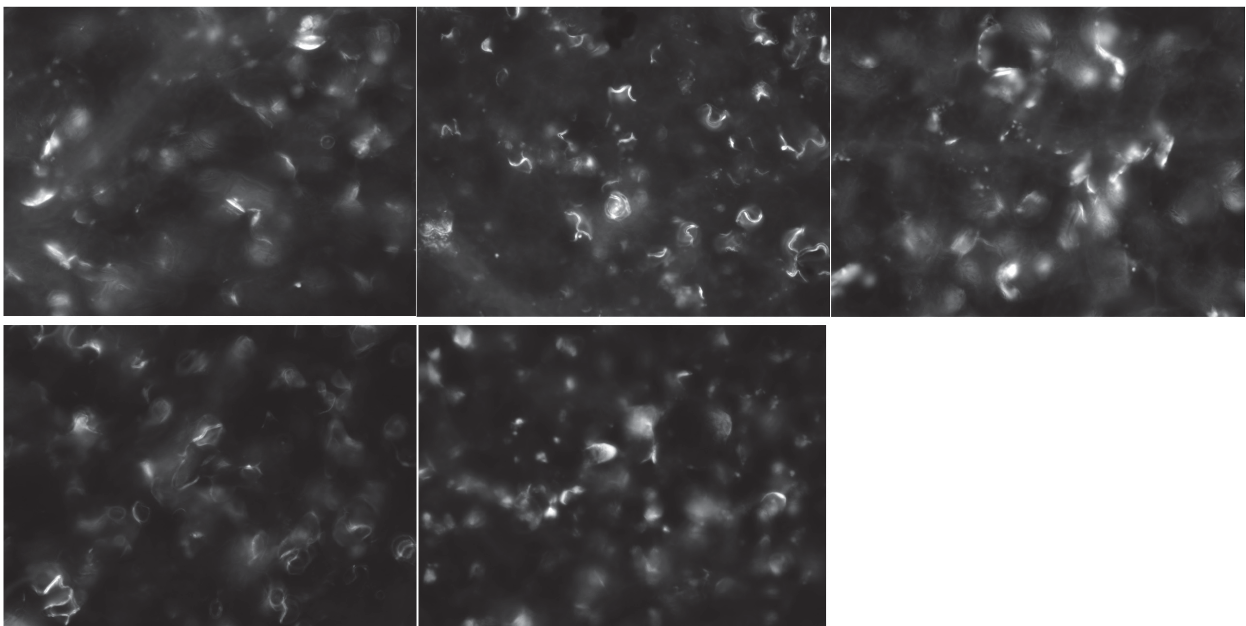
Supplementary Figure 1

Different callose deposition morphologies

<18H POI



>= 2D POI



Chapter 5

Conclusions

a. Summary and Conclusions

The evolution of the bacterial plant pathogen *Pseudomonas syringae* has been a subject of several studies (Sawada, Takeuchi et al. 1997; Sawada, Suzuki et al. 1999; Sarkar and Guttman 2004; Hwang, Morgan et al. 2005). However, many questions remained unanswered. This is in large part due to the fact that traditional typing methods were ineffective at differentiating closely related *P. syringae* strains. On the other hand, researchers studied the entire *P. syringae* species complex, which diverged from a common ancestor millions of years ago resulting in substantial diversity of strains that are very different from each other and that make a meaningful comparison problematic. Also, in most studies strains were not properly sampled to represent the true diversity of *P. syringae* in nature but selection of strains was heavily biased based on strain availability.

In regard to *P. syringae* host range evolution, earlier research focused on the identification of avirulence effectors that restrict host range due to HR (Carney and Denny 1990; Whalen, Innes et al. 1991; Ronald, Salmeron et al. 1992). Later research focused more on the functions of effectors and molecular mechanisms of their interactions with plants. Only after the publication of genome sequences of three *P. syringae* strains (Buell, Joardar et al. 2003; Joardar, Lindeberg et al. 2005; Vencato, Tian et al. 2006), researchers were able to start looking into the diversity of effector repertoires by using bioinformatics. However, the three sequenced strains are only distantly related to each other making it almost impossible to correlate the many genomic differences (effector gene differences in particular) to the differences in host range that exist between them. Early attempts of using effector information extracted from the sequenced strains to characterize other strains for which no genome sequences were available were compromised by the elevated differences in DNA sequence between strains that made the use of DNA

probes unreliable. Also, limiting probes to effector genes of a single strain, DC3000, did not come close to assaying the true effector diversity of strains (Sarkar, Gordon et al. 2006).

Aware of above limitations, we took a new approach to study evolution and mechanisms of host range in *P. syringae*. First of all, we studied strains that, based on the literature, are closely related and thus can be expected to have diverged recently from a common ancestor with a relatively small amount of genomic differences. The idea was that this choice of strains would make it possible to correlate the relatively small number of differences between strains with their host range differences and that this would make it possible to experimentally determine the role of the identified genomic differences in determining host range differences.

Therefore, we first determined the phylogenetic relationship between the chosen strains and determined their host range (Chapter 2). By using high-throughput genome sequencing technologies, we were then able to examine the differences between some of these closely related *P. syringae* strains. Specifically, by comparing Pto T1, which cannot cause disease in *A. thaliana*, to Pto DC3000, which has been widely used as model pathogen in *A. thaliana*, we could identify genes that presumably confer DC3000 compatibility with *A. thaliana* and lead to incompatibility between T1 and *A. thaliana*. Moreover, we found that resistance of *A. thaliana* to T1 is an example of multi-genic and robust non-host resistance. It thus represents an excellent system for studying the mechanisms at the basis of non-host resistance.

After sequencing 10 genomes, we narrowed down key Type III effector genes crucial for causing disease in *A. thaliana* to only nine candidates. Our tests found that these candidate effectors individually do not allow T1 to cause disease in *A. thaliana*. Since hypothetically, these 9 effectors collectively act in synergy

with other DC3000 effectors to suppress *A. thaliana* defenses, this finding is not too surprising.

Because the strains that cause disease in *A. thaliana* are located in distinct locations in the phylogenetic tree (Chapter 4 Table 3A and Chapter 2 Fig 2A), and because of the recombination we detected between these strains (Chapter 2), effectors that have variable distribution between strains ("exchangeable effectors") were likely acquired and/or lost by these strains since their divergence from their most recent common ancestor. We speculate that these exchangeable effectors are enhancements to the basic effector arsenal (or "core effectors") to enhance the overall potency of the core effectors to suppress the defenses of particular hosts. They thus may represent recent adaptations to particular hosts but may trigger defenses in other plant species. In other words, the exchangeable effectors may allow specific host–pathogen compatibility and lead to incompatibility with non-hosts. The functions of these exchangeable effectors may be more delicate, targeted and additive. The fact that the effector repertoire of these strains are not totally randomly distributed and the fact that strains with the same host do share a set of conserved exchangeable effectors, supports the notion that these exchangeable effectors act collectively. Individually, they may thus not have significant contribution to compatibility.

The reduction of T1 growth by the exchangeable effectors conserved between *A. thaliana* pathogens, however, is unexpected and intriguing. Our current hypothesis is that adding additional exchangeable effectors individually to a non-adapted pathogen could potentially disrupt the cooperation between existing core and exchangeable effectors of this pathogen, resulting in an unfavorable outcome. Alternatively, plants may recognize more effectors than we currently think and thereby activate ETI. Consequently, when adding effectors individually without the corresponding "gang" to keep them "under cover", ETI may triumph over

virulence, resulting in a slight decrease of growth. However, we would like to use caution that although our results so far lead to this hypothesis, more experiments need to be done.

When examining in detail the candidate effectors that contribute to host range, we observed that HopM1 reduces T1 growth in both *A. thaliana* and tomato. HopM1 must be considered an ancient effector since it is present in the conserved effector locus in most strains of *P. syringae*. Plant species can thus be expected to have evolved mechanisms to recognize HopM1 during co-evolution. Because HopM1 elicits an HR-like cell death response in *A. thaliana* (Nomura, Debroy et al. 2006), tomato and *N. benthamiana* (Chapter 4), and typical highly virulent tomato strains have lost *hopM1* due to mutations, we hypothesized that HopM1 might have both virulence and avirulence roles in plant cells. Our additional observation that *P. fluorescens* expressing *hopM1* and a T3SS caused disease-like symptoms and grew to over 10^6 CFU/cm^2 in the leaves of both *A. thalina* and *N. benthamiana*, further supports this hypothesis. This is a beautiful example of the importance of the genetic background of the bacterium from which an effector is expressed in determining the ultimate effect that this effector has on the plant cells and, consequently, on the bacterium form which it is secreted. The dramatic effect of *hopM1* on growth of *P. fluorescens* and on symptom formation also raises the question why *P. fluorescens* has not acquired a T3SS and *hopM1* during evolution to become a pathogen. Further investigating this question may bear insights into pathogen evolution and emergence of new pathogens.

In conclusion, our data suggest that host switching to a new host most likely requires the loss and/or acquisition of multiple effector genes (and possibly other genes), even when the non-adapted pathogen is a very close relative of pathogens of the new host. These loss and acquisition events must happen in a relatively short period of time since individually effectors may not increase growth on the new

host and, therefore, no or little selection will act on strains that acquire or lose individual effectors. Also, our data in regard to HopM1 suggest that having both virulence and avirulence roles on the same host may be a common theme for many effectors. This could explain the need for exchangeable effectors that function to cover up ETI-activity and allow them to carry out their virulence effects in plants. While the notion that effectors can suppress ETI of other effectors is nothing new, see for example Jackson et al (Jackson, Athanassopoulos et al. 1999), it has not yet been shown that effectors, which ETI activity is suppressed by other effectors, actually exert a virulence activity. Our data suggest this but further experiments are required to identify those DC3000 effectors that possibly suppress the ETI activity of the DC3000 effectors that reduce growth of T1 when ectopically expressed in this strain.

This study highlights the complexity of host range determination even when very closely related strains are compared. However, our results have already identified a small number of effector differences between T1 and DC3000 that are feasible to express in different combinations in T1 in the future making the T1–DC3000 comparison a promising pathogen pair to unravel the basis of host range determination. To our knowledge, this study is also the most in-depth, comprehensive, and systematic study of the evolution and mechanisms of host range of bacterial plant pathogens using the latest genome sequencing and analysis tools.

Finally, I would like to point out that it will be crucial to use a systems biology approach to clarify the functions of effectors, individually or collectively, in plant cells by utilizing genomics, transcriptomics, proteomics and metabolomics. The gained insight into effector function will be necessary for intelligent design of crops with more durable and robust resistance against pathogens. Systems biology approaches should ultimately answer the question of

the extent to which effectors have multiple functions by examining the changes they induce at all intersections of the plant–pathogen interaction network.

b. References

Buell, C. R., V. Joardar, et al. (2003). "The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000." Proc Natl Acad Sci U S A **100**(18): 10181-10186.

Carney, B. F. and T. P. Denny (1990). "A cloned avirulence gene from *Pseudomonas solanacearum* determines incompatibility on *Nicotiana tabacum* at the host species level." Journal of bacteriology **172**(9): 4836-4843.

Hwang, M. S. H., R. L. Morgan, et al. (2005). "Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*." Appl Environ Microbiol **71**(9): 5182-5191.

Jackson, R. W., E. Athanassopoulos, et al. (1999). "Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar *phaseolicola*." Proc Natl Acad Sci USA **96**(19): 10875-10880.

Joardar, V., M. Lindeberg, et al. (2005). "Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among

pathovars in genes involved in virulence and transposition." J Bacteriol **187**(18): 6488-6498.

Nomura, K., S. Debroy, et al. (2006). "A bacterial virulence protein suppresses host innate immunity to cause plant disease." Science **313**(5784): 220-223.

Ronald, P. C., J. M. Salmeron, et al. (1992). "The cloned avirulence gene *avrPto* induces disease resistance in tomato cultivars containing the Pto resistance gene." J Bacteriol **174**(5): 1604-1611.

Sarkar, S. F., J. S. Gordon, et al. (2006). "Comparative genomics of host-specific virulence in *Pseudomonas syringae*." Genetics **174**(2): 1041-1056.

Sarkar, S. F. and D. S. Guttman (2004). "Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen." Appl Environ Microbiol **70**(4): 1999-2012.

Sawada, H., F. Suzuki, et al. (1999). "Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of *argK* and the evolutionary stability of hrp gene cluster." J Mol Evol **49**(5): 627-644.

Sawada, H., T. Takeuchi, et al. (1997). "Comparative analysis of *Pseudomonas syringae* pv. *actinidiae* and pv. *phaseolicola* based on phaseolotoxin-resistant ornithine carbamoyltransferase gene (*argK*) and 16S-23S rRNA intergenic spacer sequences." Appl Environ Microbiol **63**(1): 282-288.

Vencato, M., F. Tian, et al. (2006). "Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A." Mol Plant Microbe Interact **19**(11): 1193-1206.

Whalen, M. C., R. W. Innes, et al. (1991). "Identification of *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean." Plant Cell **3**(1): 49-59.

Appendix

Contribution to other manuscripts

Contributed to the construction of phylogenetic tree using Bayesian interference for manuscript: Morris et al. The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. ISME J (2008) vol. 2 (3) pp. 321-34.

Contributed to the construction of phylogenetic tree, preparation of several figures, and writing for manuscript: Mohr et al. Naturally occurring nonpathogenic isolates of the plant pathogen *Pseudomonas syringae* lack a type III secretion system and effector gene orthologues. Journal of bacteriology (2008) vol. 190 (8) pp. 2858-70.

Contributed to the construction of PAMDB.org website database design, graphical user interface and design, and writing for manuscript: Almeida et al. PAMDB, A Multilocus Sequence Typing and Analysis Database and Website for Plant-Associated Microbes. Phytopathology (2010) vol. 100 (3) pp. 208-15.

Contributed to the writing for manuscript: Vinatzer and Yan. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol Plant Pathol (2008) vol. 9 (1) pp. 105-18.