Title

Characterizing the evolution and mechanisms of bacterial epitope perception and evasion of the plant immune systems

Permalink

https://escholarship.org/uc/item/2h44q8jz

Author

Stevens, Danielle

Publication Date

2024

Peer reviewed|Thesis/dissertation

Characterizing the Evolution and Mechanisms of Bacterial Epitope

Perception and Evasion of the Plant Immune Systems

By

DANIELLE M. STEVENS

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics & Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Gitta L. Coaker, Chair

Daniel J. Kliebenstein

Mary Beth Mudgett

Committee in Charge

2024

ABSTRACT

Both plants and animals are impacted by diverse biotic threats. To limit disease, plants use protein receptors to recognize and respond to pathogen protein epitopes or effectors. Pathogens have evolved strategies to circumvent recognition to proliferate and cause disease. Pathogens can also persist on non-hosts, leading to reservoir populations and subsequent costly outbreaks. Despite considerable resources focused on understanding the interactions between pathogens and model organisms, we lack considerable knowledge in how the natural diversity of bacterial pathogens, particularly Gram-positive actinobacteria, impact plant immune perception, colonization, and disease susceptibility. Using a combination of comparative genomics, genetics, and biochemistry, I leveraged natural genetic variation to understand the evolution of pathogen epitopes and elucidate a driver of pathogen evasion in a Gram-positive actinobacteria.

Pathogen recognition and receptor signaling is crucial in host-pathogen interactions, but most studies use a single pathogen epitope and thus, the impact of multi-copy epitopes on pathogen outcomes is unknown. Through comparative genomics of thousands of plant-associated bacterial genomes, I characterized the naturally-evolved bacterial epitope landscape and their impact on pathogen outcomes. I revealed that natural variation was constrained yet experimentally testable and both epitope sequence and copy number variation altered pathogen-immune outcomes. Through genetic and biochemical analyses, I uncovered a mechanism for pathogen immune evasion, intrabacterial antagonism, where a non-immunogenic epitope blocks perception of immunogenic forms encoded in a single genome. One such intrabacterial antagonist, cold shock protein CspB, was conserved in actinobacteria including *Clavibacter,* a genus comprised of several crop pathogens including tomato, potato, wheat, and corn. As a non-model system, I developed a genetic toolkit to manipulate *Clavibacter* and test the role of CspB in blocking immune perception

of one host species, tomato. While I was able to build and validate the genetic tools through deletion of several critical virulence genes, I was unable to generate a null mutant of the *cspB* gene in *C. michiganensis,* likely due to its high GC-content between 73-78%. Instead, I validated our intrabacterial antagonism model though a combination of biochemical assays and genetic transfer of *cspB* to another foliar pathogen of tomato, *Pseudomonas syringae* pathovar *tomato* DC3000. I show via bacterial titers that expression of antagonist *cspB* blocked perception of other native encoded immunogenic cold shock proteins in a receptor-dependent manner.

Collectively, I revealed a mechanism for immune evasion and showcased the importance of analyzing all epitope copies within a genome. I also provided evidence that Gram-positive actinobacteria interface with the plant immune system, a paradigm previously put into question due to insufficient evidence. Finally, I developed a genetic toolkit which may aid in characterizing other genotypic-phenotypic outcomes in the non-model bacterium. While my research has shown that we can leverage natural genetic variation to generate hypotheses and understand their impact on phenotypic outcomes, major questions remain in the evolution, functional biology, and signaling in plant-microbe interactions, which is addressed in the final chapter. Findings from the research questions posed may provide critical insights for subsequent advancements in bioengineering for disease resistance.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

After six years, nothing I have done would have been possible without those who have aided me along the way. I have received countless support through all the highs and lows. My journey began during the most expected times and with each twist and turn, while there was disappointment, there was also joy. While I could write a book to thank those involved, instead, I'll opt for a brief summary.

Thank to everyone in the Coaker lab for all the memories and moments. I could not image a better place to learn and push the boundaries of what is known. Thanks to Signe Lolle for being my bench buddy and continued friend. Thanks to Lei Lei, DongHyuk Lee, and Yi-Chang Sung for their patience and teaching me the foundation of the practical protein knowledge that continues to give back. Thanks to Domonique Lewis for an A+ friend when I needed it most and donating extra small gloves that made all future wet lab work not a complete annoyance (I am seriously more thankful for that than you could guess). Thanks to Bardo Castro for always making me laugh and willingness to answer any question, even if it costed me a beer every time (my debt is paid off by now). Thanks to Jerry Li and Charis Ramsing for always supporting me, the crazy science ideas I come up with, and always willing to think and talk. And final thanks to Alba Moreno Perez for being my right-hand woman in crime and always being 'the best'. This last year was by far the most memorable, my favorite of all, and I feel grateful to end this chapter surrounded by some of the kindest, smartest, and funniest scientists one could ever ask for.

Thank you to all the fellow students and postdocs of those I had the honor of spending time with before graduate school began. Members of the Chang lab will always be my adopted family from afar, and I will always cherish all the memories of my time in Germany. From the Chang Lab, Elizabeth Savory, Ed Davis, Skylar Fuller, Chih-Fung Wu, and Alex Weisberg and

Stukenbrock lab, Alice Feurtey: thank you for teaching me the foundational knowledge that made some much of what I accomplished in graduate school, possible.

Thank you to all those in the IGG and PLP community who have continued to support me. I have been so lucky to be surrounded by so many wonderful, thoughtful, kind, and generous individuals. You are what has made these last six years exceptional.

Thanks to my parents: you provided me a bedrock foundation in life that has continued to pay back. Your unwavering support is what helped me dream big and never look back. I would not be the strong woman I am today without you. To Brian, thanks for all the support during the late nights and long weekends, and of course, all our adventures. I love that I got to have these last six years next to someone as curious and passionate as myself. To all our friends for all the laugher, joy, and memories. Cheers to the future and all that we discover in life.

INTRODUCTION

The Evolution, Ecology, and Mechanisms of Infection by Gram-Positive, Plant-Associated Bacteria

Shree P. Thapa, Edward W. Davis II, Qingyang Lyu, Alexandra J. Weisberg, Danielle M. Stevens, Christopher R. Clarke, Gitta Coaker, and Jeff H. Chang

Abstract

Gram-positive bacteria are prominent members of plant-associated microbial communities. Although many are hypothesized to be beneficial, some are causative agents of economically important diseases of crop plants. Because the features of Gram-positive bacteria are fundamentally different relative to those of Gram-negative bacteria, the evolution and ecology as well as the mechanisms used to colonize and infect plants also differ. Here, we discuss recent advances in our understanding of Gram-positive, plant-associated bacteria and provide a framework for future research directions on these important plant symbionts.

Author Contributions

I developed Figure 1, which was remade by *Annual Reviews* computer graphics artists, and wrote its legend. Additionally, I wrote parts of the *Rhodococcus*-focused sections and provided edits throughout other parts of the manuscript.

The Evolution, Ecology, and Mechanisms of Infection by Gram-Positive, Plant-Associated Bacteria

Shree P. Thapa,[1] Edward W. Davis II,[2,3,4] Qingyang Lyu,[1] Alexandra J. Weisberg,[2] Danielle M. Stevens,[2,5] Christopher R. Clarke,[6] Gitta Coaker,[1] and Jeff H. Chang[2,3,4]

[1]Department of Plant Pathology, University of California, Davis, California 95616, USA

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331, USA; email: changj@science.oregonstate.edu

[3]Molecular and Cellular Biology Program, Oregon State University, Corvallis, Oregon 97331, USA

[4]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon 97331, USA

[5]Integrative Genetics and Genomics, University of California, Davis, California 95616, USA

[6]Genetic Improvement for Fruits and Vegetables Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, Maryland 20705, USA

Keywords

actinobacteria, phytopathogen, symbiosis, horizontal gene transfer, evolution, genomics

Abstract

Gram-positive bacteria are prominent members of plant-associated microbial communities. Although many are hypothesized to be beneficial, some are causative agents of economically important diseases of crop plants. Because the features of Gram-positive bacteria are fundamentally different relative to those of Gram-negative bacteria, the evolution and ecology as well as the mechanisms used to colonize and infect plants also differ. Here, we discuss recent advances in our understanding of Gram-positive, plant-associated bacteria and provide a framework for future research directions on these important plant symbionts.

INTRODUCTION

Many taxa of bacteria have members that are adapted to plants and can gain access to resources in the form of nutrients and habitable space. There are several types of symbiosis, and the activities of symbiotic bacteria can have profound effects on plants, with mutualist and associative symbionts reciprocating benefits to their hosts and pathogens exploiting the host for selfish gains. When fitness costs and benefits of symbioses are quantified, it becomes clear that there is inexactness in categorization and a continuum even between two symbiotic partner species. There are interactions between the location of a symbiont on a host, the health of the host, environmental conditions, and horizontal gene transfer (HGT), to name a few, that tune outcomes. Defining symbioses in the context of agriculture is even more complex. Costs and categorization of plant–microbe interactions are often defined on the basis of the expression of conspicuous symptoms and economic impact (see the two definitions of disease and their relationships to the definitions for pathogen and virulence factor). Economic impact is not a selective pressure that directly shapes the evolution of symbiotic bacteria. Hence, it is critical to consider the natural variation and ecology of populations, which provide the context for natural selection, or lack thereof, to understand why and how organisms adapt.

Gram-positive bacteria generally have a single cytoplasmic membrane and a thick peptidoglycan layer that retains crystal violet stain. Across the Bacteria domain, there is variation in the structure of the cell envelopes, and responses to Gram staining do not always follow phylogeny. Most members of Actinobacteria react positively to Gram staining, but the phylum also includes mycolic acid–containing genera of bacteria that do not. Mycolic acids form an outer mycomembrane that is impermeable to the Gram stain (67). Our use of the term Gram-positive is thus more colloquial and admittedly short on precision.

We highlight the molecular mechanisms hypothesized to be necessary for infection and examine the evolutionary ecology of Gram-positive bacteria best known for being pathogenic and causing significant economic losses. Our focus is on six groups, each currently classified as different genera within the Actinobacteria phylum (**Figure 1**). We also construct a phylogenomic framework as a resource for comparative analyses and for generating testable hypotheses.

LIFESTYLES OF GRAM-POSITIVE, PLANT-ASSOCIATED BACTERIA
Clavibacter: A Group of Xylem-Limited Bacteria

The *Clavibacter* genus has experienced multiple changes to its membership (36, 47, 89). Some members were reclassified into *Rathayibacter* and *Leifsonia* genera, leaving only a single *Clavibacter michiganensis* group with multiple subspecies. Recent recommendations suggested elevating the subspecies to species rank (89).

Each species of plant-pathogenic *Clavibacter* primarily infects one main host species. When artificially infected, pathogens can cause some symptoms on a wide range of plant species (46). Infection occurs through wounds and hydathodes or via contaminated seeds. *Clavibacter* proliferates in xylem vessels, resulting in systemic infection (**Figure 1**) (45). Symptoms include wilting, canker, necrosis, and leaf spots. Owing to its effectiveness as a seedborne pathogen, *Clavibacter* has a high risk of adverse consequences, and some species have been classified as quarantine organisms in Europe and other countries (45).

Although the genus is most known for its pathogenic members, multiple *Clavibacter* strains have been isolated as asymptomatic endophytes of monocot and dicot plant species (31, 125, 131, 150). Endophytic *Clavibacter* cells colonize the internal tissues of the plant and confer no conspicuous symptoms or negative effects on plant fitness (75, 125). Furthermore, after artificial inoculation,
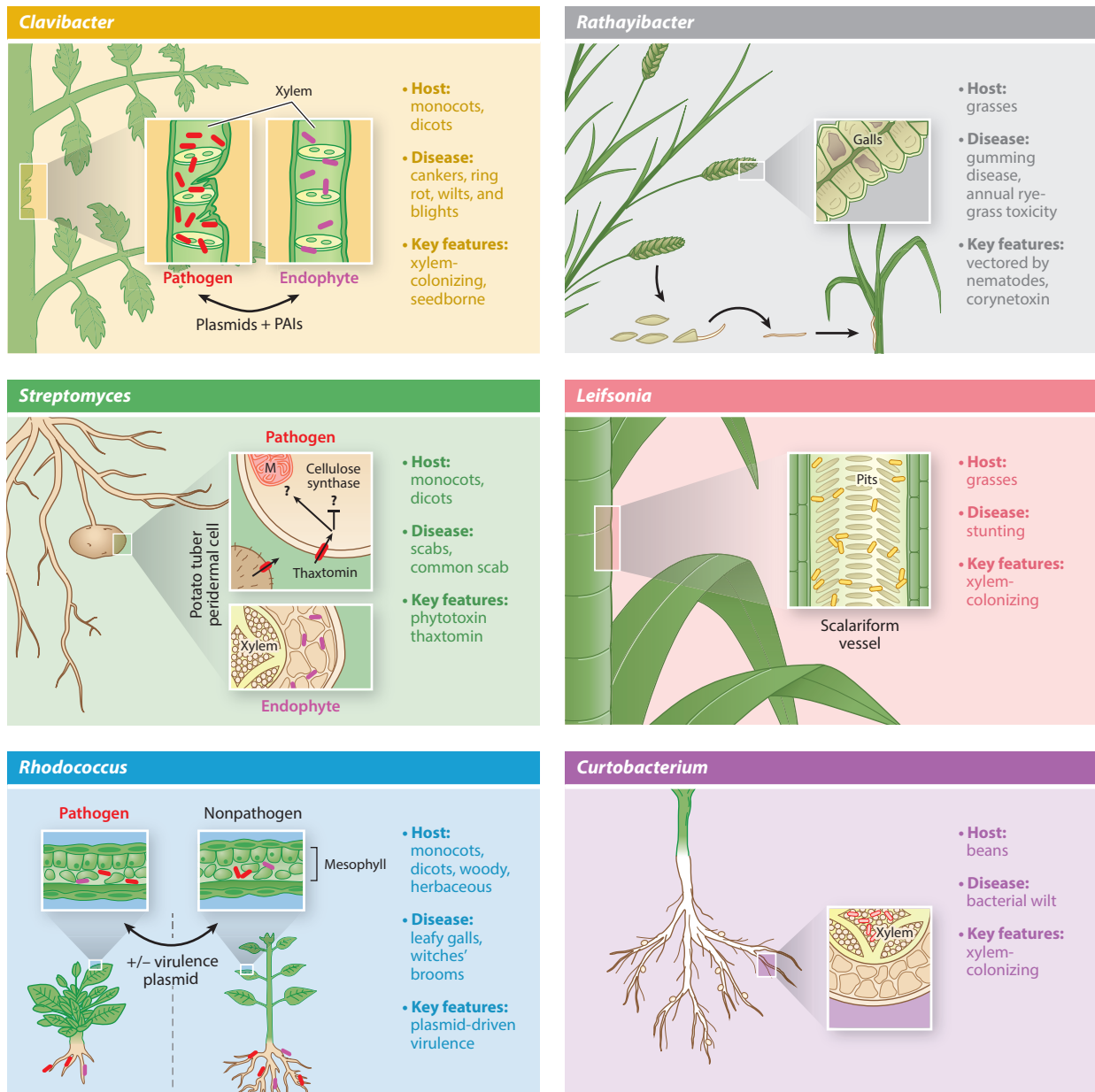
Figure 1

Gram-positive bacteria can persist in various tissues and organs of plants. Pathogenic and nonpathogenic *Clavibacter* colonize the xylem of plants. Phytopathogenicity is often associated with serine proteases and CAZymes (carbohydrate-active enzymes), which are typically encoded on plasmids or a pathogenicity island (PAI). *Streptomyces* bacteria colonize tubers and roots and, in some instances, the endophytic compartment of their host. Common scab disease occurs only when the symbiont produces thaxtomin, the loci of which are present in a PAI. Pathogenic and nonpathogenic *Rhodococcus* colonize the surface and tissues of plants. Phytopathogenicity is associated with a cluster of virulence genes, most often vectored by a plasmid. The acquisition of this plasmid is sufficient to transition *Rhodococcus* to being pathogenic. Members of the *Rathayibacter* genus cause gumming disease in grasses. These bacteria are vectored by nematodes. *Rathayibacter toxicus*, which produces corynetoxin, can be lethal to grazing animals. *Leifsonia* bacteria colonize vascular cells. Two groups currently classified as subspecies include pathogens that cause stunting diseases of grasses. *Curtobacterium* are vascular bacteria, and both pathogenic and endophytic members reside in plants.

these strains do not cause disease symptoms (125, 143). It is also the case that strains of *Clavibacter* pathogenic to one plant species can reside as nonpathogenic endophytes in other plant species (17, 125). *Clavibacter tesellarius*, which is pathogenic to wheat, has been isolated from tomato seeds (125). The potato pathogen *Clavibacter sepedonicus* has been shown to have an endophytic association with sugar beet and *Solanum rostratum* (17, 130). Thus, the spread of pathogenic *Clavibacter* species may be promoted via their dissemination as endophytes associated with plants that do not succumb to their influence as pathogens.

Plasmids are a significant source of variation in the *Clavibacter* genus. Strain NCPPB382 of the tomato pathogen *C. michiganensis* has been well characterized and is a reference for this discussion. Strain NCPPB382 possesses two plasmids, pCM1 and pCM2, both of which are implicated in virulence, and their absence results in a delay in or complete loss of the ability of the bacteria to cause disease symptoms (56, 95). The main virulence genes on the plasmids of NCPPB382 are *celA* (pCM1), a cellulase, and *pat-1* (pCM2), a serine protease (42, 66). *Clavibacter insidiosus* can carry up to three plasmids, two of which are homologous to pCM1 and pCM2 (91).

Plasmid-free derivatives of *C. michiganensis*, *C. sepedonicus*, and *Clavibacter capsici* fail to induce disease symptoms, demonstrating that plasmids are required for virulence on solanaceous hosts (65, 125). But the presence/absence of plasmids is not always predictive of outcomes of symbioses. Some strains of *C. michiganensis* lacking pCM2 are still pathogenic on tomato, indicating that pCM2 is not always necessary for pathogenicity of *C. michiganensis* (125). It is also the case that pCM-like plasmids in pathogenic strains isolated from field-grown plants differ considerably with respect to size and composition compared to those in NCPPB382 (125). Plasmids are not universally associated with virulence in other species of *Clavibacter*. *Clavibacter nebraskensis* is typically free of plasmids, although strains containing plasmids have been reported (1, 4). The putative virulence factors are encoded by genes located in the chromosome of this pathogen (4, 45, 122). Plasmids are rare in *C. tesellarius* and the mechanisms of virulence are unknown (25).

Within the xylem: life in a nutrient-poor environment and the role of carbohydrate-active enzymes and serine proteases. *C. michiganensis* resides, moves, and grows in xylem vessels (45). The xylem has extremely low quantities of organic and inorganic compounds, and vascular pathogens must fulfill their nutritional requirements by acquiring the scarce nutrients, enzymatically digesting host cell walls, invading neighboring cells, and inducing nutrient leakage from surrounding tissues (140). The genome sequences of pathogenic *Clavibacter* are enriched in genes that encode carbohydrate-active enzymes (CAZymes) (23, 55, 125). Relative to other pathogenic or endophytic *Clavibacter* species, *C. michiganensis* is predicted to have the highest number of CAZymes (**Figure 2**). Some families of CAZymes degrade and metabolize glycoconjugates, oligosaccharides, and polysaccharides of plant cell walls and may be used to facilitate bacterial spread, colonization, and acquisition of nutrients (23). *C. michiganensis*, for example, possesses larger numbers of CAZymes involved in degrading cellulose (GH1, GH3) and hemicellulose (GH35) than other members of this genus (125). Glycome profiles of tomato plants inoculated with *C. michiganensis* strain CASJ002 compared to those of uninoculated plants revealed differences in cell wall chemistry with reduced arabinogalactan epitopes and loosening of xyloglucan epitopes, indicating severe modification of the integrity of the plant cell wall (125).

Streptomyces: Taxa of Root-Associated Bacteria

Streptomyces is a large and diverse genus with more than 800 named species. This group is best known for the production of secondary metabolites (129). Fewer than twenty species are known

---

Endophytes: bacteria that do not cause disease and are present within the tissue of plants

Virulence genes: nonessential genes that are necessary for a pathogenic organism to gain a fitness advantage at a cost to the host

Pathogenicity: the ability of an organism to cause disease

Carbohydrate active enzymes (CAZymes): enzymes that degrade, modify, or cleave glycosidic bonds
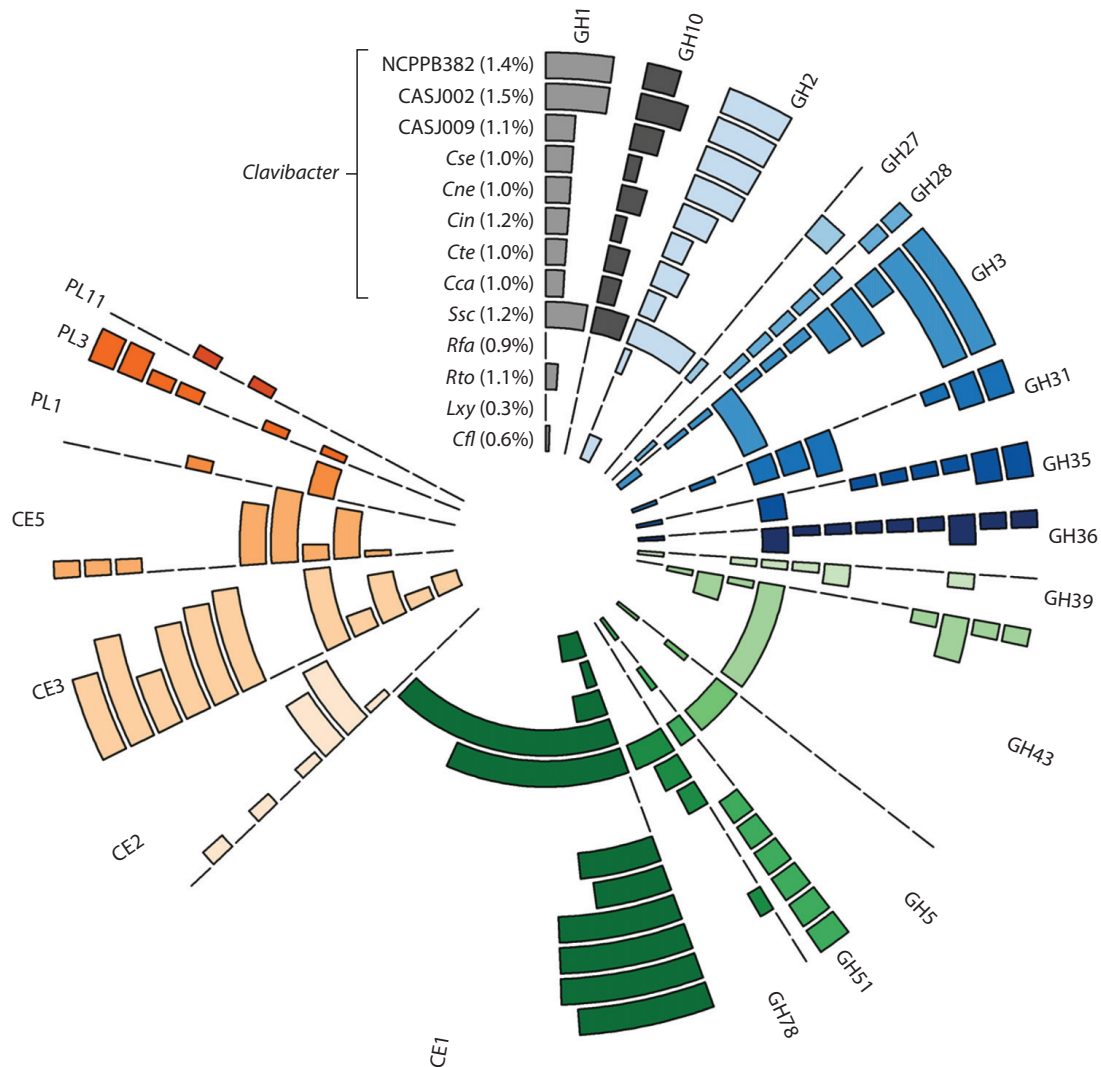
Figure 2

Distribution of predicted plant cell wall–degrading enzymes across Gram-positive bacteria. *Clavibacter michiganensis* subsp. *michiganensis* (NCPPB382, CASJ002), *Clavibacter* tomato endophyte (CASJ009), *Clavibacter sepdonicus* (*Cse*; ATCC33133), *Clavibacter nebraskensis* (*Cne*; NCPPB2581), *Clavibacter insidiosus* (*Cin*; R1–1), *Clavibacter tesellarius* (*Cte*; DOAB 609), *Clavibacter capsici* (*Cca*; PF008), *Streptomyces scabiei* (*Ssc*; 87.22), *Rhodococcus fascians* (*Rfa*; D188), *Rathayibacter toxicus* (*Rto*; WAC3373), *Leifsonia xyli* (*Lxy*; CTCB07), and *Curtobacterium flaccumfaciens* (*Cfl*; UCD-AKU). The percentage of predicted enzymes was calculated relative to the total number of encoded proteins. The bars represent total numbers of predicted enzymes within each category and scale from 1 (shortest bar) to 32 (longest bar) predicted enzymes. Abbreviations: CE, carbohydrate esterase; GH, glycoside hydrolase; PL, polysaccharide lyase.

to include strains that cause common scab disease in plants (**Figure 1**) (15, 136). Some of the more commonly studied pathogenic species include *Streptomyces scabiei* (syn. *Streptomyces scabies*), *Streptomyces turgidiscabiei*, and *Streptomyces acidiscabiei*. Diseases of potato and other root crops caused by this group are prevalent, endemic, and costly because of the expression of raised or pitted lesions that make the tubers nonmarketable (15).

Members of the Actinobacteria phylum, particularly *Streptomyces*, are prominent in microbiota of plants and potentially actively recruited by some plant species (20, 44, 92). It has been speculated that nonpathogenic *Streptomyces* can promote plant growth, as they often produce plant

growth–promoting hormones, but direct effects have not been demonstrated (133). Non-pathogenic *Streptomyces* can also directly, via production of secondary metabolites, and indirectly, via triggering induced systemic resistance, suppress the growth of plant pathogens (105, 133).

Thaxtomin: the root of scab diseases. Thaxtomin is the most important virulence molecule associated with plant-pathogenic lineages of *Streptomyces* (90). Thaxtomin is a phytotoxin, a chemical compound that compromises the normal development of plant cells. Evidence suggests that the thaxtomin locus is necessary for expression of common scab disease, as disruption of *txt* and other genes associated with thaxtomin biosynthesis eliminates thaxtomin production, and mutants fail to elicit symptom formation on plants (61, 70, 79). The thaxtomin locus is located within a *txt* pathogenicity island (PAI) and some strains of *Streptomyces* modified to carry the *txt* PAI gain the ability to cause symptoms typical of common scab disease (18, 78, 90, 147).

Cellobiose and cellotriose are considered the primary triggers for thaxtomin production. These di- and trisaccharides of glucose bind to the CebE protein, characterized in *S. scabiei*, and are shepherded through the CebEFG-MsiK ABC transporter into the cytoplasm (73). Cellobiose and cellotriose then relieve CebR, the cellulose utilization repressor protein, from two binding sites in the thaxtomin biosynthetic cluster (49). This leads to transcriptional upregulation of the thaxtomin biosynthetic genes *txtA* and *txtB* and the thaxtomin biosynthesis transcriptional activator gene *txtR* (70).

The ability of *S. scabiei*, *S. acidiscabiei*, and *S. turgidiscabiei* to hydrolyze cellulose, a process that yields cellobiose and cellotriose, could not be demonstrated (68). This led to speculation that the primary triggers for thaxtomin production originate in the plant, not from the action of the bacteria. This is surprising considering that the *S. scabiei* genome encodes several putative cellulases and other potential cell wall–degrading enzymes (**Figure 2**). Furthermore, these genes are specifically induced by the presence of suberin, a component of the periderm of potato tubers (101). It is possible that previous attempts at quantifying *S. scabiei* cellulase activity did not include the specific elicitor necessary to initiate expression of the cellulases of *Streptomyces*. Therefore, we cannot discount the possibility that the action of scab-causing *Streptomyces* leads to the release of cellobiose and cellotriose and control the transition between saprophyte and pathogen. How detection of cellulose-degradation by-products controls the transition from saprophyte to pathogen remains unknown (72).

The mode of toxicity of thaxtomin is not well established. In potato, symptom development occurs only if the pathogen is present and producing thaxtomin during early tuber development (80). In *Arabidopsis*, thaxtomin causes severe stunting as well as hypertrophy and demonstrably inhibits cellulose synthesis (51, 115), although the inhibition of cellulose synthesis is potentially indirect. Alternatively, the toxicity of thaxtomin may be a consequence of its targeting of Pam16 or associated proteins that likely regulate induction of mitochondrial-generated cell death signals (**Figure 1**) (57, 64).

Rhodococcus: Root- and Leaf-Associated Bacteria

The *Rhodococcus* genus is genetically diverse, and its members have a cosmopolitan distribution (86). They are mycolic acid–containing bacteria and have enzymes belonging to the carbohydrate esterase family 1 (CE1) that are necessary for the biosynthesis of their cell walls. This may explain the high number of predicted CE1 family members in this group (**Figure 2**). *Rhodococcus* has been predominantly studied for their potential uses in bioremediation and biofuels (6, 37). Some are associated with plants, and the majority of such strains are present within a single major group that can be further divided into four clades (113).

7

Phytopathogenic lineages of *Rhodococcus* are present in two of the clades. These bacteria cause proliferations of differentiated shoots called leafy galls or witches' brooms (**Figure 1**) (107). Diseased tissues are hypothesized to be nutrient sinks that can be exploited by the pathogen (119). More than 100 plant taxa, including both monocots and dicots that have herbaceous and woody growth habits, can be infected by pathogenic *Rhodococcus*, and individual plant species can be host to diverse lineages of *Rhodococcus* (107, 113). Pathogenic *Rhodococcus* can be distinguished from nonpathogenic lineages by the unique presence of a cluster of three virulence loci. In the majority of sequenced pathogenic lineages, these virulence loci are harbored on a plasmid (33, 34). Strains cured of the plasmid are not pathogenic, mutants with insertions in the *att* locus have attenuated virulence, and those in which *fasR* and *fas* loci are disrupted are not pathogenic (33).

The data are consistent with a scenario in which plants that associate with nonpathogenic *Rhodococcus* have a selective advantage over those that do not. *Rhodococcus* was identified in surveys of roots and leaves of *Arabidopsis*, rice, clover, and soybean and hypothesized to be selected for by these plant species (20, 38, 83, 92, 135). When inoculated onto seedlings, nonpathogenic *Rhodococcus* can cause significant changes in root architecture and increases in aboveground biomass (50, 113). *Rhodococcus* also has antimicrobial activities and shows potential in protecting plants against pathogens (63).

Two sides to the role of cytokinins in the virulence of *Rhodococcus*. Pathogenic *Rhodococcus* is hypothesized to secrete a mixture of cytokinins that upset signaling in plants (119). Several lines of evidence provide correlative support for this model. Cytokinins are a class of phytohormone for which disruptions of their normal levels can lead to plant growth anomalies, as is the case with symptoms associated with *Rhodococcus* (81). Furthermore, the levels of cytokinins are altered in plant tissues infected with *Rhodococcus* (39, 54). Likewise, cytokinins are detectable in *Rhodococcus* cultures (9, 32, 34, 62, 82, 103, 104). Most importantly, three of the six plasmid-borne *fas* virulence genes in *Rhodococcus* encode enzymes involved in synthesizing and modifying cytokinins (34, 103, 104).

However, not all of the evidence is congruent with such a model of pathogenesis. The *Rhodococcus*-associated changes in levels of most cytokinin types in plants are minor, the temporal patterns are erratic, and their source cannot be definitively associated with the pathogen (39–41, 54). Plants encode an abundance of enzymes that modify and degrade cytokinins to guard against excessive levels (81). Transgenic plants overexpressing cytokinin dehydrogenases, which irreversibly degrade cytokinins, should be more immune to *Rhodococcus* if bacterial-synthesized cytokinins are secreted virulence factors (53). However, contrary to predictions, such transgenic plants exhibited more severe symptoms (39). The necessity of *fas* in the accumulation of cytokinins is also subject to interpretation. Most cytokinins accumulate to minuscule levels in media of culture-grown bacteria and do not accumulate in a *fas*-dependent manner (32, 41, 43). An important confounding factor that cannot be overlooked is that cytokinins have ancient origins and are present in bacteria and eukaryotes as modifications on base 37, 3′-adjacent to the anticodon, of a subset of tRNAs (116). In fact, several studies have associated the accumulation of cytokinin in *Rhodococcus* cultures with the degradation of tRNAs (93, 110, 114).

Rathayibacter: Hitchhiking to Grass Seedheads

Members of the *Rathayibacter* genus cause gumming disease of grasses (98). These bacteria are vectored by juvenile parasitic nematodes of the subfamily Anguininae and gain access to the ovaries of grasses (**Figure 1**). The bacteria are hypothesized to require the nematode to first establish seed galls that they then inhabit, as *Rathayibacter* is not as capable of colonizing plant hosts when

inoculated directly (111). *Rathayibacter* may then outcompete the nematode to occupy the pre-established gall. The relatively low inventory of CAZymes may reflect the reliance of *Rathayibacter* on nematodes for plant colonization (**Figure 2**).

Rathayibacter toxicus is the most infamous species of *Rathayibacter* (98). Grazing animals that consume grasses contaminated with *R. toxicus* can succumb to annual ryegrass toxicity, which is associated with often lethal neurological disorders. This toxicity is due to corynetoxin, a member of the tunicamycin family of nucleotide antibiotics that inhibit the first step of protein glycosylation in eukaryotes (106).

Bacteriophages are another partner in the ecology of *R. toxicus*. The bacteriophages appear to exist in multiple copies as independent replicons in a pseudolysogenic state, which is an extended, arrested, and nonreplicative developmental state (26, 100). Bacteria with such pseudolysogenic phages are also associated with the production of corynetoxin. However, ecological sampling of *R. toxicus* failed to correlate the presence of toxin with the presence of a phage (85). In addition, a putative corynetoxin-associated locus is present in the genome of the bacteria (117). Thus, it is more likely that bacteriophages are indirect triggers for the production of the toxin.

Leifsonia: Xylem-Limited Bacteria

There are 11 species of *Leifsonia*, with only two subspecies of *Leifsonia xyli* recognized as plant pathogens (47, 142). *L. xyli* subsp. *xyli* is the causal agent of ratoon stunting disease of sugarcane and contributes to significant economic losses worldwide (**Figure 1**) (128). *L. xyli* subsp. *xyli* adheres to the surface of scalariform xylem vessels and interferes with water transport, leading to the reduction of height and diameter of sugarcane (145). *L. xyli* subsp. *cynodontis* colonizes the xylem of Bermudagrass, causing Bermudagrass stunting disease (36). Given that these pathogens and those in the genus *Curtobacterium* are similar to *Clavibacter* species in having xylem-limited lifestyles, it was somewhat unexpected that the examined genome sequences had so few predicted CAZymes (**Figure 2**). However, the decay in the genome of *L. xyli* subsp. *xyli* and the fastidious nature of this pathogen could also account for the lower number of predicted CAZymes (16). In many grasses, *L. xyli* subsp. *cynodontis* is an endophyte (97).

Curtobacterium: Xylem-Limited Bacteria

Curtobacterium flaccumfaciens is a seed-transmitted vascular pathogen causing bacterial wilt of common bean (**Figure 1**) (2). This disease is the most important bacterial disease of dry beans in the United States and is a quarantine pathogen in many countries (2). Like other groups of bacteria discussed in this review, the outcomes of the symbioses of members of *Curtobacterium* with plants are variable. Strains of *Curtobacterium* have been isolated as nonpathogenic endophytes, whereas pathogenic *C. flaccumfaciens* has been implicated as being beneficial to plant species other than bean (7, 21, 22, 127).

EVOLUTION OF VIRULENCE

Do Gram-Positive Bacteria and Their Plant Hosts Zig and Zag?

A zigzag model has been developed that fits mechanisms of virulence and immunity into an evolutionary framework of an arms race between microbes and plants (69). Depending on the microbial elicitor, different sectors of the immune system can activate and elicit a response that varies in timing and amplitude. Pattern-triggered immunity is elicited upon the perception of

features conserved in microbes. Many phytopathogenic organisms deploy so-called effector proteins to strike key immunity nodes, dampening responses below a threshold and rendering pattern-triggered immunity ineffective. However, certain genotypes of host species encode proteins that perceive an effector and can mount a rapid effector-triggered immune response. The latter immune response imposes a strong selective pressure, leading to continual loss and gains of effector genes.

If the zigzag model is to serve as a paradigm, Gram-positive phytopathogens are expected to elicit pattern-triggered immunity. Gram-positive bacteria cause changes in host plants indicative of an immune response (10, 11, 112). Their cell envelope includes various antigenic features conserved in bacteria, and peptidoglycan of *Staphylococcus aureus* and *Streptomyces* species, for example, can cause changes in *Arabidopsis* that are consistent with an immune response (48, 59). Gram-positive bacteria also express Ef-Tu, characterized in Gram-negative bacteria as an elicitor of plant immunity, on their surfaces (138). It is also the case that the composition of Gram-positive bacteria in microbiota associated with *Arabidopsis* mutants affected in immune signaling is significantly different relative to that of wild-type plants (87). However, it is unclear whether such differences reflect immunity acting directly on Gram-positive taxa or indirect consequences from changes in priority effects, which are the effects that the sequence in which species arrive have on community composition. Last, it has been suggested that Gram-positive bacteria have promise in biocontrol because of their potential to elicit immunity (105). But we are not aware of any demonstration that a purified conserved microbial feature from a pathogenic member of Actinobacteria is sufficient as an elicitor.

Effector-triggered immunity has also not been demonstrated to confer resistance to Gram-positive bacteria. For *Clavibacter*, quantitative trait loci (QTLs) impacting the severity of disease symptoms have been mapped in wild tomato species (74). Resistance does not appear to be due to classic immunity. For example, the Rcm2.0 QTL overlaps with another QTL controlling an increased rate of stem vascular development and may affect the maturation of the vascular system (29, 74).

Cultivars of potato resistant to *Streptomyces* are available for controlling common scab. But no dominant resistance genes have been identified and only a few putative resistance-associated QTLs have been discovered in *Solanum* germplasm (15). Resistance to scab is hypothesized to be due to morphological changes to the tuber. For example, common scab-tolerant potatoes developed through somatic cell selection with thaxtomin as the selectable marker uniformly have thicker periderms, potentially limiting the accessibility of the toxin and the pathogen into the tuber (139).

For *Rhodococcus*, the only form of resistance that has been characterized is via a secondary metabolite extracted from the bark of the resistant legume *Dalbergia pervillei* (5, 109). This metabolite is hypothesized to interfere with virulence gene signaling.

Sugarcane varieties with the highest levels of resistance to *L. xyli* subsp. *xyli* exhibit highly branched xylem vessels, which presumably restrict bacterial colonization of new vascular bundles (123).

Resistance to *C. flaccumfaciens* is rare. The combined results from two separate screens of more than 2,100 accessions yielded fewer than 40 accessions with reduced symptoms (60). It is unknown whether these are conferred by dominant resistant genes or via QTLs.

In all, classic immunity has not been demonstrated to be a barrier to the proliferation of these bacteria. Thus, the strength of plant immunity as a selective pressure on Gram-positive bacteria is unknown, and there is little context for understanding the mechanisms hypothesized to contribute to their virulence.

10

What Other Evolutionary Processes Could Have Shaped Gram-Positive Bacteria?

One traditional model predicts that the emergence of exploitive lineages occurs in a step-wise fashion, with HGT first innovating cells with novel functions and followed subsequently by more long-term evolutionary changes (126). Virulence evolution of Gram-positive bacteria might, however, be more appropriately modeled according to alternative paradigms.

In species of *Clavibacter*, the combination of both plasmids and PAIs shapes their plant-associated lifestyle (45). The *chp/tomA* PAI is conserved, required for pathogenicity of *C. michiganensis*, and absent from genomes of endophytic strains (45, 55, 125). The *chp/tomA* PAI is predicted to encode many secreted serine proteases and CAZymes. The *tomA* region encodes tomatinase, which is involved in deglycosylating an antibacterial saponin (77). The *tomA* region also encodes several proteins potentially involved in the degradation and transport of host oligosaccharides. But the PAI and even the genomes have undergone substantial modification. In *C. sepedonicus*, the PAI appears to have been fragmented or disrupted (13). *C. capsici* lacks the *chp/tomA* PAI, and the region corresponding to a portion of *chp* that encodes CAZymes and serine proteases that contribute to pathogenicity is present on a plasmid (65). *C. nebraskensis* has few homologs of genes of the *chp/tomA* PAI and must rely on different virulence strategies, with toxin production hypothesized to be important (96, 122). The *C. sepedonicus* genome has a high number of direct repeats due to the presence of 106 insertion elements, which promote pseudogenization and genome plasticity and adaptations to new environments (13, 122).

The divergence of pathogenic *Clavibacter* species, and possibly the emergence of pathogenic *Leifsonia* (see below), may be explained by a model proposed for *Bordetella pertussis* (102). The causative agent of whooping cough has a narrow host range and is a highly virulent pathogen. Relative to sister taxa of *Bordetella*, the production of pertussis toxin is exclusive to *B. pertussis*. However, few other genome gains could be identified that could be associated with the change in the lifestyle of *B. pertussis*. Instead, it was suggested that virulence evolution was shaped by massive losses to the genome, likely due to a proliferation of insertion sequences. The losses were hypothesized to have led to a lineage incapable of curbing determinants that when misregulated are detrimental to the health of hosts. For example, genome decay could have led to the overexpression or misexpression of virulence genes.

In some Gram-positive taxa, HGT seems to operate on a shorter timescale than in Gram-negative bacteria. Several lines of evidence suggest the *txt* PAI of *Streptomyces* is highly mobile and its acquisition leads to the rapid emergence of new lineages that can cause common scab disease. First, the *txt* PAI is polyphyletic and incongruent with the phylogeny of the genus (19, 146). Second, new lineages associated with disease are potentially emerging in real time. A lineage of *Streptomyces bottropensis*, a species once considered to be nonpathogenic, was recently identified on symptomatic plants (149). In a separate study, several strains of *S. bottropensis* with the *txt* PAI were collected, although their ability to cause scab disease has not yet been confirmed (136). Third, *Streptomyces reticuliscabiei* and *Streptomyces turgidiscabies* differ in that the former causes a less severe netted scab disease, whereas the latter can cause severe common scab. *S. reticuliscabiei* and *S. turgidiscabies*, despite their binomial names, are the same species, whereas only members of the latter group have the *txt* PAI (14). Fourth, the *Streptomyces txt* PAIs are diverse, likely because of recombination of multiple subregions (146, 148). These variants can include other genes such as *tomA* or *fas*, which are implicated in the virulence of other Gram-positive bacteria and have been associated with the ability of *Streptomyces* to cause symptoms (71, 118, 146, 147).

It is notable that following acquisition of the *txt* PAI, not all *Streptomyces* species can cause symptoms (78, 147). Perhaps only certain lineages of plant-symbiotic *Streptomyces* encode the traits

11

necessary for the *txt* PAI to function. Such traits could include specific cellulases necessary for the release of the thaxtomin-inducing elicitors. Traits could also be those that enable the pathogen to access the subperiderm, as the structure of the tuber periderm is likely a critical determinant of susceptibility to thaxtomin. There could also be compatibility issues between *txt* PAIs and genomic backgrounds, as *Streptomyces coelicolor* transconjugants with a *txt* PAI integrated into their chromosomes fail to produce thaxtomin (78). Identifying the genetic background necessary for the *txt* PAI to exert its effect is critical for predicting which species of *Streptomyces* have the potential to become common scab pathogens.

The virulence plasmid of *Rhodococcus* is key to driving evolutionary transitions between being beneficial and pathogenic. Eviction of the plasmid from a pathogenic strain led to one that behaves like an associative symbiont that causes changes in the plant predicted to promote plant growth and development (50, 113). The opposite was also observed, as introduction of the plasmid into what were originally associative symbionts converted strains into pathogens that cause leafy galls and terminally arrest seedlings (113). Focus has been on *fas*, but *fasR* is also necessary for pathogenicity (124). FasR is predicted to be a member of the AraC-type transcriptional regulators (124, 141). A model predicts that in the absence of a ligand, AraC-type regulators bind as monomers to a specific DNA sequence and repress expression of target genes. Binding of the regulator to a cognate ligand stimulates a structural change and activates transcription of genes. Despite its name, there is no evidence to suggest that FasR directly regulates the *fas* locus (124). Hence, the mechanism of FasR may be explained by the virulence gene co-option model, in which horizontally acquired genes reprogram core genes and transition recipient bacterial cells into those that can exploit the host (88). Strains that were once beneficial or, at the very least, not detrimental lose control of features that when unchecked can cause symptoms to their hosts.

Disease Symptoms Caused by Gram-Positive Bacteria: Fitness Advantage or Collateral Damage?

A fundamental assumption is that Gram-positive pathogens gain a fitness advantage at a cost to the host. But in the symbioses described here, the benefits reaped and costs exacted on the hosts have not all been sufficiently quantified.

Of the bacterial taxa discussed here, *Clavibacter* most easily meets the strict definition of a pathogen. Pathogenic *Clavibacter* tend to have higher bacterial loads compared to endophytes. *C. michiganensis* strains that experienced a spontaneous deletion of the entire *chp/tomA* PAI have significantly lower growth (55). But bacterial growth and virulence are not always positively correlated (99, 125). It is also the case that the deletion of key genes encoding CAZymes or serine proteases or loss of plasmids that compromise that ability of the bacteria to cause disease symptoms does not lead to large decreases in bacterial titers (99, 125).

In *Streptomyces*, the extent to which the acquisition of the *txt* PAI increases the fitness of the bacteria has not been measured. Tissue lesions may be a novel niche that promotes *Streptomyces* colonization. Alternatively, the lesions may reflect damage that is inconsequential to the bacteria. For example, a selective advantage for producing thaxtomin may be due to its antimicrobial activity against competing rhizosphere microbes. This hypothesized scenario is analogous to one that may also explain the selective pressures acting on *R. toxicus* to produce corynetoxin, an antibiotic that interferes with cell wall biosynthesis (106).

Pathogenic and nonpathogenic *Rhodococcus* achieve similar levels of growth on the surface of plants (30). Within tissues, and relative to nonpathogenic *Rhodococcus*, a strain with the virulence plasmid achieves higher population densities. However, the onset of disease occurred prior to extensive colonization of the interior tissues. Whether entry is necessary for causing symptoms is

unknown. Symptomatic mature plants infected with *Rhodococcus* can be maintained in greenhouses for years without any obvious deleterious fitness effects. In laboratory settings and on seedlings as well as plants propagated via tissue culture, *Rhodococcus* can terminally arrest or stunt primary growth (40, 113, 132). But it has not been determined whether pathogens cause similar effects in native environments.

It is valid to ask whether *Rathayibacter* fits the strict definition of a pathogen. *Rathayibacter* is dependent on the priority effects of plant-parasitic nematodes and participates in the ecological succession of already established diseased plant tissues. The plant does provide habitable space for *Rathayibacter*, but whether this interaction comes at an additional cost to the host that is exacted by *Rathayibacter* is unknown.

The emergence of *L. xyli* subsp. *xyli* as a pathogen is speculated to be an accident resulting from artificial, not natural, selection. It is suggested that an endophyte introduced by one parent species of an interspecies cross causes symptoms in only the hybrids used for commercial sugarcane (16, 142). A survey of more than 100 isolates cultured from diseased sugarcane growing in various parts of the world showed there to be little genetic diversity (142). This is consistent with the possibility of a recent single emergence and clonal expansion of this population.

Some of the Gram-positive bacteria may therefore lack bona fide virulence traits. This does not undermine the importance of these Gram-positive bacteria, as their economic costs to agriculture are undeniable. But recognizing them as pathogens in a looser sense compels us to consider different niches, processes, and selective pressures to understand the evolution of these bacteria.

Sequencing More Genomes: When Is Enough, Enough?

The combination of broad and deep taxon sampling is extremely powerful for providing context for informing on the mechanisms of evolution. At least one member from each of the known species within the *Clavibacter* genus has been sequenced. Comparisons between species have revealed significant variation. *C. nebraskensis* lacks plasmids and has the smallest genome size, consistent with its specialized lifestyle within leaf tissue of corn (122). The coding genes of *C. sepedonicus* have the lowest similarity to homologs in other members of the genus, indicating increased divergence (122). Of the species groups, only *C. michiganensis* has been deeply sampled. Comparisons within genome sequences of pathogenic *C. michiganensis* revealed that all possess the *chp/tomA* PAI, the virulence plasmid pCM1, and a variety of CAZyme-encoding genes, indicating these genetic features are critical for pathogenicity (122, 125). More than 1,200 genome sequences from *Streptomyces* are available. But, regrettably, the quality for many are below levels sufficient for comparative genomics and only a handful are from plant pathogens (120).

The *Rhodococcus* data set is large and diverse (32, 113). Analyses revealed the diversity of pathogens and the general lack of diversity in virulence plasmids. But natural variation in virulence loci was revealed. Two pathogenic strains were discovered that have virulence genes integrated into their chromosomes. One of these experienced a nonorthologous displacement of the *fas* locus with a single gene that is predicted to encode a single protein with two of the *fas*-encoded functions (32). This variant may provide valuable insights into the mechanisms of *fas*-dependent virulence. The whole genome data set was also analyzed to model transmission of pathogens and identified epidemiological links, reservoir populations, and the potential for dissemination of virulence plasmids between strains. In addition, the data set included genetically related nonpathogenic strains, which were critical as contrasts for developing models of HGT-driven, rapid evolutionary transitions within clades (113).

Comparisons between deeply sampled taxa of *Rathayibacter* revealed remarkably little natural variation in *R. toxicus* (35). Within species, the comparisons of genome sequences and spacer arrays

of clustered regularly interspaced short palindromic repeat (CRISPR) loci led to a hypothesis that *R. toxicus* populations are under periodic hard selection (35). In this scenario, *R. toxicus* populations experience repeated bottlenecks caused by blooms of bacteriophages (121, 137). Associated with such blooms is a diversity of bacteriophages, of which some evade CRISPR immunity, and drive susceptible genotypes of *R. toxicus* to extinction while selecting for the rare lineages with spacers that confer immunity. The result is geographically separated groups of *R. toxicus* with extremely depressed levels of genetic variation (3, 8).

USING PHYLOGENOMICS TO GENERATE HYPOTHESES

How Are the Strains of Gram-Positive Bacteria Related?

We used multiple gene loci to construct a maximum-likelihood tree of strains with high-quality genome sequences (**Figure 3a**; original tree provided in Newick format as **Supplemental Material 1**). Two whole genome–based methods, for operationally defining genera and species, were then used to provide context and cross-validate conclusions (**Figure 3b,c**; **Supplemental Material 2 and 3**) (84, 108). Despite filtering, we cannot exclude the possibility that the quality of genome assemblies affected the results (**Supplemental Material 4**).

Streptomyces and *Rhodococcus* are both cohesive units that can be operationally classified as distinct genus-level groups (**Figure 3**). The relationship among genera within the Microbacteriaceae family is more contentious (**Figure 3b**). All species of *Clavibacter* cluster together and with most species of the cohesive *Rathayibacter* group. *Leifsonia* does not form a cohesive group. It can generally be divided into three subgroups. Subgroup 1, consisting of endophytic *Leifsonia* strains, clusters with all *Clavibacter*, most *Rathayibacter*, and subgroup 2 of *Leifsonia*. Subgroup 2, also consisting of endophytic strains, clusters with some species of *Rathayibacter*, most species of *Clavibacter*, and all *Leifsonia*. Pathogenic strains of subgroup 3 of *Leifsonia* cluster only with subgroup 2 of *Leifsonia* and *R. toxicus*. *Curtobacterium* clusters with *Clavibacter*, most *Rathayibacter*, and subgroups 1 and 2 of *Leifsonia*. *Pseudoclavibacter* forms a distinct group.

Subgroup 3 and *R. toxicus* consistently confounded classification and the importance of their comparisons was down-weighted because their significantly smaller genome sizes likely compromised the whole genome–based method used for operational classification. The findings therefore suggest *Clavibacter*, *Rathayibacter*, *Curtobacterium*, and even *Leifsonia* are more closely related, as was originally proposed, than their current classification conveys (36).

There are also notable observations at the species level. Some results are consistent with previously reported findings. Strains CF11 and LMG 26808 of *Clavibacter* are distinctly different from *C. michiganensis* (89). The type strain of *S. reticuliscabiei* (NRRL B-24446) grouped with strains *S. turgidiscabies* T45 and Car8 (14). In addition, results here show the type strain of *S. scabiei* (NRRL B-16523) grouped with only 10 of the 15 strains classified as *S. scabiei*. In *Leifsonia*, two pathogenic strains of *L. xyli* subsp. *xyli* cultured from sugarcane are significantly different from the *L. xyli* subsp. *cynodontis* pathogenic strain cultured from Bermudagrass, and the growth-promoting *L. xyli* SE134 strain is different relative to each of the three aforementioned *Leifsonia* strains (76).

Genetic distance influences the granularity of data, and conclusions need to be framed around the evolutionary relationships of the strains being compared (12). As exemplified in the next section, comparing two phenotypically different strains, e.g., pathogenic versus nonpathogenic, of the same species has power in allowing one to generate a hypothesis on the genetic bases for differences. However, comparing strains that were incorrectly assigned to the same species has the potential to lead to erroneous conclusions. In the following, we use this new phylogenomic framework to generate some testable hypotheses.
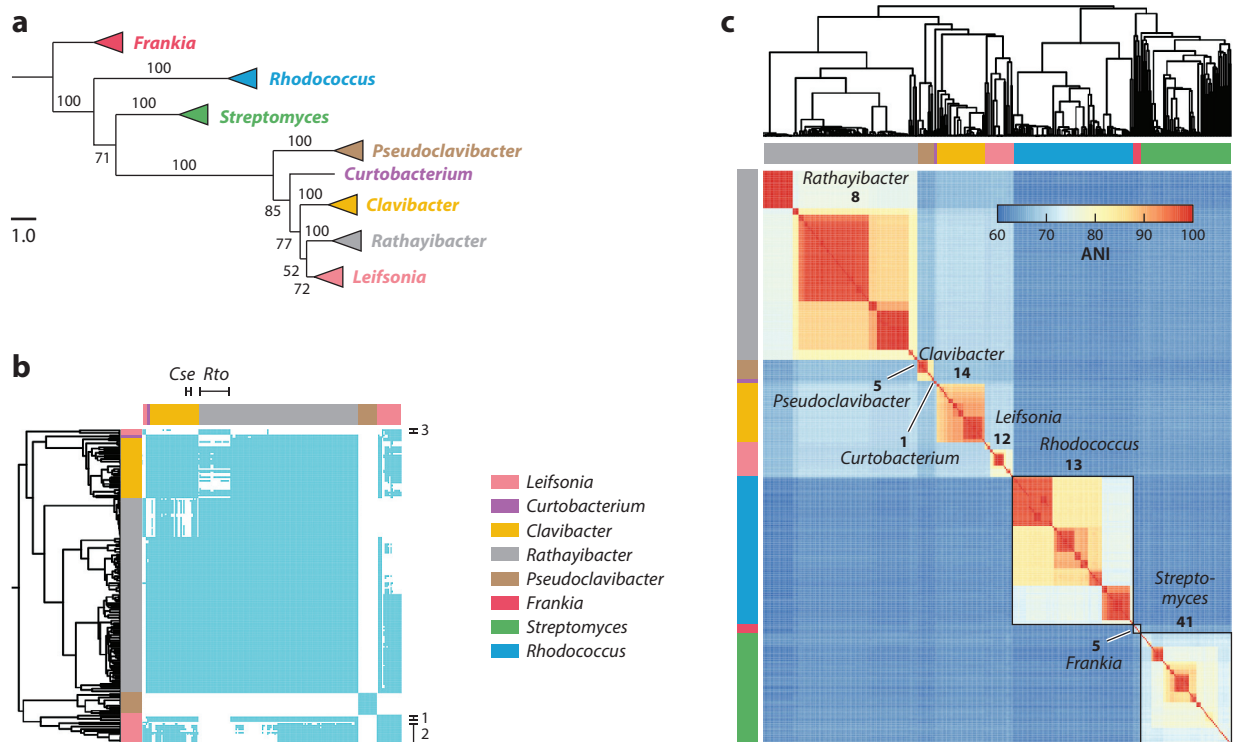
Figure 3

Relationship among genera with phytopathogenic members within the Actinomycetales order. (*a*) A maximum-likelihood phylogenetic tree constructed on the basis of the genes: *ftsY*, *infH*, *rpoB*, *rsmA*, *secY*, *tsaD*, and *ychF*. The symbiosis phenotypes of these strains have not been characterized. Pathogenic *Leifsonia* were not included because some of the homologs of genes failed to meet the coverage threshold used. *Frankia* of the Frankiaceae family was used to root the tree. Nodes were collapsed at the genus level. Colors are coded to represent the taxonomic units and used throughout the figure. Bar indicates amino acid substitutions per site. (*b*) A heatmap constructed on the basis of the percentage of conserved proteins (POCP). A threshold of ≥50% POCP values (*cyan*) was used to operationally classify genus-level groups; only members of the Microbacteriaceae are shown because *Streptomyces* and *Rhodococcus* were clearly delineated (see panel *c*). The 1–3 subgroups of *Leifsonia*, *Clavibacter sepdonicus* (*Cse*), and *Rathayibacter toxicus* (*Rto*) are indicated. A fourth group consisting of *Leifsonia aquatica* was not denoted because members are not known to interact with plants. (*c*) A heatmap constructed on the basis of average nucleotide identity (ANI) values (*see scale*). A threshold of ≥94% ANI values was used to operationally classify species-level groups. Genera in which all members met POCP thresholds are boxed in black. The numbers of ANI groups for each of the currently recognized taxonomic units are indicated.

How Did Virulence Evolve in Taxa of Gram-Positive Bacteria?

Previously sequenced endophytic strains of *Clavibacter* cluster with pathogenic *C. insidiosus*, *C. capsici*, and *C. tesellarius* but not *C. michiganensis* (125). On the basis of thresholds used here, results suggest the endophytes are in the same species-level groups as pathogens, thereby providing contrasts to study the evolution of virulence in these species of *Clavibacter*. The uncharacterized strains Z001 and Z002 cluster closely with *C. michiganensis* (35). We found no evidence that these two strains carry pCM1 or pCM2 or have homologs of *celA* and *pat-1*. In addition, both lack the *chp/tomA* PAI. Assuming the two strains are not pathogenic, this is consistent with the hypothesis that acquisition of both the plasmid and *chp/tomA* is necessary to transition *C. michiganensis* into being pathogenic. S. *reticuliscabiei* and *S. turgidiscabies* cluster into a single species-level group (14). We hypothesize that the horizontal acquisition of the *txt* PAI by members of this group is

sufficient to confer upon recipients the ability to cause common scab disease. Lastly, on the basis of having a small genome size and the relationships among *Leifsonia*, we hypothesize that a process analogous to what occurred in *B. pertussis* led to the two subspecies of pathogenic *L. xyli*. Given the low pairwise values, we cannot exclude the possibility that the two subspecies independently emerged from different recent ancestors.

Is There Evidence for Gram-Positive Bacteria Encoding Elicitors of Plant Immunity?

Under current models, populations under conflict are predicted to be rapidly diversifying. Gram-negative pathogens present evidence for diversifying selection acting on regions of candidate elicitors of plant immunity (94, 134). We therefore used a similar approach as previously published for a preliminary examination of whether adaptive changes can be similarly detected in Gram-positive bacteria (94). We focused on the deeply sampled *Rhodococcus* group. With the exception of Ef-Tu, none of the candidate elicitors previously identified have a homolog in *Rhodococcus*. Members of clades I and II are pathogenic, whereas those in clades III and IV are not, and *Rhodococcus opacus* are soil bacteria (113). There is a core (present in ≥95% of the sequences analyzed) of 1,317 single-copy genes (**Supplemental Material 5**). The core genes have an average $dN/dS = 0.076$, which is consistent with the core genome being under strong negative selection. A total of 377 (28.6%) translated sequences had at least one site with evidence for diversifying selection. We focused on those in which there was evidence for positive selection acting on clusters of two or more sites within a window of 30 codons. A total of 85 genes present in members of clades I + II and 27 genes present in members of clade III + IV met these criteria. In *R. opacus*, 54 genes met the criteria. Of the genes associated with clades I and II, only three were identified in clades III and IV, and seven were identified in *R. opacus*. A total of 76 were unique to members of clades I and II; none of the genes encodes Ef-Tu but five are predicted to encode membrane-associated proteins (WP_027496706.1, WP_032383528.1, WP_027495305.1, WP_032382129.1, WP_027496936.1). These results are consistent with the possibility that *Rhodococcus* encodes elicitors of plant immunity that have not been previously characterized, but these candidates will need to be experimentally tested.

Is the Type VII Secretion System Used for Virulence?

A specialized type III secretion system is a necessary virulence factor for many Gram-negative bacterial pathogens of plants (27). This system deploys protein effectors into plants cells, typically to suppress immunity. Similarly, three of the five type VII secretion systems (T7SSs) of *Mycobacterium tuberculosis* function to secrete protein effectors and are necessary for or contribute to virulence (58). Thus, the T7SS could possibly function in the virulence of phytopathogenic Gram-positive bacteria. But the T7SS is also essential for viability and is involved in conjugation. In addition, mutants of *S. scabiei* affected in T7SS-associated genes still caused symptoms but were altered in development (52).

We used *essC/eccC*, a marker gene, to mine genome sequences for putative T7SS loci. A tree constructed on the basis of translated sequences formed two clades (**Figure 4**). Most sequences clustered with homologs from *Mycobacterium*. Inspection of these loci suggests that members of *Streptomyces*, *Rhodococcus*, and subgroup 1 of *Leifsonia* have T7SS loci that are most similar in composition to the ESX-4 T7SS locus, which encodes a system that has no known role in virulence and was only recently shown in *Mycobacterium smegmatis* to be involved in mediating social interactions between bacteria (**Figure 4b**) (28).
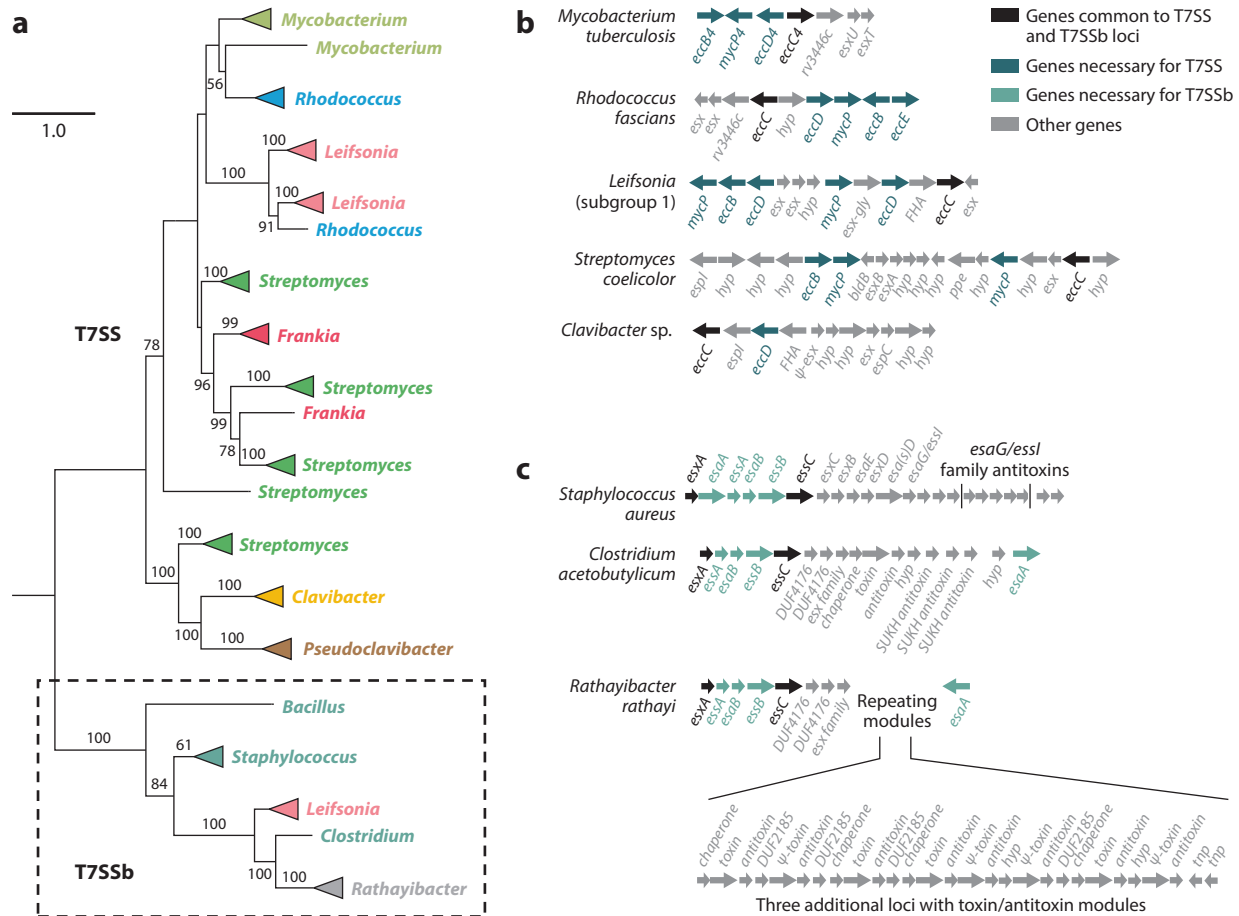
Figure 4

Type VII secretion system (T7SS)-encoding loci of plant-associated members of the Actinomycetales order were independently acquired. (*a*) A maximum-likelihood phylogenetic tree constructed on the basis of translated sequences of *essC/eccC*. Nodes in which bootstrap support are not shown have values below 50. Bar indicates amino acid substitutions per site. (*b*) Structures of representative T7SS loci of Actinobacteria. (*c*) Structure of representative T7SSb loci of Firmicutes (*Bacillus*, *Staphylococcus*, and *Clostridium*). For both panel *b* and panel *c*, translated sequences of T7SS-associated genes were used as queries to search the protein sequences. The corresponding loci were manually inspected. The *esx* and *ppe* genes encode domains associated with T7SS-secreted proteins. Arrows represent coding sequences, with directions indicating strand. Abbreviations: *hyp*, hypothetical; Ψ, predicted pseudogene.

Unexpectedly, homologs from two species of *Rathayibacter* and those of subgroup 2 of *Leifsonia* clustered with *essC/eccC* of T7SSb loci in members of Firmicutes. In this phylum, the T7SSb variant functions in interbacterial competition and the locus includes toxin genes and cognate antitoxin genes (24). These toxin genes typically also encode an LXG domain, which has been used to predict polymorphic toxins potentially translocated via the T7SS apparatus (24, 144). The T7SSb loci of *Rathayibacter* include repetitive exported toxin–antitoxin modules, each of which has a corresponding chaperone-encoding gene (**Figure 4***c*). Most of the translated sequences of the putative toxin-encoding genes have identifiable LXG domains. The modularity is unique relative to other T7SS and T7SSb loci. We therefore suggest the putative T7SSs of these plant-associated bacteria are not specialized for virulence but function to influence interactions between bacteria.

17

There is no evidence for T7SS or T7SSb loci in the genome sequences of members of subgroup 3 of *Leifsonia*. The observation that the subgroups of *Leifsonia* are polymorphic with respect to this secretion system is consistent with the inferred divergence within the group reflecting true biological relationships as opposed to an artifact of sequence assembly or analyses (**Figure 3**). Lastly, the locus we examined for *Clavibacter* does not appear to encode a functional system.

SUMMARY POINTS

1. Gram-positive bacteria are important symbionts that positively and negatively impact plants.

2. Gram-positive bacteria are different from Gram-negative bacteria in how they interact with plants.

3. HGT can rapidly enable members of some groups of Gram-positive bacteria to transition to being pathogenic.

4. Diseases caused by some Gram-positive bacteria have economic costs, but fitness costs to the host have yet to be quantified.

5. Current models in molecular plant pathology are useful for comparing and contrasting, but not explicitly for predicting the evolution and mechanisms of interactions between Gram-positive bacteria and plants.

6. There is currently no available classic, dominant resistance gene that confers resistance to Gram-positive pathogens.

7. The understanding of plant symbioses with Gram-positive bacteria will be accelerated by framing studies on the bases of their natural variation and autecology.

8. Phylogeny and whole genome–based methods can be used to provide a strong framework for studying Gram-positive bacteria.

FUTURE ISSUES

1. Models for the interactions between Gram-positive bacteria and plants need to be developed.

2. It will be important to identify the selective advantages for Gram-positive bacteria in adopting a plant-associated or pathogenic lifestyle.

3. The ability of Gram-positive bacteria to trigger and suppress plant immune responses must be investigated.

4. The diversity of populations of plant-associated, Gram-positive bacteria need to be quantified.

5. We need to characterize the mechanisms by which CelA and Pat-1 perturb plants.

6. We need to characterize the mechanism by which thaxtomin causes symptoms.

7. The genes co-opted by FasR must be identified.

8. The ecological features that drive HGT of plasmids and PAIs need to be identified.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

1. Agarkova IV, Lambrecht PA, Vidaver AK. 2011. Genetic diversity and population structure of *Clavibacter michiganensis* subsp. *nebraskensis*. *Can. J. Microbiol.* 57(5):366–74

2. Agarkova IV, Lambrecht PA, Vidaver AK, Harveson RM. 2012. Genetic diversity among *Curtobacterium flaccumfaciens* pv. *flaccumfaciens* populations in the American high plains. *Can. J. Microbiol.* 58(6):788–801

3. Agarkova IV, Vidaver AK, Postnikova EN, Riley IT, Schaad NW. 2006. Genetic characterization and diversity of *Rathayibacter toxicus*. *Phytopathology* 96(11):1270–77

4. Ahmad A, Mbofung GY, Acharya J, Schmidt CL, Robertson AE. 2015. Characterization and comparison of *Clavibacter michiganensis* subsp. *nebraskensis* strains recovered from epiphytic and symptomatic infections of maize in Iowa. *PLOS ONE* 10(11):e0143553

5. Almabruk KH, Chang JH, Mahmud T. 2016. Total synthesis of (±)-isoperbergins and correction of the chemical structure of perbergin. *J. Nat. Prod.* 79(9):2391–96

6. Alvarez HM, Steinbuchel A. 2002. Triacylglycerols in prokaryotic microorganisms. *Appl. Microbiol. Biotechnol.* 60(4):367–76

7. Araújo WL, Marcon J, Maccheroni W, Van Elsas JD, Van Vuurde JWL, Azevedo JL. 2002. Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. *Appl. Environ. Microbiol.* 68(10):4906–14

8. Arif M, Busot GY, Mann R, Rodoni B, Liu S, Stack JP. 2016. Emergence of a new population of *Rathayibacter toxicus*: an ecologically complex, geographically isolated bacterium. *PLOS ONE* 11(5): e0156182

9. Armstrong DJ, Scarbrough E, Skoog F. 1976. Cytokinins in *Corynebacterium fascians* cultures: isolation and identification of 6-(4-Hydroxy-3-methyl-cis-2-butenylamino)-2-methylthiopurine. *Plant Physiol.* 58(6):749–52

10. Balaji V, Mayrose M, Sherf O, Jacob-Hirsch J, Eichenlaub R, et al. 2008. Tomato transcriptional changes in response to *Clavibacter michiganensis* subsp. *michiganensis* reveal a role for ethylene in disease development. *Plant Physiol.* 146(4):1797–809

11. Balaji V, Sessa G, Smart CD. 2011. Silencing of host basal defense response-related gene expression increases susceptibility of *Nicotiana benthamiana* to *Clavibacter michiganensis* subsp. *michiganensis*. *Phytopathology* 101(3):349–57

12. Baltrus DA. 2016. Divorcing strain classification from species names. *Trends Microbiol*. 24(6):431–39

13. Bentley SD, Corton C, Brown SE, Barron A, Clark L, et al. 2008. Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. *J. Bacteriol*. 190(6):2150–60

14. Bouchek-Mechiche K, Gardan L, Andrivon D, Normand P. 2006. *Streptomyces turgidiscabies* and *Streptomyces reticuliscabiei*: one genomic species, two pathogenic groups. *Int. J. Syst. Evol. Microbiol*. 56(Pt. 12):2771–76

15. Braun S, Gevens A, Charkowski A, Allen C, Jansky S. 2017. Potato common scab: a review of the causal pathogens, management practices, varietal resistance screening methods, and host resistance. *Am. J. Potato Res*. 94(4):283–96

16. Brumbley SM, Petrasovits LA, Hermann SR, Young AJ, Croft BJ. 2006. Recent advances in the molecular biology of *Leifsonia xyli* subsp. *xyli*, causal organism of ratoon stunting disease. *Australas. Plant Pathol*. 35(6):681–89

17. Bugbee WM, Gudmestad NC, Secor GA, Nolte P. 1987. Sugar beet as a symptomless host for *Corynebacterium sepedonicum*. *Phytopathology* 77:765–70

18. Bukhalid RA, Loria R. 1997. Cloning and expression of a gene from *Streptomyces scabies* encoding a putative pathogenicity factor. *J. Bacteriol*. 179(24):7776–83

19. Bukhalid RA, Takeuchi T, Labeda D, Loria R. 2002. Horizontal transfer of the plant virulence gene, *nec1*, and flanking sequences among genetically distinct *Streptomyces* strains in the Diastatochromogenes cluster. *Appl. Environ. Microbiol*. 68(2):738–44

20. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, et al. 2012. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* 488(7409):91–95

21. Bulgari D, Casati P, Brusetti L, Quaglino F, Brasca M, et al. 2009. Endophytic bacterial diversity in grapevine (*Vitis vinifera* L.) leaves described by 16S rRNA gene sequence analysis and length heterogeneity-PCR. *J. Microbiol*. 47(4):393–401

22. Bulgari D, Minio A, Casati P, Quaglino F, Delledonne M, Bianco PA. 2014. *Curtobacterium* sp. genome sequencing underlines plant growth promotion-related traits. *Genome Announc*. 2(4):e00592-14

23. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res*. 37(Database Issue):D233–38

24. Cao Z, Casabona MG, Kneuper H, Chalmers JD, Palmer T. 2016. The type VII secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria. *Nat. Microbiol*. 2:16183

25. Carlson RR, Vidaver AK. 1982. Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *Int. J. Syst. Bacteriol*. 32(3):315–26

26. Cenens W, Makumi A, Mebrhatu MT, Lavigne R, Aertsen A. 2013. Phage-host interactions during pseudolysogeny: lessons from the Pid/dgo interaction. *Bacteriophage* 3(1):e25029

27. Chang JH, Desveaux D, Creason AL. 2014. The ABCs and 123s of bacterial secretion systems in plant pathogenesis. *Annu. Rev. Phytopathol*. 52:317–45

28. Clark RR, Judd J, Lasek-Nesselquist E, Montgomery SA, Hoffmann JG, et al. 2018. Direct cell-cell contact activates SigM to express the ESX-4 secretion system in *Mycobacterium smegmatis*. *PNAS* 115(28):E6595–603

29. Coaker GL, Meulia T, Kabelka EA, Jones AK, Francis DM. 2002. A QTL controlling stem morphology and vascular development in *Lycopersicon esculentum* × *Lycopersicon hirsutum* (Solanaceae) crosses is located on chromosome 2. *Am. J. Bot*. 89(12):1859–66

30. Cornelis K, Ritsema T, Nijsse J, Holsters M, Goethals K, Jaziri M. 2001. The plant pathogen *Rhodococcus fascians* colonizes the exterior and interior of the aerial parts of plants. *Mol. Plant-Microbe Interact*. 14(5):599–608

31. Cottyn B, Regalado E, Lanoot B, De Cleene M, Mew TW, Swings J. 2001. Bacterial populations associated with rice seed in the tropical environment. *Phytopathology* 91(3):282–92

32. Creason AL, Vandeputte OM, Savory EA, Davis EW, Putnam ML, et al. 2014. analysis of genome sequences from plant pathogenic *Rhodococcus* reveals genetic novelties in virulence loci. *PLOS ONE* 9(7):e101996

33. Crespi M, Messens E, Caplan AB, Van Montagu M, Desomer J. 1992. Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. *EMBO J.* 11(3):795–804

34. Crespi M, Vereecke D, Temmerman W, Van Montagu M, Desomer J. 1994. The *fas* operon of *Rhodococcus fascians* encodes new genes required for efficient fasciation of host plants. *J. Bacteriol.* 176(9):2492–501

35. Davis EW, Tabima JF, Weisberg AJ, Lopes LD, Wiseman MS, et al. 2018. Evolution of the U.S. Biological Select Agent *Rathayibacter toxicus*. *mBio* 9(4):e01280-18

36. Davis MJ, Gillaspie AG, Vidaver AK, Harris RW. 1984. *Clavibacter*: a new genus containing some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. *xyli* sp. nov., subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease. *Int. J. Syst. Bacteriol.* 34(2):107–17

37. de Carvalho CC, Costa SS, Fernandes P, Couto I, Viveiros M. 2014. Membrane transport systems and the biodegradation potential and pathogenicity of genus *Rhodococcus*. *Front. Physiol.* 5:133

38. Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, et al. 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *PNAS* 106(38):16428–33

39. Depuydt S, Dolezal K, Van Lijsebettens M, Moritz T, Holsters M, Vereecke D. 2008. Modulation of the hormone setting by *Rhodococcus fascians* results in ectopic KNOX activation in *Arabidopsis. Plant Physiol.* 146(3):1267–81

40. Dhandapani P, Song J, Novák O, Jameson PE. 2016. Infection by *Rhodococcus fascians* maintains cotyledons as a sink tissue for the pathogen. *Ann. Bot.* 119:841–52

41. Dhandapani P, Song J, Novák O, Jameson PE. 2018. Both epiphytic and endophytic strains of *Rhodococcus fascians* influence transporter gene expression and cytokinins in infected *Pisum sativum* L. seedlings. *Plant Growth Regul.* 85(2):231–42

42. Dreier J, Meletzus D, Eichenlaub R. 1997. Characterization of the plasmid encoded virulence region *pat-1* of phytopathogenic *Clavibacter michiganensis* subsp. *michiganensis. Mol. Plant-Microbe Interact.* 10(2):195–206

43. Eason J, Morris R, Jameson P. 1996. The relationship between virulence and cytokinin production by *Rhodococcus fascians* (Tilford 1936) Goodfellow 1984. *Plant Pathol.* 45(2):323–31

44. Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, et al. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *PNAS* 112(8):E911–20

45. Eichenlaub R, Gartemann K-H. 2011. The *Clavibacter michiganensis* subspecies: molecular investigation of gram-positive bacterial plant pathogens. *Annu. Rev. Phytopathol.* 49:445–64

46. Eichenlaub R, Gartemann K-H, Burger A. 2006. *Clavibacter michiganensis*, a group of gram-positive phytopathogenic bacteria. In *Plant-Associated Bacteria*, ed. SS Gnanamanickam, pp. 385–421. Dordrecht, Neth.: Springer

47. Evtushenko LI, Dorofeeva LV, Subbotin SA, Cole JR, Tiedje JM. 2000. *Leifsonia poae* gen. nov., sp. nov., isolated from nematode galls on *Poa annua*, and reclassification of "*Corynebacterium aquaticum*" Leifson 1962 as *Leifsonia aquatica* (ex Leifson 1962) gen. nov., nom. rev., comb. nov. and *Clavibacter xyli* Davis et al. 1984 with two subspecies as *Leifsonia xyli* (Davis et al. 1984) gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 50(Pt. 1):371–80

48. Felix G, Boller T. 2003. Molecular sensing of bacteria in plants. The highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. *J. Biol. Chem.* 278(8):6201–8

49. Francis IM, Jourdan S, Fanara S, Loria R, Rigali S. 2015. The cellobiose sensor CebR is the gatekeeper of *Streptomyces scabies* pathogenicity. *mBio* 6(2):e02018-14

50. Francis IM, Stes E, Zhang Y, Rangel D, Audenaert K, Vereecke D. 2016. Mining the genome of *Rhodococcus fascians*, a plant growth-promoting bacterium gone astray. *New Biotechnol.* 33(5 Pt. B):706–17

51. Fry BA, Loria R. 2002. Thaxtomin A: evidence for a plant cell wall target. *Physiol. Mol. Plant Pathol.* 60(1):1–8

52. Fyans JK, Bignell D, Loria R, Toth I, Palmer T. 2013. The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen *Streptomyces scabies. Mol. Plant Pathol.* 14(2):119–30

53. Galis I, Bilyeu KD, Godinho MJG, Jameson PE. 2005. Expression of three *Arabidopsis* cytokinin oxidase/dehydrogenase promoter::GUS chimeric constructs in tobacco: response to developmental and biotic factors. *Plant Growth Regul.* 45(3):173–82

54. Galis I, Bilyeu K, Wood G, Jameson PE. 2005. *Rhodococcus fascians*: shoot proliferation without elevated cytokinins? *Plant Growth Regul.* 46(2):109–15

55. Gartemann K-H, Abt B, Bekel T, Burger A, Engemann J, et al. 2008. The genome sequence of the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382 reveals a large island involved in pathogenicity. *J. Bacteriol.* 190(6):2138–49

56. Gartemann K-H, Kirchner O, Engemann J, Gräfen I, Eichenlaub R, Burger A. 2003. *Clavibacter michiganensis* subsp. *michiganensis*: first steps in the understanding of virulence of a Gram-positive phytopathogenic bacterium. *J. Biotechnol.* 106(2–3):179–91

57. Gray J, Rustgi S, von Wettstein D, Reinbothe C, Reinbothe S. 2016. Common functions of the chloroplast and mitochondrial co-chaperones cpDnaJL (CDF1) and mtDnaJ (PAM16) in protein import and ROS scavenging in *Arabidopsis thaliana*. *Commun. Integr. Biol.* 9(5):e1119343

58. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* 14(11):677–91

59. Gust AA, Biswas R, Lenz HD, Rauhut T, Ranf S, et al. 2007. Bacteria-derived peptidoglycans constitute pathogen-associated molecular patterns triggering innate immunity in *Arabidopsis*. *J. Biol. Chem.* 282(44):32338–48

60. Harveson RM, Schwartz HF, Urrea CA, Yonts CD. 2015. Bacterial wilt of dry-edible beans in the Central High Plains of the U.S.: past, present, and future. *Plant Dis.* 99(12):1665–77

61. Healy FG, Wach M, Krasnoff SB, Gibson DM, Loria R. 2000. The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol.* 38(4):794–804

62. Helgeson JP, Leonard NJ. 1966. Cytokinins: identification of compounds isolated from *Corynebacterium fascians. PNAS* 56(1):60–63

63. Hong CE, Jeong H, Jo SH, Jeong JC, Kwon S-Y, et al. 2016. A leaf-inhabiting endophytic bacterium, *Rhodococcus* sp. KB6, enhances sweet potato resistance to black rot disease caused by *Ceratocystis fimbriata*. *J. Microbiol. Biotechnol.* 26(3):488–92

64. Huang Y, Chen X, Liu Y, Roth C, Copeland C, et al. 2013. Mitochondrial AtPAM16 is required for plant survival and the negative regulation of plant immunity. *Nat. Commun.* 4:2558

65. Hwang IS, Oh E-J, Kim D, Oh C-S. 2018. Multiple plasmid-borne virulence genes of *Clavibacter michiganensis* ssp. *capsici* critical for disease development in pepper. *New Phytol.* 217(3):1177–89

66. Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R. 2000. The endo-beta-1,4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. *Mol. Plant-Microbe Interact.* 13(7):703–14

67. Jankute M, Cox JAG, Harrison J, Besra GS. 2015. Assembly of the mycobacterial cell wall. *Annu. Rev. Microbiol.* 69:405–23

68. Johnson EG, Joshi MV, Gibson DM, Loria R. 2007. Cello-oligosaccharides released from host plants induce pathogenicity in scab-causing *Streptomyces* species. *Physiol. Mol. Plant Pathol.* 71(1–3):18–25

69. Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444(7117):323–29

70. Joshi MV, Bignell DRD, Johnson EG, Sparks JP, Gibson DM, Loria R. 2007. The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and virulence in *Streptomyces scabies*. *Mol. Microbiol.* 66(3):633–42

71. Joshi MV, Loria R. 2007. *Streptomyces turgidiscabies* possesses a functional cytokinin biosynthetic pathway and produces leafy galls. *Mol. Plant-Microbe Interact.* 20(7):751–58

72. Jourdan S, Francis IM, Deflandre B, Loria R, Rigali S. 2017. Tracking the subtle mutations driving host sensing by the plant pathogen *Streptomyces scabies*. *mSphere* 2(2):e00367-16

73. Jourdan S, Francis IM, Kim MJ, Salazar JJC, Planckaert S, et al. 2016. The CebE/MsiK transporter is a doorway to the cello-oligosaccharide-mediated induction of *Streptomyces scabies* pathogenicity. *Sci. Rep.* 6:27144

74. Kabelka E, Franchino B, Francis DM. 2002. Two loci from *Lycopersicon hirsutum* LA407 confer resistance to strains of *Clavibacter michiganensis* subsp. *michiganensis*. *Phytopathology* 92(5):504–10

75. Kaneshiro WS, Mizumoto CY, Alvarez AM. 2006. Differentiation of *Clavibacter michiganensis* subsp. *michiganensis* from seed-borne saprophytes using ELISA, Biolog and 16S rDNA sequencing. *Eur. J. Plant Pathol.* 116(1):45–56

76. Kang S-M, Asaf S, Kim S-J, Yun B-W, Lee I-J. 2016. Complete genome sequence of plant growth-promoting bacterium *Leifsonia xyli* SE134, a possible gibberellin and auxin producer. *J. Biotechnol.* 239:34–38

77. Kaup O, Gräfen I, Zellermann E-M, Eichenlaub R, Gartemann K-H. 2005. Identification of a tomatinase in the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. *Mol. Plant-Microbe Interact.* 18(10):1090–98

78. Kers JA, Cameron KD, Joshi MV, Bukhalid RA, Morello JE, et al. 2005. A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Mol. Microbiol.* 55(4):1025–33

79. Kers JA, Wach MJ, Krasnoff SB, Widom J, Cameron KD, et al. 2004. Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature* 429(6987):79–82

80. Khatri BB, Tegg RS, Brown PH, Wilson CR. 2011. Temporal association of potato tuber development with susceptibility to common scab and *Streptomyces scabies*-induced responses in the potato periderm. *Plant Pathol.* 60(4):776–86

81. Kieber JJ, Schaller GE. 2014. Cytokinins. *Arabidopsis Book* 12:e0168

82. Klämbt D, Thies G, Skoog F. 1966. Isolation of cytokinins from *Corynebacterium fascians*. *PNAS* 56(1):52–59

83. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, et al. 2012. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J.* 6(7):1378–90

84. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *PNAS* 102(7):2567–72

85. Kowalski MC, Cahill D, Doran TJ, Colegate SM. 2007. Development and application of polymerase chain reaction-based assays for *Rathayibacter toxicus* and a bacteriophage associated with annual ryegrass (*Lolium rigidum*) toxicity. *Aust. J. Exp. Agric.* 47(2):177–83

86. Larkin MJ, De Mot R, Kulakov LA, Nagy I. 1998. Applied aspects of *Rhodococcus* genetics. *Antonie van Leeuwenhoek* 74(1–3):133–53

87. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, et al. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349(6250):860–64

88. Letek M, González P, Macarthur I, Rodríguez H, Freeman TC, et al. 2010. The genome of a pathogenic *Rhodococcus*: cooptive virulence underpinned by key gene acquisitions. *PLOS Genet.* 6(9):e1001145

89. Li X, Tambong J, Yuan KX, Chen W, Xu H, et al. 2018. Re-classification of *Clavibacter michiganensis* subspecies on the basis of whole-genome and multi-locus sequence analyses. *Int. J. Syst. Evol. Microbiol.* 68(1):234–40

90. Loria R, Bignell DRD, Moll S, Huguet-Tapia JC, Joshi MV, et al. 2008. Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*. *Antonie van Leeuwenhoek* 94(1):3–10

91. Lu Y, Ishimaru CA, Glazebrook J, Samac DA. 2018. Comparative genomic analyses of *Clavibacter michiganensis* subsp. *insidiosus* and pathogenicity on *Medicago truncatula*. *Phytopathology* 108(2):172–85

92. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, et al. 2012. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488(7409):86–90

93. Matsubara S, Armstrong DJ, Skoog F. 1968. Cytokinins in tRNA of *Corynebacterium fascians*. *Plant Physiol.* 43(3):451–53

94. McCann HC, Nahal H, Thakur S, Guttman DS. 2012. Identification of innate immunity elicitors using molecular signatures of natural selection. *PNAS* 109(11):4215–20

95. Meletzus D, Bermphol A, Dreier J, Eichenlaub R. 1993. Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. *J. Bacteriol.* 175(7):2131–36

96. Michalke A, Galla H-J, Steinem C. 2001. Channel activity of a phytotoxin of *Clavibacter michiganense* ssp. *nebraskense* in tethered membranes. *Eur. Biophys. J.* 30(6):421–29

97. Mills L, Leaman TM, Taghavi SM, Shackel L, Dominiak BC, et al. 2001. *Leifsonia xyli*-like bacteria are endophytes of grasses in eastern Australia. *Australas. Plant Pathol.* 30(2):145–51

98. Murray TD, Schroeder BK, Schneider WL, Luster DG, Sechler A, et al. 2017. *Rathayibacter toxicus*, other *Rathayibacter* species inducing bacterial head blight of grasses, and the potential for livestock poisonings. *Phytopathology* 107(7):804–15

99. Nissinen R, Xia Y, Mattinen L, Ishimaru CA, Knudson DL, et al. 2009. The putative secreted serine protease Chp-7 is required for full virulence and induction of a nonhost hypersensitive response by *Clavibacter michiganensis* subsp. *sepedonicus*. *Mol. Plant-Microbe Interact.* 22(7):809–19

100. Ophel KM, Bird AF, Kerr A. 1993. Association of bacteriophage particles with toxin production by *Clavibacter toxicus*, the causal agent of annual ryegrass toxicity. *Phytopathology* 83(6): 676–81

101. Padilla-Reynaud R, Simao-Beaunoir A-M, Lerat S, Bernards MA, Beaulieu C. 2015. Suberin regulates the production of cellulolytic enzymes in *Streptomyces scabies*, the causal agent of potato common scab. *Microbes Environ.* 30(3):245–53

102. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* 35(1):32–40

103. Pertry I, Václavíková K, Depuydt S, Galuszka P, Spíchal L, et al. 2009. Identification of *Rhodococcus fascians* cytokinins and their modus operandi to reshape the plant. *PNAS* 106(3):929–34

104. Pertry I, Václavíková K, Gemrotová M, Spíchal L, Galuszka P, et al. 2010. *Rhodococcus fascians* impacts plant development through the dynamic *fas*-mediated production of a cytokinin mix. *Mol. Plant-Microbe Interact.* 23(9):1164–74

105. Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. *Annu. Rev. Phytopathol.* 52:347–75

106. Price NPJ, Tsvetanova B. 2007. Biosynthesis of the tunicamycins: a review. *J. Antibiot.* 60(8):485–91

107. Putnam ML, Miller ML. 2007. *Rhodococcus fascians* in herbaceous perennials. *Plant Dis.* 91(9):1064–76

108. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, et al. 2014. A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196(12):2210–15

109. Rajaonson S, Vandeputte OM, Vereecke D, Kiendrebeogo M, Ralambofetra E, et al. 2011. Virulence quenching with a prenylated isoflavanone renders the Malagasy legume *Dalbergia pervillei* resistant to *Rhodococcus fascians*. *Environ. Microbiol.* 13(5):1236–52

110. Rathbone MP, Hall RH. 1972. Concerning the presence of the cytokinin, N^6-(Δ^2-isopentnyl) adenine, in cultures of *Corynebacterium fascians*. *Planta* 108(2):93–102

111. Riley IT, McKay AC. 1991. Inoculation of *Lolium rigidum* with *Clavibacter* sp., the toxigenic bacteria associated with annual ryegrass toxicity. *J. Appl. Bacteriol.* 71:302–6

112. Savidor A, Teper D, Gartemann K-H, Eichenlaub R, Chalupowicz L, et al. 2012. The *Clavibacter michiganensis* subsp. *michiganensis*-tomato interactome reveals the perception of pathogen by the host and suggests mechanisms of infection. *J. Proteome Res.* 11(2):736–50

113. Savory EA, Fuller SL, Weisberg AJ, Thomas WJ, Gordon MI, et al. 2017. Evolutionary transitions between beneficial and phytopathogenic *Rhodococcus* challenge disease management. *eLife* 6:e30925

114. Scarbrough E, Armstrong DJ, Skoog F, Frihart CR, Leonard NJ. 1973. Isolation of *cis*-zeatin from *Corynebacterium fascians* cultures. *PNAS* 70(12 Pt 1–2):3825–29

115. Scheible W-R, Fry B, Kochevenko A, Schindelasch D, Zimmerli L, et al. 2003. An *Arabidopsis* mutant resistant to thaxtomin A, a cellulose synthesis inhibitor from *Streptomyces* species. *Plant Cell* 15(8):1781–94

116. Schweizer U, Bohleber S, Fradejas-Villar N. 2017. The modified base isopentenyladenosine and its derivatives in tRNA. *RNA Biol.* 14(9):1197–208

117. Sechler AJ, Tancos MA, Schneider DJ, King JG, Fennessey CM, et al. 2017. Whole genome sequence of two *Rathayibacter toxicus* strains reveals a tunicamycin biosynthetic cluster similar to *Streptomyces chartreusis*. *PLOS ONE* 12(8):e0183005

118. Seipke RF, Loria R. 2008. *Streptomyces scabies* 87–22 possesses a functional tomatinase. *J. Bacteriol.* 190(23):7684–92

119. Stes E, Vandeputte OM, El Jaziri M, Holsters M, Vereecke D. 2011. A successful bacterial coup d'état: how *Rhodococcus fascians* redirects plant development. *Annu. Rev. Phytopathol.* 49:69–86

120. Studholme DJ. 2016. Genome update. Let the consumer beware: *Streptomyces* genome sequence quality. *Microb. Biotechnol.* 9(1):3–7

121. Sun CL, Thomas BC, Barrangou R, Banfield JF. 2016. Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J.* 10(4):858–70

122. Tambong JT. 2017. *Clavibacter michiganensis* subspecies, pathogens of important agricultural crops. *PLOS ONE* 12(3):e0172295

123. Teakle DS, Appleton JM, Steindl DRL. 1978. An anatomical basis for resistance of sugar cane to ratoon stunting disease. *Physiol. Plant Pathol.* 12(1):83–91

124. Temmerman W, Vereecke D, Dreesen R, Van Montagu M, Holsters M, Goethals K. 2000. Leafy gall formation is controlled by *fasR*, an AraC-type regulatory gene in *Rhodococcus fascians*. *J. Bacteriol.* 182(20):5832–40

125. Thapa SP, Pattathil S, Hahn MG, Jacques M-A, Gilbertson RL, Coaker G. 2017. Genomic analysis of *Clavibacter michiganensis* reveals insight into virulence strategies and genetic diversity of a gram-positive bacterial pathogen. *Mol. Plant-Microbe Interact.* 30(10):786–802

126. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. 2016. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14(4):235–50

127. Ulrich K, Ulrich A, Ewald D. 2008. Diversity of endophytic bacterial communities in poplar grown under field conditions. *FEMS Microbiol. Ecol.* 63(2):169–80

128. Urashima AS, Marchetti LBL. 2013. Incidence and severity of *Leifsonia xyli* subsp. *xyli* infection of sugarcane in Sao Paulo state, Brazil. *J. Phytopathol.* 161(7–8):478–84

129. van der Meij A, Worsley SF, Hutchings MI, van Wezel GP. 2017. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev.* 41(3):392–416

130. van der Wolf JM, van Beckhoven J, Hukkanen A, Karjalainen R, Müller P. 2005. Fate of *Clavibacter michiganensis* ssp. *sepedonicus*, the causal organism of bacterial ring rot of potato, in weeds and field crops. *J. Phytopathol.* 153(6):358–65

131. Vega FE, Pava-Ripoll M, Posada F, Buyer JS. 2005. Endophytic bacteria in *Coffea arabica* L. *J. Basic Microbiol.* 45(5):371–80

132. Vereecke D, Burssens S, Simón-Mateo C, Inzé D, Van Montagu M, et al. 2000. The *Rhodococcus fascians*–plant interaction: morphological traits and biotechnological applications. *Planta* 210(2): 241–51

133. Viaene T, Langendries S, Beirinckx S, Maes M, Goormachtig S. 2016. *Streptomyces* as a plant's best friend? *FEMS Microbiol. Ecol.* 92(8):fiw119

134. Vinatzer BA, Monteil CL, Clarke CR. 2014. Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annu. Rev. Phytopathol.* 52:19–43

135. Vorholt JA. 2012. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* 10(12):828–40

136. Wanner LA. 2009. A patchwork of *Streptomyces* species isolated from potato common scab lesions in North America. *Am. Potato J.* 86(4):247–64

137. Weinberger AD, Sun CL, Pluciński MM, Denef VJ, Thomas BC, et al. 2012. Persisting viral sequences shape microbial CRISPR-based immunity. *PLOS Comp. Biol.* 8(4):e1002475

138. Widjaja M, Harvey KL, Hagemann L, Berry IJ, Jarocki VM, et al. 2017. Elongation factor Tu is a multifunctional and processed moonlighting protein. *Sci. Rep.* 7(1):11227

139. Wilson CR, Tegg RS, Wilson AJ, Luckman GA, Eyles A, et al. 2010. Stable and extreme resistance to common scab of potato obtained through somatic cell selection. *Phytopathology* 100(5):460–67

140. Yadeta KA, Thomma BPHJ. 2013. The xylem as battleground for plant hosts and vascular wilt pathogens. *Front. Plant Sci.* 4:97

141. Yang J, Tauschek M, Robins-Browne RM. 2011. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends Microbiol.* 19(3):128–135

142. Young AJ, Petrasovits LA, Croft BJ, Gillings M, Brumbley SM. 2006. Genetic uniformity of international isolates of *Leifsonia xyli* subsp. *xyli*, causal agent of ratoon stunting disease of sugarcane. *Australas. Plant Pathol.* 35(5):503–11

143. Zaluga J, Van Vaerenbergh J, Stragier P, Maes M, de Vos P. 2013. Genetic diversity of non-pathogenic *Clavibacter* strains isolated from tomato seeds. *Syst. Appl. Microbiol.* 36(6):426–35

144. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. 2012. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* 7:18

145. Zhang X, Chen M, Liang Y, Xing Y, Yang L, et al. 2016. Morphological and physiological responses of sugarcane to *Leifsonia xyli* subsp. *xyli* infection. *Plant Dis.* 100(12):2499–506

146. Zhang Y, Bignell DRD, Zuo R, Fan Q, Huguet-Tapia JC, et al. 2016. Promiscuous pathogenicity islands and phylogeny of pathogenic *Streptomyces* spp. *Mol. Plant-Microbe Interact.* 29(8):640–50

147. Zhang Y, Jiang G, Ding Y, Loria R. 2018. Genetic background affects pathogenicity island function and pathogen emergence in *Streptomyces*. *Mol. Plant Pathol.* 19(7):1733–41

148. Zhang Y, Loria R. 2017. Emergence of novel pathogenic *Streptomyces* species by site-specific accretion and *cis*-mobilization of pathogenicity islands. *Mol. Plant-Microbe Interact.* 30(1):72–82

149. Zhou B, Zhang MS, Ma X-K. 2016. First report of *Streptomyces bottropensis* causing potato common scab in Hebei Province, China. *Plant Dis.* 101(3):502

150. Zinniel DK, Lambrecht P, Harris NB, Feng Z, Kuczmarski D, et al. 2002. Isolation and characterization of endophytic colonizing bacteria from agronomic crops and prairie plants. *Appl. Environ. Microbiol.* 68(5):2198–208

CHAPTER 1

A Genetic Toolkit for Investigating *Clavibacter* Species: Markerless Deletion, Permissive Site Identification, and an Integrative Plasmid

Danielle M. Stevens[1,2], Andrea Tang[2], and Gitta Coaker[2*]

Abstract

The development of knockout mutants and expression variants are critical for understanding genotype-phenotype relationships. However, advancements of these techniques in Gram-positive actinobacteria have stagnated over the last decade. Actinobacteria in the *Clavibacter* genus are composed of diverse crop pathogens which cause a variety of wilt and cankering diseases. Here, we present a suite of tools for genetic manipulation in the tomato pathogen *C. michiganensis* including a markerless deletion system, an integrative plasmid, and an R package for identification of permissive sites for plasmid integration. The vector pSelAct-KO is a recombination based, markerless knockout system that uses dual selection to engineer seamless deletions of a region of interest, providing opportunities for repeated higher-order genetic knockouts. The efficacy of pSelAct-KO was demonstrated in *C. michiganensis* and confirmed using whole genome sequencing. We developed permissR, an R package to identify permissive sites for chromosomal integration, which can be used in conjunction with pSelAct-Express, a non-replicating integrative plasmid that enables recombination into a permissive genomic location. Expression of eGFP by pSelAct-Express was verified in two candidate permissive regions predicted by permissR in *C. michiganensis*. These molecular tools are essential advancements for investigating Gram-positive actinobacteria, particularly for important pathogens in the *Clavibacter* genus.

Author contributions

I wrote the initial drafts of the manuscript and addressed subsequent edits as suggested by my co-authors. AT developed the vector pSelAct-Express from pSelAct-KO and pKGT-GFP for integrative chromosomal expression. I adapted pSelAct-KO to make unmarked deletions in *Clavibacter,* developed the R package permissR to find permissive sites in whole bacterial genomes, and tested both pSelAct-KO and pSelAct-Express in *C. michiganensis* type strain NCPPB382. I developed all figures in the manuscript.

*e-*X tra*

A Genetic Toolkit for Investigating *Clavibacter* Species: Markerless Deletion, Permissive Site Identification, and an Integrative Plasmid

Danielle M. Stevens,[1,2] Andrea Tang,[2] and Gitta Coaker[2,†]

[1] Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA 95616, U.S.A.
[2] Department of Plant Pathology, University of California, Davis, Davis, CA 95616, U.S.A.

The development of knockout mutants and expression variants are critical for understanding genotype-phenotype relationships. However, advances in these techniques in gram-positive actinobacteria have stagnated over the last decade. Actinobacteria in the *Clavibacter* genus are composed of diverse crop pathogens that cause a variety of wilt and cankering diseases. Here, we present a suite of tools for genetic manipulation in the tomato pathogen *Clavibacter michiganensis* including a markerless deletion system, an integrative plasmid, and an R package for identification of permissive sites for plasmid integration. The vector pSelAct-KO is a recombination-based, markerless knockout system that uses dual selection to engineer seamless deletions of a region of interest, providing opportunities for repeated higher-order genetic knockouts. The efficacy of pSelAct-KO was demonstrated in *C. michiganensis* and was confirmed using whole-genome sequencing. We developed permissR, an R package to identify permissive sites for chromosomal integration, which can be used in conjunction with pSelAct-Express, a nonreplicating integrative plasmid that enables recombination into a permissive genomic location. Expression of enhanced green fluorescent protein by pSelAct-Express was verified in two candidate permissive regions predicted by permissR in *C. michiganensis*. These molecular tools are essential advances for investigating gram-positive actinobacteria, particularly for important pathogens in the *Clavibacter* genus.

Gram-positive actinobacteria in the *Clavibacter* genus are economically important xylem-colonizing bacterial pathogens that can infect both monocots and dicots, resulting in canker and

[†]Corresponding author: G. Coaker; glcoaker@ucdavis.edu

The author(s) declare no conflict of interest.

wilting diseases (Eichenlaub and Gartemann 2011; Thapa et al. 2019). Despite current disease control measures including good horticultural practices and monitoring seed stock, outbreaks of *Clavibacter* pathogens have occurred in recent years (Peritore-Galve et al. 2021). The corn pathogen *Clavibacter nebraskensis* has been problematic for many corn-producing Midwestern states and the tomato pathogen *C. michiganensis* has caused notable losses in years past (Ahmad et al. 2015; Nandi et al. 2018; Peritore-Galve et al. 2021). While outbreaks are sporadic, four of the six species are considered quarantine organisms by the European and Mediterranean Plant Protection Organization, with *C. michiganensis* acting as a potential threat to both greenhouse and field production (Nandi et al. 2018; Peritore-Galve et al. 2021). Despite their agricultural impact, little is known about how these pathogens cause disease (Nandi et al. 2018; Thapa et al. 2019).

In other bacterial pathogens, genes important for pathogenicity and host range include secreted protein effectors that act either in the apoplast or inside cells to suppress host immunity, alter host metabolism, and enable colonization, providing a fitness advantage. Bacterial effectors are released by a variety of secretion systems including type II, III, IV, and V as well as Sec and Tat (Dalio et al. 2018; Ponciano et al. 2003). The most well-studied *Clavibacter* species is the tomato pathogen *C. michiganensis*. *C. michiganensis* carries effectors on two plasmids, pCM1 and pCM2, and within the approximately 129-kb *chp/tomA* pathogenicity island (PAI) (Meletzus et al. 1993). While the entire *chp/tomA* PAI is only found in *C. michiganensis*, fragments or homologs of PAI members, or both, can be found in the genomes of other pathogenic *Clavibacter* species (Bentley et al. 2008; Hwang et al. 2018; Lu et al. 2018; Tambong 2017). The *chp/tomA* PAI encodes suites of putative serine proteases belonging to the Sbt, Chp, and Ppa families and carbohydrate activating enzymes (CAZymes) (Thapa et al. 2017). A few *C. michiganensis* effectors have been characterized, including the tomatinase *tomA* and the cellulase *celA* (Jahr et al. 2000; Meletzus et al. 1993). However, the plant targets of most effectors are unknown (Nandi et al. 2018).

Pioneering work in the late 1990s through the early 2010s resulted in vectors for gene deletion and expression in *Clavibacter* spp., but there is still reliance on these tools despite known limitations. Mutants depend on gene replacement with an antibiotic cassette or by transposon mutagenesis (Gartemann and Eichenlaub 2001; Kirchner et al. 2001; Peritore-Galve et al. 2021; Thapa et al. 2017). Engineering gene expression relies on using a plasmid modified from the backbone of native pCM1/pCM2 plasmids from *C. michiganensis*. This requires recipient strains to lack the native plasmid, a confounding factor because known virulence

genes are carried on pCM1/pCM2. Gene expression can also be achieved by transposition via a transposase-based vector, which can have unexpected effects depending on the genetic location inserted (Chalupowicz et al. 2012; Laine et al. 1996; Mullens and Jamann 2021; Tancos et al. 2013).

Genetic tool kits in other bacterial systems enabled functional characterization of genetic drivers in disease development. The original pSelAct vector was designed to make gene deletions in the foal actinobacterial pathogen, *Rhodococcus equi* (van der Geize et al. 2008). pSelAct was modified through the addition of a Gateway cassette into the multiple cloning site and was used to make single gene deletions in the distantly related actinobacterial plant pathogen *Rhodococcus fascians* (later referred to as pSelAct-KO [Savory et al. 2020]). A second group independently developed pMP201, for deletions in *Streptomyces* spp., which functions similarly to the pSelAct vector, using a positive selection marker and counter selection (Dubeau et al. 2008). Integrative expression vectors have been developed for gram-negative bacteria, but there are few examples in gram-positives. In *Ralstonia solanacearum* GMI1000 and related strains, the integrative plasmid series pRC enables insertion and expression of genetic material in a defined location (Monteiro et al. 2012). Finally, a knock-out vector using the sucrose counter-selection marker *sacB* was modified for integrative expression in *Pseudomonas syringae* pv. *tomato* DC3000 (Lee et al. 2018). Flexible genetic tools facilitate rapid advances in studying pathogens regardless of the system.

Here, we present a suite of genetic tools to facilitate investigation of *Clavibacter* spp., including a markerless deletion system, an integrative plasmid, and an R package for identification of permissive sites for plasmid integration. We optimized pSelAct-KO, a recombination-based system that allows for development of higher-order markerless deletions. pSelAct-KO was used to make a 5.6-kb deletion in the *C. michiganensis chp/tomA* PAI comprising five effectors (*chpE* through *ppaC*), and the deletion was confirmed using whole-genome sequencing. We developed the R package permissR and pSelAct-Express to work in conjunction to allow for identification of permissive genomic locations and subsequent targeted plasmid integration. These genetic tools represent a clear advance in our ability to genetically investigate *Clavibacter* spp. by allowing for markerless deletions (decrease polar effects that occur from antibiotic insertions), enable the generation of higher-order mutants, and enable chromosomal expression (allows for gene expression and complementation without removal of endogenous plasmids). In total, these tools should accelerate reverse genetics and subsequent genotype-phenotype studies in *Clavibacter* spp.

RESULTS AND DISCUSSION

pSelAct system for markerless deletions in *Clavibacter* spp.

The utility of two-step systems has been demonstrated in other actinobacteria including *Rhodococcus* and *Streptomyces* bacteria as well as other gram-positive bacteria (Dubeau et al. 2008; Kostner et al. 2017; Savory et al. 2020). Thus, generating markerless deletions using selection and counter-selective methods could likely be applied *Clavibacter* and other related taxa. We employed the pSelAct-KO vector to *C. michiganensis* with the strategy illustrated in Figure 1 for construct design, selection, and counter-selection. In order to develop the knockout construct, approximately 1.5-kb regions flanking the gene or gene locus of interest are amplified (Fig. 1A). Fragments from the 5′ and 3′ flanking regions are cloned into the pSelAct-KO Gateway C Cassette, through either Gateway cloning via LR Clonase from a Gateway pDONR vector or a homology-based cloning approach (InFusion or Gibson) to generate pSelActΔgoi (Fig. 1B)

(Savory et al. 2020). pSelActΔgoi is transformed into *Escherichia coli*, is screened via PCR, and is verified via Sanger sequencing, and is subsequently transformed into the wild-type *Clavibacter* strain (Fig. 1C). Positive recombinants are resistant to the antibiotic apramycin and are screened for their merodiploid state (haploid cells that have duplicated genetic material) using PCR (Fig. 1D). Growing the merodiploid on minimal medium supplemented with 5-fluorocytosine (5-FC) will promote a recombination event to generate a clean (scarless) deletion (Fig. 1E).

Determining apramycin resistance in C. michiganensis. The pSelAct-KO system employs selection using the aminoglycoside antibiotic apramycin to select for chromosomal integration, followed by counter-selection with 5-FC to recombine out. To determine the minimum inhibitory concentration (MIC) of apramycin, MIC growth curves were performed in *C. michiganensis*. In liquid rich tryptone broth with yeast (TBY) media, the MIC necessary to limit growth was 50 µg/ml (Fig. 2B). On solid agar TBY plates, the MIC required was higher at 100 µg/ml (Supplementary Fig. S1). Notably, if plates were incubated beyond 5 days post–*C. michiganensis* transformation, smaller colonies appeared to form (data not shown). This dissipated if *Clavibacter* transformants were plated on TBY that had double the MIC, at 200 µg/ml. However, TBY plates with 100 µg/ml apramycin were sufficient for downstream screening (Supplementary Table S1). Susceptibility to the antibiotic apramycin in *C. michiganensis* differed on plates compared with other actinobacterial pathogens, which require lower concentrations (Savory et al. 2020). A TblastN search of the apramycin-resistance gene in *Clavibacter* spp. (taxid: 1573) yielded no significant results and antibiotics within the same drug class, neomycin and gentamycin, are known to be effective (Supplementary Table S2) (Meletzus and Eichenlaub 1991). While it is unclear why a higher concentration is required, apramycin is effective against *C. michiganensis*.

Deletion of a 5.6-kb gene cluster in the tomato pathogen C. michiganensis. To demonstrate the utility of the pSelAct-KO for gene deletion, we focused on the *chp* region in the PAI of *C. michiganensis* strain CASJ002. PAI-localized effectors have been of interest since their discovery and are hypothesized to drive disease development in the *C. michiganensis*–tomato pathosystem (Eichenlaub and Gartemann 2011; Nandi et al. 2018; Thapa et al. 2019). Within the *chpE-ppaC* region, there are several putative virulence genes, including four serine proteases and the virulence CAZyme *pelA1* (Fig. 2A). Three proteases are from the Ppa family and one from the Chp family. Four of the five genes of interest have been individually deleted (Chalupowicz et al. 2017; Thapa et al. 2017). Mutants of the proteases *chpE*, *ppaA*, and *ppaC* had no reported effect on disease development when tested on tomato. Others have noted, however, that there may be functional redundancy and deleting one protease effector may not be sufficient for disease reduction (Peritore-Galve et al. 2021). The CAZyme *pelA1* mutant had reduced wilting symptoms on tomato. The protease *ppaB1* has not yet been tested for its contribution in disease development.

To test pSelAct-KO in *C. michiganensis*, we developed primers to amplify flanking regions before *chpE* and after *ppaC* (Fig. 2A). Using homology-based inFusion cloning, the fragments replaced the Gateway cassette of pSelAct-KO and a positive clone was selected and sequence-verified (Fig. 1A and B). Upon transformation of *C. michiganensis* wild-type strain CASJ002, the vector will recombine with one of the two homology arms (Fig. 1C). PCR was used to screen for the merodiploid state; first recombinants will provide two bands, one that is the same as wild type and one that is smaller, representative of the recombination event (Figs. 1D, a gray box highlights two

amplicons, and 2C). It is important to note that, as the deletion size increases, it may be more difficult to amplify the upper band representing the wild-type state. Since the merodiploid confers antibiotic resistance whereas the knockout does not, this phenotype can be used as a quick guide. However, if necessary, in the case of multigene deletions, it may be worthwhile to design a probe that binds to one of the flanking regions and to run a Southern blot.

To develop the knockout through secondary recombination, the merodiploid was plated in dilutions from 10^{-6} to 10^{-7} on minimal M9 (mM9) medium supplemented with the counter-selectant 5-FC. Strains are required to be insensitive to the cysteine analog 5-FC but sensitive to its more toxic counterpart 5-florourouracil (5-FU) and 5-fUMP (Dubeau et al. 2008). Multiple actinobacteria have been demonstrated to exhibit resistance to 5-FC at 100 µM, including *Rhodococcus* spp., *Streptomyces* spp., and *C. sepedonicus* (Dubeau et al. 2008; Savory et al. 2020; Syverson 2011; van der Geize et al. 2008). For counter-selection, the gene cassette *codA::upp* converts 5-FC into the toxic compound 5-FU and subsequently 5-fUMP (van der Geize et al. 2008). After 7 to 14 days of growth, single colonies are visible. As a quick and easy screen, colonies from M9 5-FC plates were re-streaked on both TBY and apramycin-containing TBY plates. Recombination of the vector will inhibit growth on
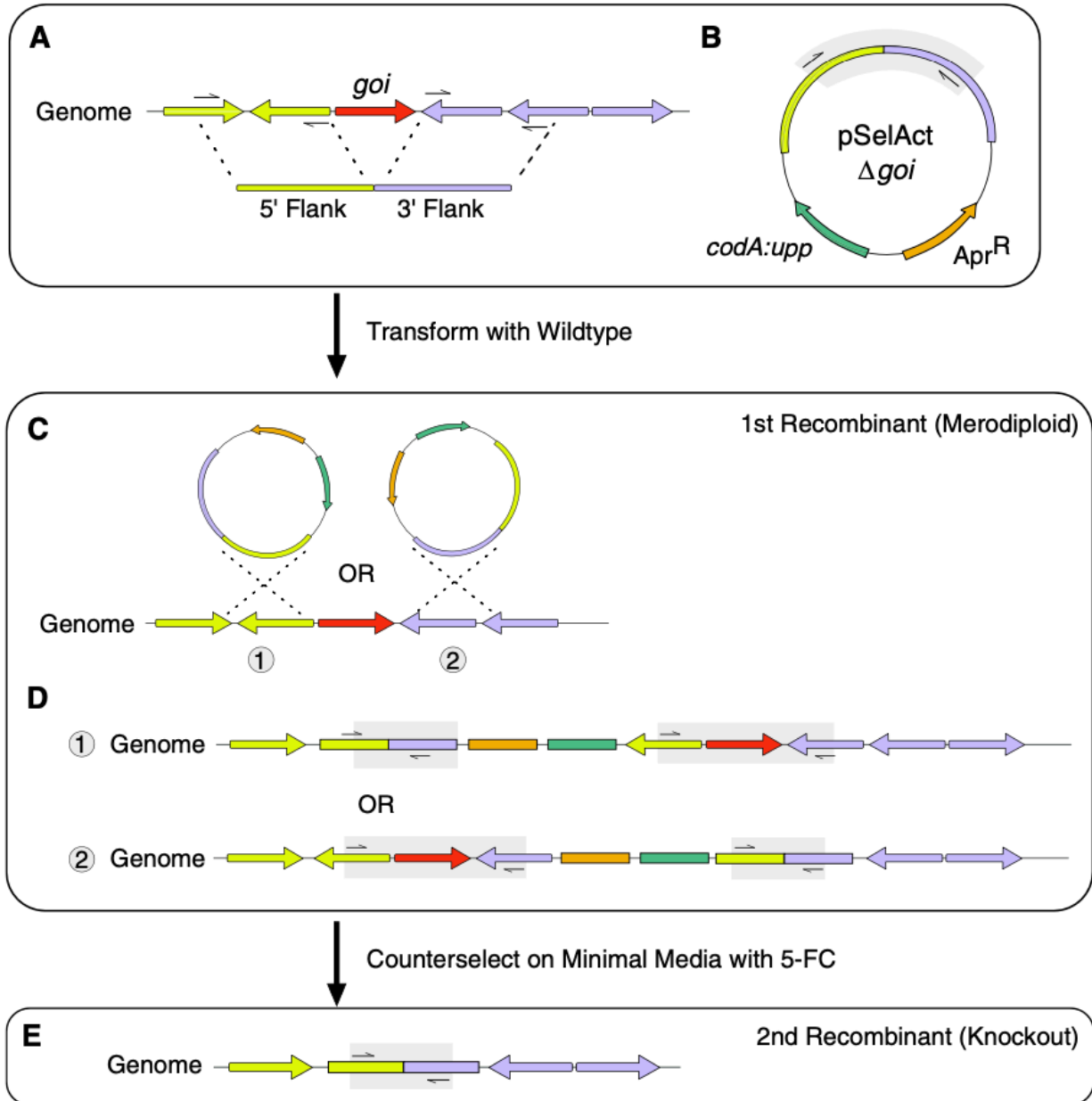
Fig. 1. Pipeline for markerless knockouts using the vector pSelAct-KO. Gray-shaded area signifies the amplified region using primers spanning the region of interest. **A,** Amplification of flanking regions before and after genes of interest. **B,** Cloning of 5′ and 3′ flanking regions into pSelAct vector through either Gateway or homology-based cloning. **C,** Possible sites of recombination during transformation and selection of nonreplicating vector into the genome. **D,** Possible gene structure of merodiploid. **E,** Gene structure after recombination where gene or gene locus of interest is deleted.

apramycin. Therefore, colonies growing on TBY but not TBY with apramycin either reverted back to wild type or are a true knockout (Fig. 1E). After incubation at 28°C on M9 5-FC plates, several hundred colonies from the *C. michiganensis* CASJ002 *chpE-ppaC* merodiploid were screened for second recombinants after transferring to apramycin-containing TBY plates. All colonies grew on both TBY and apramycin-containing TBY plates, indicating they remained merodiploids, likely showing ineffective compound uptake as no recombination events occurred.

We hypothesize that rapid growth could signal inefficient uptake of 5-FC leading to a nonselective environment. Therefore, adjustments to the incubation temperature were made. We grew the *C. michiganensis* CASJ002 *chpE-ppaC* merodiploid at 24°C on M9 5-FC plates, which results in slower colony growth. Of the 63 colonies subsequently screened on both TBY and apramycin-containing TBY plates at 28°C, three were confirmed to be true knockouts (Supplementary Fig. S2). Previous deletions using pSelAct or other similar systems tend to delete one to two genes at a time, so while we do not know the upper

limits of the system, we were able to make a deletion as large as 5,636 bp in *C. michiganensis* CASJ002 (Fig. 2) (Savory et al. 2020). These results highlight the importance of environmental conditions impacting 5-FC uptake and effectiveness as means for counter-selection.

Admittedly, converting the merodiploid to a knockout mutant can be the most laborious and unpredictable step of the system. However, there are known approaches that can enhance recombination efficiency. First, the concentration of 5-FC could be optimized. We did not choose to alter this concentration, based on its success in distantly related actinobacteria and early reports of 5-FC resistance, 5-FU susceptibility in *C. sepedonicus* (Savory et al. 2020; Syverson 2011). Second, a similar system in *Bacillus llicheniformis*, pKVM4, noted that the addition of *codB*, an encoded cytosine permease, notably increased efficiency of selection (Kostner et al. 2017). Kostner et al. (2017) hypothesized that the addition of *codB* increased 5-FC uptake, leading to an increase in the concentration of 5-FU and 5-fUMP present and, thus, greater selection. Re-engineering the pSelAct-KO to include a cytosine permease homolog may improve
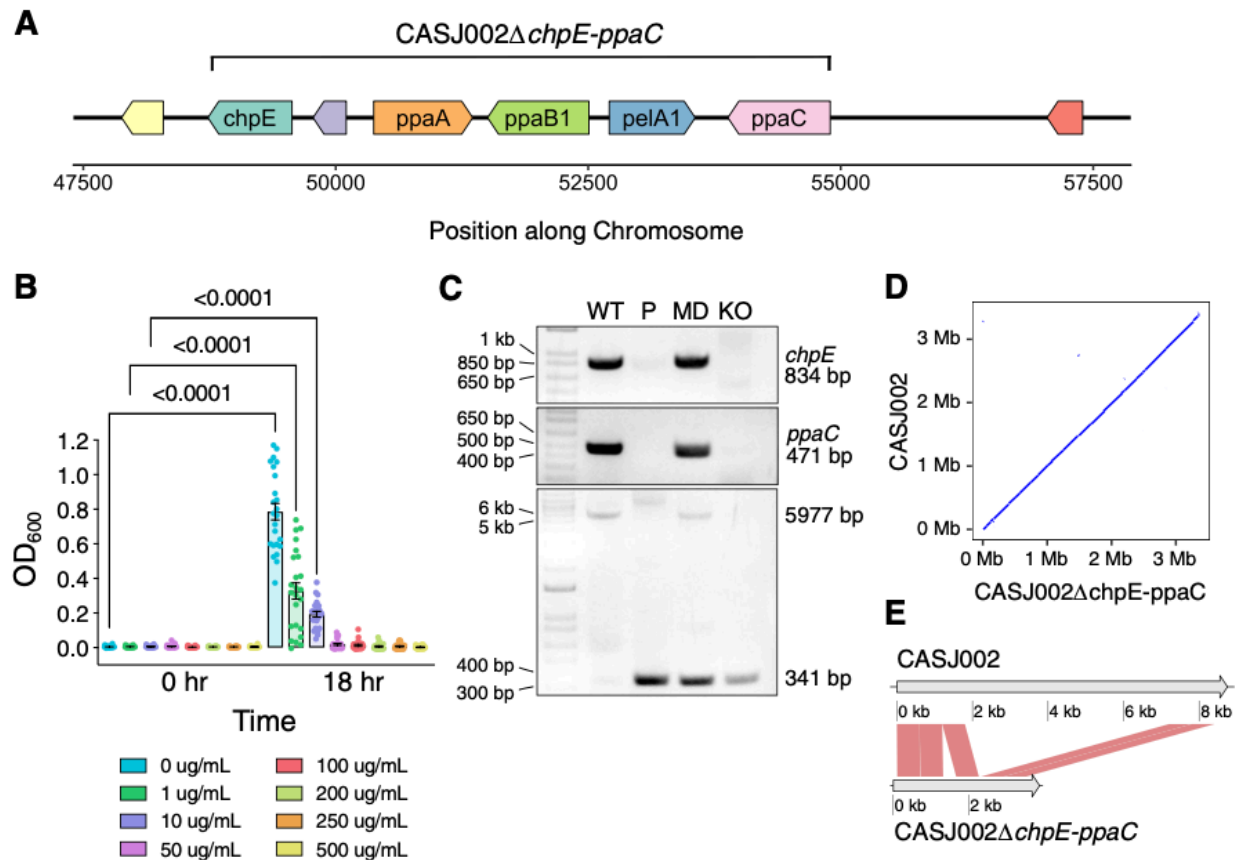

Fig. 2. Markerless deletion of *chpE* through *ppaC* in the tomato pathogen *Clavibacter michiganensis* CASJ002. **A,** Map of genes of interest for deletion and their flanking regions. **B,** Minimum inhibitory concentration of the antibiotic apramycin in tryptone broth with yeast liquid culture for *C. michiganensis* isolate CASJ002. Eight replicates per treatment were performed with media only as a negative control. The experiment was repeated independently three times. A one-way analysis of variance and a post-means Tukey's multiple comparisons test was computed. Comparisons of 0- and 18-h values from 50 μg/ml upward were not significantly different. **C,** PCR-based verification of deletion on 1% Tris-acetate-EDTA agarose gel. The top panel, gene-specific primers for *chpE*; the middle panel, gene specific primers for *ppaC*; and the bottom panel, primers that bind to the flanking regions and span the region deleted. Lanes are products from the reactions: WT = wild type genomic DNA from CASJ002 (WT), P = plasmid DNA from *Escherichia coli* carrying the pSelAct_CASJ002_chpE-ppaC_KO, MD = genomic DNA from the merodiploid, amd KO = genomic DNA from CASJ002Δ*chpE-ppaC*. The estimated sizes of the PCR products are depicted (bp = base pairs). **D,** Dot plot of CASJ002Δ*chpE-ppaC* against reference CASJ002 genome sequence. **E,** Syntenic comparison of the region of interest from CASJ002 against the engineered deletion in CASJ002Δ*chpE-ppaC*. Comparison fragment size is 600 bp. The gap in the 5′ flanking region of CASJ002Δ*chpE-ppaC* represents a gap between the scaffolded contigs.

recombination efficiency. Finally, modified environmental conditions can have dramatic effects. In the original pSelAct vector developed for *R. equi*, growing the merodiploid on rich medium over minimal medium demolished the counter-selective effects from 5-FC, likely due to a lack of 5-FC uptake related to the pyrimidine salvage pathway. While it is unclear whether decreasing the temperature had the same effect, it is clear that environmental conditions can impact recombination frequency.

Collectively, these results demonstrate that pSelAct-KO can be used in *C. michiganensis* to generate markerless deletions. Furthermore, optimizing the growth temperature when the bacteria are grown on M9 with 5-FC resulted in relatively high counter-selection. It is important to note that pSelAct-KO has been used in other gram-positive actinomycetes. Therefore, it is likely that this system will be functional in other *Clavibacter* species and related genera.

Sequence confirmation of C. michiganensis *CASJ002∆chpE-ppaC reveal no major structural changes.* The strong selection pressure from 5-FC compound conversion to 5-FU and 5-fUMP may result in plasmid loss or other structural changes. Plasmid loss, particularly pCM2, has been noted in past studies that genetically altered *C. michiganensis* (Stork et al. 2008). Of the three confirmed *C. michiganensis* CASJ002∆*chpE-ppaC*, only one lost the native plasmid pCM2 (Supplementary Fig. S2). In order to investigate for any potential large structural rearrangements, Illumina paired-end reads of *C. michiganensis* CASJ002∆*chpE-ppaC* colony 9 were assembled and compared with the wild-type genome (detailed methods below as well as in the GitHub repository pSelAct_KO_Clavibacter page). Even with low 13× Illumina sequencing coverage, most contigs were syntenic (Fig. 2D). A clean deletion in the region of interest was further confirmed between the wild type and deletion (Fig. 2E). Therefore, we recommend screening via PCR for one or more native plasmids as a background assessment and, if possible, low coverage whole-genome Illumina sequencing to compare against the wild-type genome as a precautionary measure.

Taken together, these data demonstrate the utility of using the pSelAct-KO system to generate markerless deletions in *Clavibacter* spp. This system has several advantages over antibiotic integration, including reduced potential for polar effects, and should facilitate generation of higher-order mutants. Since this system is markerless, higher-order deletions can be made, showcasing the potential strength of the system especially for characterizing redundant virulence genes or effectors.

Prediction of permissive sites via permissR.

In addition to establishing a system for markerless deletions in *Clavibacter* spp., we sought to establish an integrative plasmid for gene expression. Unlike previous integrative plasmids that were designed with a specific genome in mind, we wanted to design a system that was flexible for future genomic-driven functional studies. This required developing the necessary tools to predict potential permissive, noncoding target regions for plasmid integration.

We developed an R package, permissR, to predict permissive sites for potential plasmid integration using draft genomes. permissR minimally requires two input files, a GenBank file (.gbff or .gbk format) and a whole-genome FASTA file (Supplementary Fig. S3). We also recommend running ISEScan to avoid cloning and recombining into an insertion element that could cause unwanted structural changes. Insertion element prediction can be run within the R console or separately in the terminal and imported into permissR. Upon running permissR, either by providing paths to the required files or using the pop-up graphical use interface (GUI) to select the files, the pipeline will provide several outputs. Outputs include a whole-genome plot with information on the annotation, insertion elements (if present),

and a scanning 1,000-bp window of Shannon entropy and GC content, two different measurements of complexity (Fig. 3A) (Akhter et al. 2013). Predicted candidate permissive sites (otherwise known as intergenic regions) are at least 1.5 kb in size, are not within insertion elements, and are located on contigs a minimum 15 kb in size. While sites for homologous recombination can be smaller in size (i.e., 1 kb), since 1.5-kb regions have been reliable in terms of recombination efficiency and specificity and other studies using similar approaches selected similar sizes, we opted to maintain candidate sites be a minimum 1.5 kb in size (Dubeau et al. 2008; Monteiro et al. 2012; Savory et al. 2020). Plots of candidate sites are output such that the user can refer to the neighboring genes (Fig. 4B). Predicted sites are ranked and supplied to the user as a tab-delimited text file. Additionally, candidate sites are also compared against the whole genome and, if close hits are present (i.e., the maximum number of mismatches equal 15% or less difference based on permissive hit length), they are exported in a separate tab-delimited text file. Permissive regions predicted by permissR are not assessed or filtered in the context of lethality or having a potential defect to fitness. Therefore, users interested in permissR should test several regions for effects on bacterial fitness, especially if the genome has not been used with permissR before.

Bacterial assembly and annotation quality should be independently assessed before running permissR. Since the pipeline relies on space between coding information, permissR will not predict sites on bacterial genomes that are too fragmented. Contigs less than the 15-kb minimum size will not be assessed. While permissR itself does not assess annotation quality, results are dependent on a reliable annotation. Common measurements of annotation quality include BUSCO scores, the number of coding genes, and the number of hypothetical proteins. If the BUSCO scores are low or if the number of coding genes or hypothetical proteins becomes unreasonable compared with related taxa, then the genome should be improved before proceeding (Richardson and Watson 2013).

To test permissR, we used the *C. michiganensis* CASJ002 strain to predict potential sites. permissR predicted 10 candidate sites in the CASJ002 genome (Fig. 3A; Supplementary Table S3). Of the 10 candidates, one had a homologous region within the genome, which happened to be the second top predicted site (Fig. 3B). The GC content of sites ranged from 55 to 75%. Additional *Clavibacter* genomes were tested against the pipeline to see if sites could be detected in other species (Supplementary Table S4). The number of predicted sites differed across *Clavibacter* strains and species, including bacteria outside the *Clavibacter* genus (Supplementary Table S4). This pipeline could, in theory, be used to help other researchers design integrated plasmids for their system of interest as permissR is not genera dependent.

**Green fluorescent protein (GFP) expression
in the *Clavibacter* chromosome, using pSelAct Express.**

Next, we sought to experimentally validate two candidate sites identified by permissR. First, we generated an integrative plasmid, pSelAct-Express containing GFP. pSelAct-Express was built from the pSelAct-KO vector by replacing the *codA::upp* counter-selection cassette with the expression cassette from pKGT-GFP, which includes the phage promoter pCMP1 and enhanced GFP (eGFP) (Fig. 4A) (Chalupowicz et al. 2012; Tancos et al. 2013). We developed primers with homologous ends to amplify the pSelAct fragment and pCMP1::eGFP cassette and cloned them together, using inFusion. Similar to pSelAct-KO, to use pSelAct-Express, the Gateway cassette is replaced with an approximately 1.5-kb region amplified from within one of the predicted permissive sites to facilitate recombination.

For testing the system in *C. michiganensis* CASJ002, approximately 1.5-kb regions within two permissive regions, hit regions 1 and 3 (HR1 and HR3), were cloned into pSelAct-Express (Fig. 4B; Supplementary Table S5). Each construct was separately transformed into CASJ002 and construct integration was verified by PCR (Fig. 4B and C). Expression of eGFP was confirmed by confocal microscopy, using wild-type CASJ002 as a negative control (Fig. 4D). To verify plasmid integration did not affect disease development and bacterial titers, 3-week-old tomato cultivar M82 plants were stab-inoculated. There were no significant differences between wild type and the integrated variants when measuring the area of stem canker, a signature symptom in *C. michiganensis* disease development in tomato (Fig. 4E and F). Wilting, another common disease phenotype from *C. michiganensis*–tomato infections, was not explicitly measured; however, no drastic changes in wilting were observed. Similarly, no significant differences were detected in bacterial titer 1 cm above the site of inoculation (Fig. 4G). While we did not observe alterations in bacterial fitness or disease development upon plasmid integration in HR1 or HR3, we recommend testing several candidate regions, as there is always a potential for unpredictable effects. These results demonstrate permissR can be used to discover promising sites for plasmid integration using pSelAct-Express. We were able to identify two chromosomal regions for integrative expression in *C. michiganensis* that do not affect bacterial virulence.

Conclusion.

The tools optimized and developed in this work, pSelAct-KO, permissR, and pSelAct-Express, provide flexibility and some degree of throughput for genetic manipulation of *Clavibacter* bacteria. Molecular tools that enable the easier manipulation for genotype-phenotype studies have greatly advanced our understanding of these organisms, and we conclude that the tools
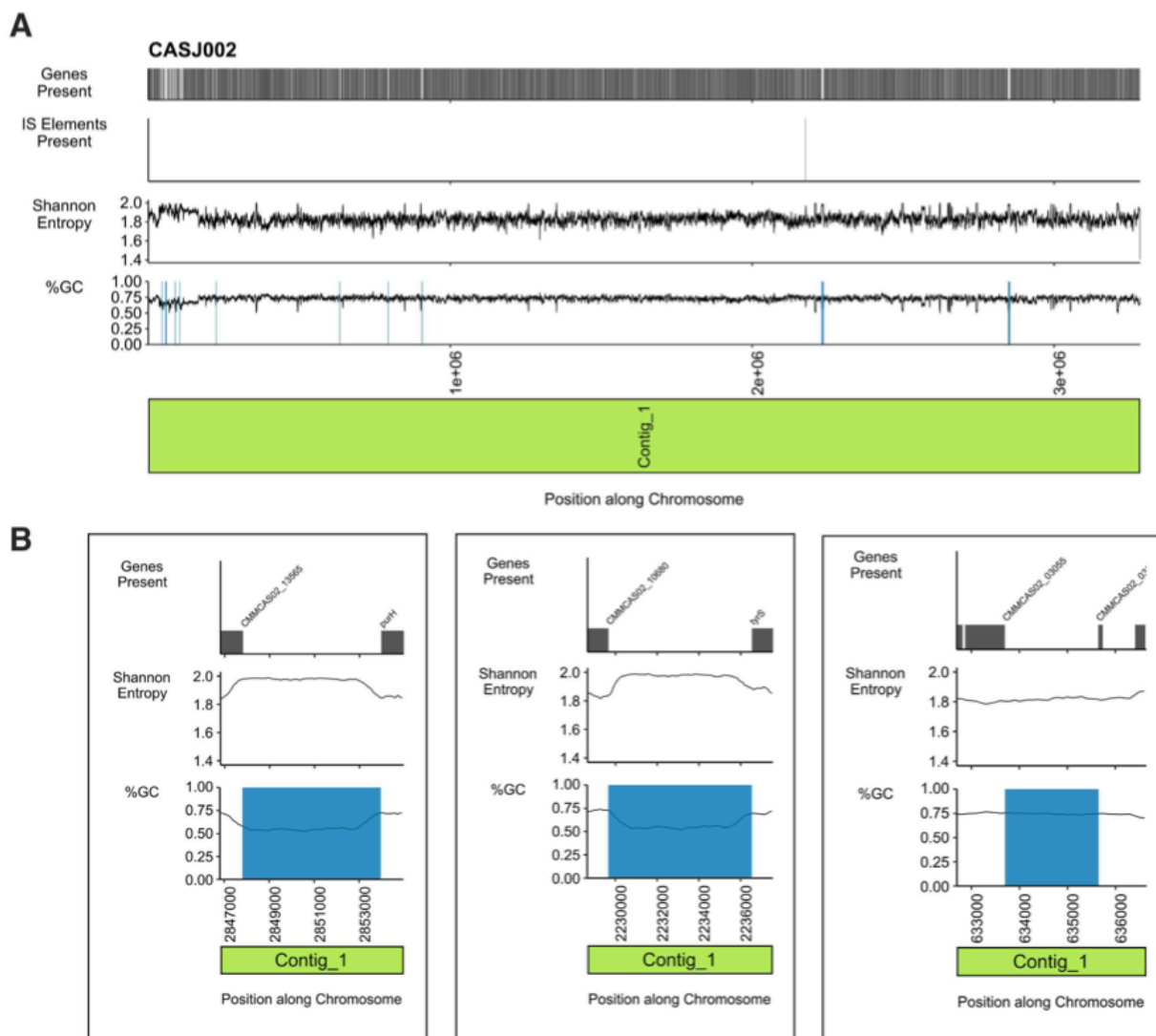

Fig. 3. The R package permissR detects permissive sites in the tomato pathogen *Clavibacter michiganensis* CASJ002. **A,** Whole-genome analyses to detect permissive sites. From top to bottom, respectively, the presence and absence of genes encoded, the functional insertion elements (IS element) predicted in the genome by ISEScan, a 1,000-bp scanning window calculation of Shannon entropy, and a 1,000-bp scanning window of GC content calculated along the genome. Candidate permissive sites highlighted in blue. **B,** Subplots of predicted permissive sites ranked by GC content and Shannon entropy. Left to right, the top two ranked sites and the last ranked site.
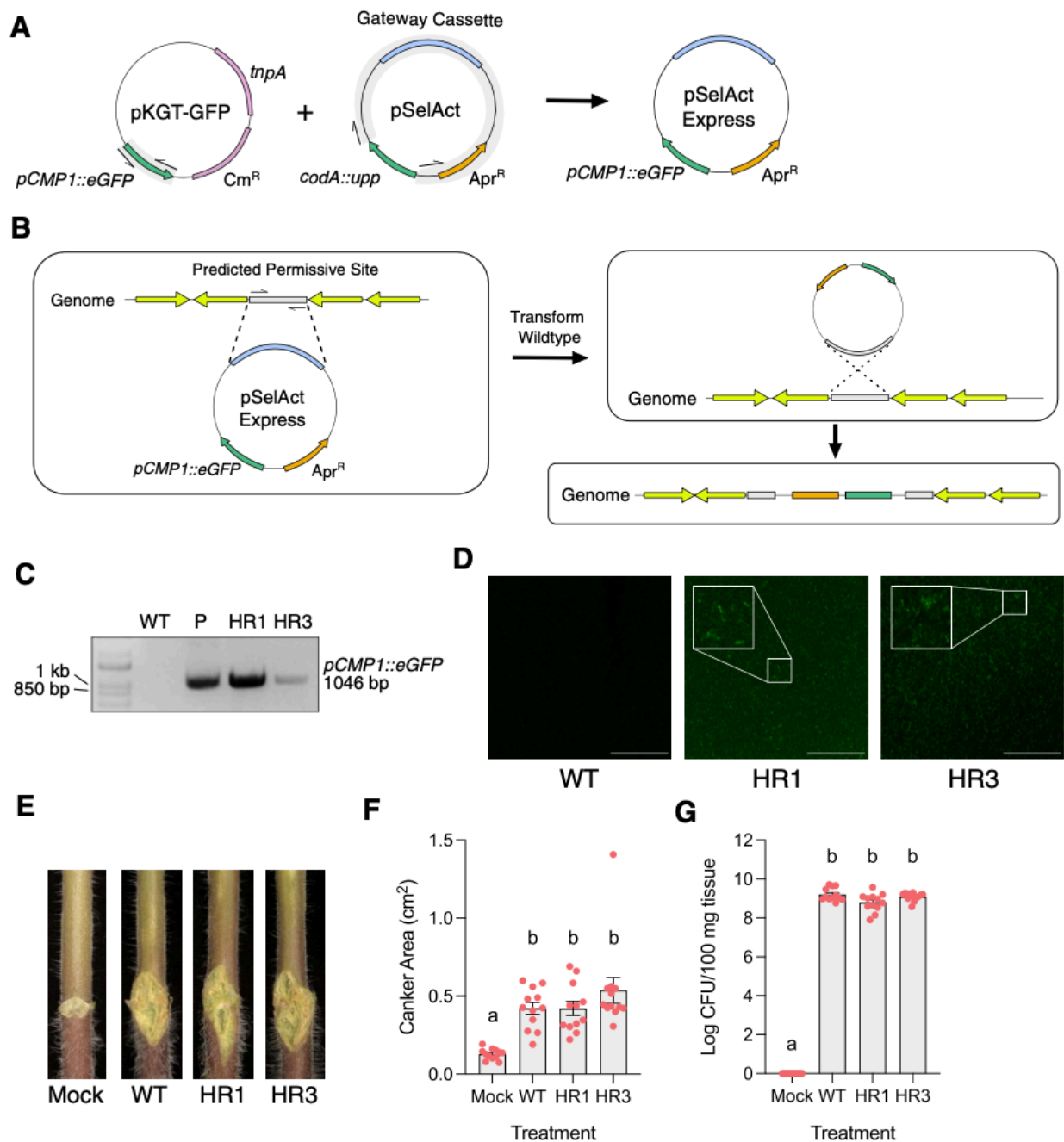
Fig. 4. Enhanced green fluorescent protein (eGFP) expression in two *Clavibacter* permissR target regions using an integrative plasmid. **A,** pSelAct Express was developed by replacing the *codA::upp* counter-selection cassette in pSelAct-KO with eGFP under a phage promoter from pKGT-GFP. **B,** On the left, amplification of a predicted permissive site from permissR into pSelAct-Express via Gateway or homology-based cloning. On the right, transformation with the modified nonreplicating vector into the wild-type (WT) strain causes recombination into the genome at the predicted permissive site. **C,** PCR-based verification of the plasmid insertion via amplification of *pCMP1::eGFP*. WT = genomic DNA from CASJ002, P = plasmid DNA from *Escherichia coli* carrying the pSelAct Express, HR1 = genomic DNA from CASJ002 integration variant at hit region 1, HR3 = genomic DNA from CASJ002 integration variant at hit region 3. **D,** Confocal micrococopy of eGFP expression in *C. michiganensis* CASJ002 (WT) and transformed integration variants (HR1 and HR3). White bar represents 40 μm in length. **E,** Representative tomato canker symptoms after inoculation with mock (water), WT CASJ002, and integrative variants HR1 and HR3. Stems were stabbed with a needle and were inoculated with 5 μl of inoculum (either water or bacteria at optical density at 600 nm = 1). Images were taken at 14 days postinoculation (dpi). Four plants per treatment per experiment were conducted with the entire experiment repeated three independent times. **F,** Quantification of canker area shown in E. Measurements were normalized with a 1-cm ruler (not shown). A one-way analysis of variance (ANOVA) and a post-means Tukey's multiple comparisons test were computed. **G,** Bacterial titers in stem tissue 14 dpi at 1 cm above the site of inoculation from tomatoes shown in E and F. A one-way ANOVA and a post-means Tukey's multiple comparisons test were computed.

presented here will open the door to investigating *Clavibacter* spp. Designed with flexibility in mind, we think there is a potential to be adapted to other orphan systems beyond bacteria in the *Clavibacter* genus.

MATERIALS AND METHODS

Plasmids, stains, and culture conditions.

Clavibacter michiganensis strains were grown in tryptone broth with yeast (TBY) (Kirchner et al. 2001). *Clavibacter* transformants were selected on TBY rich media supplemented with 200 µg of apramycin per milliliter and were screened on TBY with only 100 µg of apramycin per milliliter (Alfa Aesar, Haverhill, MA, U.S.A.). *Escherichia coli* was grown in Luria broth at 37°C. All strains were grown under standard temperatures, as listed in Supplementary Table S1, except when selecting for second recombinants. This process requires growth on a minimal media, M9, supplemented with 100 µg of of 5-FC and grown at 24°C, as listed in Supplementary Table S1 (Stork et al. 2008) (VWR, Radnor, PA, U.S.A.). A 10-mg/ml stock solution of 5-FC was prepared in distilled water and was filter-sterilized. Apramycin was prepared similarly, at a stock concentration of 100 mg/ml.

MIC of apramycin in *Clavibacter* spp.

Initial starter cultures were grown overnight at 28°C with shaking at 200 rpm in 5 ml of TBY media. Cultures were spun down at max speed for 2 min and were washed with and resuspended in sterilized water. The optical density at 600 nm (OD_{600}) was taken and was adjusted to a final concentration of 0.01 in TBY with different antibiotic concentrations. Eight 200-µl wells were used for each antibiotic concentration in a clear 96-well plate (Beckman Coulter, Pasadena, CA, U.S.A.) and were incubated at 28°C with shaking at 200 rpm. OD_{600} measurements were taken at 0 and 18 h Experiments were repeated a minimum three times.

pSelAct-KO and pSelAct-Express.

For generation of the multigene deletion plasmid, genomic DNA was extracted from 5-ml cultures of *C. michiganensis* CASJ002 grown overnight at 28°C with shaking at 200 rpm, using a similar protocol of extraction (Murray and Thompson 1980). Approximately 1.5-kb regions flanking clustered genes of interest were amplified using PCR. It is important to select the position of the knockout carefully to avoid deletion into nearby protein-coding genes to prevent unwanted polar effects. The reaction mixture for PCRs was as follows: 1× iProof GC buffer, 0.2 µM of each primer, 50 to 200 ng of genomic DNA, 2% dimethyl sulfoxide, 0.2 µM of DNTPs, and double distilled water to a final volume of 20 µl. Extension times were 15 to 30 s per kilobase of DNA amplified. Primer sequences, amplicon size, and optimized annealing temperatures for each primer pair are listed in Supplementary Table S5. The backbone of pSelAct was linearized using Phusion polymerase with GC-buffer, requiring no more than 15 ng/µl. Amplicons were gel-purified using the Zymo gel extraction kit according to the manufacturer instructions (Irvine, CA, U.S.A.). The knockout construct was developed through inFusion, using the purified amplicons, based on the manufacturer instructions, with the insert to vector ratio of 2:1 (Takara Bio, Mountain View, CA, U.S.A.) and transformed into *E. coli* DH5α. The resulting plasmid insertions were amplified and confirmed via Sanger sequencing.

For generation of pSelAct Express, pKGT-GFP plasmid DNA was extracted from *E. coli*, using a QIAprep Spin miniprep kit, according to manufacturer instructions (Qiagen, Hilden,

Germany). The phage promoter eGFP expression cassette (*pCMP1::eGFP*) was amplified via PCR (Supplementary Table S5). The pSelAct plasmid was linearized without *codA::upp*, using conditions similar to those stated above. Similar to generating the knockout construct, the plasmid was generated using inFusion, was transformed into *E. coli* DH5α, was screened via PCR, and was confirmed via Sanger Sequencing.

For cloning into hit regions from permissR predictions in *C. michiganensis* CASJ002, primers were designed from output sequence to amplify regions approximately 1.4 to 1.5 kb in size. HR1 and HR3 were amplified from genomic DNA and were cloned to replace the Gateway cassette via InFusion (Supplementary Tables S3 and S5). Positive clones were confirmed as above.

Transformation and selection.

Clavibacter competent cells were prepared and transformed as previously described (Kirchner et al. 2001). Briefly, cells were transformed with 100 to 200 ng of plasmid DNA and were plated on TBY with 200 µg of apramycin per milliliter. Colonies were picked and re-streaked onto TBY with 100 µg of apramycin per milliliter. Colonies were cultured overnight in TBY with 100 µg of apramycin per milliliter, for genomic DNA extraction (Murray and Thompson 1980).

Selection for second recombinant.

Positive merodiploids were grown overnight at their respective temperature and media. The culture was centrifuged at max speed, and the pellet resuspended in 1 ml of sterile water. Ten-fold dilutions from 10^{-1} to 10^{-7} were made and 100 µl from 10^{-6} and 10^{-7} dilutions were plated onto mM9 medium supplemented with 100 µg per milliliter of 5-FC. Plates were incubated at their indicative lower temperature, 24°C and were optimized for counter-selection (Supplementary Table S1) for 9 to 14 days. To screen quickly for recombination events, colonies were replated onto their indicated rich media with and without apramycin selection. Positive colonies, which only grew on the rich media, were additionally screened via PCR using both span and gene-specific primers, for a true knockout and not a recombinant that has returned to wild-type state.

Bioinformatic analyses.

C. michiganensis CASJ002Δ*chpE-ppaC* genomic DNA was prepared from 5-ml overnight cultures (Murray and Thompson 1980). The Microbial Genome Sequencing Center (MIGS) library prepped the DNA for paired-end read sequencing via an Illumina platform. Raw paired-end reads were checked for quality and any contamination using FastQC (v0.11.9) and the sendsketch.sh script from bbtools, respectively. Since contamination was present, genomes closest to the top one to two hits from the sendsketch.sh output were downloaded using the bioinf_tools package (v1.8.17) (Lee 2018). Reads were then mapped to the genome of the *Clavibacter* reference and contaminants and were binned to their respective hit using bbtools' bbsplit.sh script. Successful binning was confirmed using sourmash (v3.5.0) based on a kmer size of 31 (Brown and Irber 2016). Merged reads were reformed via reformate.sh from bbtools and were trimmed using trimmomatic (v0.39). Trimmed reads were de novo assembled via SPAdes (v3.14.1) with default parameters for short-read assembly (Bankevich et al. 2012). Contigs were scaffolded using MeDuSa based on the reference wild-type genome and were oriented based on the reference, using CONTIGuator2 (Bosi et al. 2015; Galardini et al. 2011). The scaffolded contigs and the region of interest were mapped to their associate reference via Minimap2 (v2.17-r941) and FastANI (v1.32.0), and structural changes were assessed using a custom R script that uses the R package pafr (0.0.2) and genoplotR (0.8.11) (Charif and Lobry 2007; Jain et al. 2018;

36

Li 2018). Detailed methods and scripts can be found in the GitHub repository.

The R package permissR.

permissR was written in the R language (R Core Team 2019) and requires only two input files, at minimum, a GenBank file (gbk or gbff file formate) and a FASTA file. While the package does not require any external programs to run, it is recommended running and including the output from ISEScan, an annotation-independent software to find insertion elements in bacterial genomes, to avoid the potential of cloning into a functional insertion element that may cause unwanted gene movement (Xie and Tang 2017). To run ISEScan, we recommend using the package management software bioconda.

Detailed installation and usage can be found in the permissR GitHub repository. Briefly, installation of the package requires installing the *devtools* R package. To run, calling the function permissR in the package on the R console, the user can provide the path (relative or absolute) for the GenBank file and then FASTA file of interest or leave the function empty, which will cause a GUI to appear to allow the user to select each file. The function will ask the name of the strain, which is used to label the output files, as well as if there are any outputs for ISEScan.

Plant growth and pathogenicity assays.

The tomato cultivar M82 (*Solanum lycopersicum* cv. M82) was used for all assays. Tomato plants were grown in a growth chamber under 16-h light and 8-h dark conditions with 50% humidity. Plants were grown for 3 weeks, by which time at least two true leaves had fully emerged, were pricked between the two cotyledons on the stem using a sterilized needle, and 5 µl of bacterial inoculum or water (mock) was dropped into the wound. The inoculum was prepared by initially streaking out *Clavibacter* stocks 4 days before inoculation on either TBY or TBY plates with apramycin. The day of inoculation, bacteria were suspended from the plate in sterile water and the OD_{600} was adjusted to 1. Symptoms were photographed at 14 days postinoculation (dpi) and the canker was measured, using ImageJ, based on a 1-cm ruler to normalize.

To determine bacterial titers in the tomato stem, stem segments were cut surrounding the inoculation site at 14 dpi. Stem pieces were surface-sterilized in 70% ethanol for 10 to 15 s, followed by suspension in sterile water for 10 to 15 s. Sections were taken 1 cm above the inoculation site. Tissue was weighed at 100 mg, was suspended in sterile water, and was ground. Sterile dilutions of the ground tissue were plated on D2 medium supplemented with 20 mg of cycloheximide and were incubated for 5 to 6 days at 28°C (Thapa et al. 2017).

Confocal microscopy.

Before imaging, *Clavibacter* stocks were streaked onto either TBY or TBY plates with apramycin and were grown at 28°C. Bacteria was smeared onto glass slides, 2 µl of sterile water was pipetted on, and glass cover slips were added on top. Confocal images were taken with a Leica SP8 TCS with a 63× oil objective, laser power set to 6%, and excitation and emission wavelengths set to 488 and 509 nm, respectively. Fiji with the Bio-Formats Importer was used to adjust the brightness and contrast of the images.

Statistical analyses.

All raw data were plotted and error bars in plots represent standard error of mean. Data across independent replicates was colligated and a one-way analyis of variance and a post-means Tukey's multiple comparisons test was computed, using Graphpad Prism 9 software. Significant differences between groups include P values ≤ 0.001. Outputs of statistical analyses from Figures 2B and 4F and 4G can be found in Supplementary Tables S6, S7, and S8, respectively.

Data availability.

Paired-end Illumina reads of *C. michiganensis* CASJ002Δ*chpE-ppaC* were submitted to figshare and the assembled genome is available in the GitHub repository, along with the scripts necessary for analyzing the data. The R package permissR can also be found in the GitHub repository, including installation and usage instructions.

AUTHOR-RECOMMENDED INTERNET RESOURCES

figshare Illumina reads:
    https://figshare.com/articles/dataset/CASJ002_chpE-ppaC_Paired_End_Illumina_Raw_Reads/14810322
GitHub repository, assembled *Clavibacter* genome:
   https://github.com/DanielleMStevens/pSelAct_KO_Clavibacter
GitHub repository, permissR:
    https://github.com/DanielleMStevens/permissR

LITERATURE CITED

Ahmad, A., Mbofung, G. Y., Acharya, J., Schmidt, C. L., and Robertson, A. E. 2015. Characterization and comparison of *Clavibacter michiganensis* subsp. *nebraskensis* strains recovered from epiphytic and symptomatic infections of maize in Iowa. PLoS One 10:e0143553.

Akhter, S., Bailey, B. A., Salamon, P., Aziz, R. K., and Edwards, R. A. 2013. Applying Shannon's information theory to bacterial and phage genomes and metagenomes. Sci. Rep. 3:1033.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455-477.

Bentley, S. D., Corton, C., Brown, S. E., Barron, A., Clark, L., Doggett, J., Harris, B., Ormond, D., Quail, M. A., May, G., Francis, D., Knudson, D., Parkhill, J., and Ishimaru, C. A. 2008. Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. J. Bacteriol. 190:2150-2160.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R., and Fondi, M. 2015. MeDuSa: A multi-draft based scaffolder. Bioinformatics 31:2443-2451.

Brown, C. T., and Irber, L. 2016. sourmash: A library for MinHash sketching of DNA. J. Open Source Softw. 1:27.

Chalupowicz, L., Barash, I., Reuven, M., Dror, O., Sharabani, G., Gartemann, K., Eichenlaub, R., Sessa, G., and Manulis-Sasson, S. 2017. Differential contribution of *Clavibacter michiganensis* ssp. *michiganensis* virulence factors to systemic and local infection in tomato. Mol. Plant Pathol. 18:336-346.

Chalupowicz, L., Zellermann, E.-M., Fluegel, M., Dror, O., Eichenlaub, R., Gartemann, K.-H., Savidor, A., Sessa, G., Iraki, N., Barash, I., and Manulis-Sasson, S. 2012. Colonization and movement of GFP-labeled *Clavibacter michiganensis* subsp. *michiganensis* during tomato infection. Phytopathology 102:23-31.

Charif, D., and Lobry, J. R. 2007. SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. Pages 207-232 in: Structural Approaches to Sequence Evolution. U. Bastiola, M. Porto, E. Roman, and M. Vendruscolo, eds. Springer-Verlag, Berlin.

Dalio, R. J. D., Herlihy, J., Oliveira, T. S., McDowell, J. M., and Machado, M. 2018. Effector biology in focus: A primer for computational prediction and functional characterization. Mol. Plant-Microbe Interact. 31:22-33.

Dubeau, M.-P., Ghinet, M. G., Jacques, P.-É., Clermont, N., Beaulieu, C., and Brzezinski, R. 2008. Cytosine deaminase as a negative selection

marker for gene disruption and replacement in the genus *Streptomyces* and other actinobacteria. Appl. Environ. Microbiol. 75:1211-1214.

Eichenlaub, R., and Gartemann, K.-H. 2011. The *Clavibacter michiganensis* subspecies: Molecular investigation of gram-positive bacterial plant pathogens. Annu. Rev. Phytopathol. 49:445-464.

Galardini, M., Biondi, E. G., Bazzicalupo, M., and Mengoni, A. 2011. CONTIGuator: A bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol. Med. 6:11.

Gartemann, K.-H., and Eichenlaub, R. 2001. Isolation and characterization of IS1409, an insertion element of 4-chlorobenzoate-degrading *Arthrobacter* sp. strain TM1, and development of a system for transposon mutagenesis. J. Bacteriol. 183:3729-3736.

Hwang, I. S., Oh, E., Kim, D., and Oh, C. 2018. Multiple plasmid-borne virulence genes of *Clavibacter michiganensis* ssp. *capsici* critical for disease development in pepper. New Phytol. 217:1177-1189.

Jahr, H., Dreier, J., Meletzus, D., Bahro, R., and Eichenlaub, R. 2000. The endo-β-1,4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. Mol. Plant-Microbe Interact.. 13:703-714.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9:5114.

Kirchner, O., Gartemann, K.-H., Zellermann, E.-M., Eichenlaub, R., and Burger, A. 2001. A highly efficient transposon mutagenesis system for the tomato pathogen *Clavibacter michiganensis* subsp. *michiganensis*. Mol. Plant-Microbe Interact.. 14:1312-1318.

Kostner, D., Rachinger, M., Liebl, W., and Ehrenreich, A. 2017. Markerless deletion of putative alanine dehydrogenase genes in *Bacillus licheniformis* using a *codBA*-based counterselection technique. Microbiology 163:1532-1539.

Laine, M. J., Nakhei, H., Dreier, J., Lehtilä, K., Meletzus, D., Eichenlaub, R., and Metzler, M. C. 1996. Stable transformation of the gram-positive phytopathogenic bacterium *Clavibacter michiganensis* subsp. *sepedonicus* with several cloning vectors. Appl. Environ. Microbiol. 62:1500-1506.

Lee, M. D. 2018. Bioinformatics tools (bit). Zenodo. Geneva, Switzerland. Published online.

Lee, Y.-C., Chien, C.-F., and Lin, N.-C. 2018. Knock-out or knock-in? Converting a SacB-based gene disruption system for site-specific chromosomal integration in *Pseudomonas syringae* pv. *tomato* DC3000. J. Microbiol. Methods 145:50-58.

Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.

Lu, Y., Ishimaru, C. A., Glazebrook, J., and Samac, D. A. 2018. Comparative genomic analyses of *Clavibacter michiganensis* subsp. *insidiosus* and pathogenicity on *Medicago truncatula*. Phytopathology 108:172-185.

Meletzus, D., Bermphol, A., Dreier, J., and Eichenlaub, R. 1993. Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. J. Bacteriol. 175:2131-2136.

Meletzus, D., and Eichenlaub, R. 1991. Transformation of the phytopathogenic bacterium *Clavibacter michiganense* subsp. *michiganense* by electroporation and development of a cloning vector. J. Bacteriol. 173: 184-190.

Monteiro, F., Solé, M., van Dijk, I., and Valls, M. 2012. A chromosomal insertion toolbox for promoter probing, mutant complementation, and pathogenicity studies in *Ralstonia solanacearum*. Mol. Plant-Microbe Interact. 25:557-568.

Mullens, A., and Jamann, T. 2021. Colonization and movement of green fluorescent protein-labeled *Clavibacter nebraskensis* in maize. Plant Dis. 105:1422-1431.

Murray, M. G., and Thompson, W. F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8:4321-4326.

Nandi, M., Macdonald, J., Liu, P., Weselowski, B., and Yuan, Z. 2018. *Clavibacter michiganensis* ssp. *michiganensis*: Bacterial canker of tomato, molecular interactions and disease management. Mol. Plant Pathol. 19:2036-2050.

Peritore-Galve, F. C., Tancos, M. A., and Smart, C. D. 2021. Bacterial canker of tomato: Revisiting a global and economically damaging seedborne pathogen. Plant Dis. Published online.

Ponciano, G., Ishihara, H., Tsuyumu, S., and Leach, J. E. 2003. Bacterial effectors in plant disease and defense: Keys to durable resistance? Plant Dis. 87:1272-1282.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, E. J., and Watson, M. 2013. The automatic annotation of bacterial genomes. Brief. Bioinform. 14:1-12.

Savory, E. A., Weisberg, A. J., Stevens, D. M., Creason, A. L., Fuller, S. L., Pearce, E. M., and Chang, J. H. 2020. Phytopathogenic *Rhodococcus* have diverse plasmids with few conserved virulence functions. Front. Microbiol. 11:1022.

Stork, I., Gartemann, K., Burger, A., and Eichenlaub, R. 2008. A family of serine proteases of *Clavibacter michiganensis* subsp. *michiganensis*: chpC plays a role in colonization of the host plant tomato. Mol. Plant Pathol. 9:599-608.

Syverson, R. L. 2011. Multiple approaches towards understanding virulence in *Clavibacter michiganensis* subsp. *sepedonicus*, causal agent of bacterial ring rot of potato. Ph.D. dissertation. University of Minnesota, St. Paul, MN, U.S.A.

Tambong, J. T. 2017. Comparative genomics of *Clavibacter michiganensis* subspecies, pathogens of important agricultural crops. PLoS One 12:e0172295.

Tancos, M. A., Chalupowicz, L., Barash, I., Manulis-Sasson, S., and Smart, C. D. 2013. Tomato fruit and seed colonization by *Clavibacter michiganensis* subsp. *michiganensis* through external and internal routes. Appl. Environ. Microbiol. 79:6948-6957.

Thapa, S. P., Davis, E. W., II, Lyu, Q., Weisberg, A. J., Stevens, D. M., Clarke, C. R., Coaker, G., and Chang, J. H. 2019. The evolution, ecology, and mechanisms of infection by gram-positive, plant-associated bacteria. Annu. Rev. Phytopathol. 57:341-365.

Thapa, S. P., Pattathil, S., Hahn, M., Jacques, M.-A., Gilbertson, B., and Coaker, G. 2017. Genomic analysis of *Clavibacter michiganensis* reveals insight into virulence strategies and genetic diversity of a gram-positive bacterial pathogen. Mol. Plant-Microbe Interact. 30:786-802.

van der Geize, R., de Jong, W., Hessels, G. I., Grommen, A. W. F., Jacobs, A. A. C., and Dijkhuizen, L. 2008. A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. Nucleic Acids Res. 36:e151.

Xie, Z., and Tang, H. 2017. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 33: 3340-3347.

Cm CASJ002

Concentration Apramycin
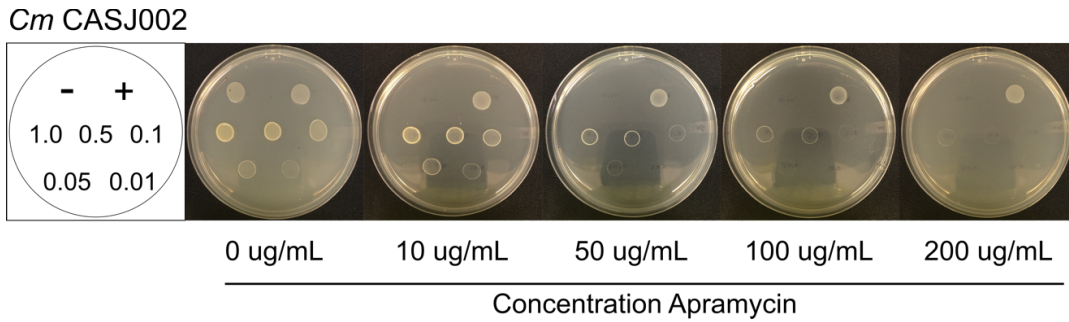
Supplemental Figure 1: Concentration of apramycin required to restrict *C. michiganensis* on agar plates. Left: Schematic for plating. The negative sign represents the negative *E. coli* control. The positive sign represents the positive control, the pSelAct KO vector in *E. coli.* Both controls are tested at an OD_{600} of 1. The numbers 1.0 through 0.01 are the tested optical densities (600 nm) of *C. michiganensis* isolate CASJ002. Right: Apramycin concentrations in TBY media.
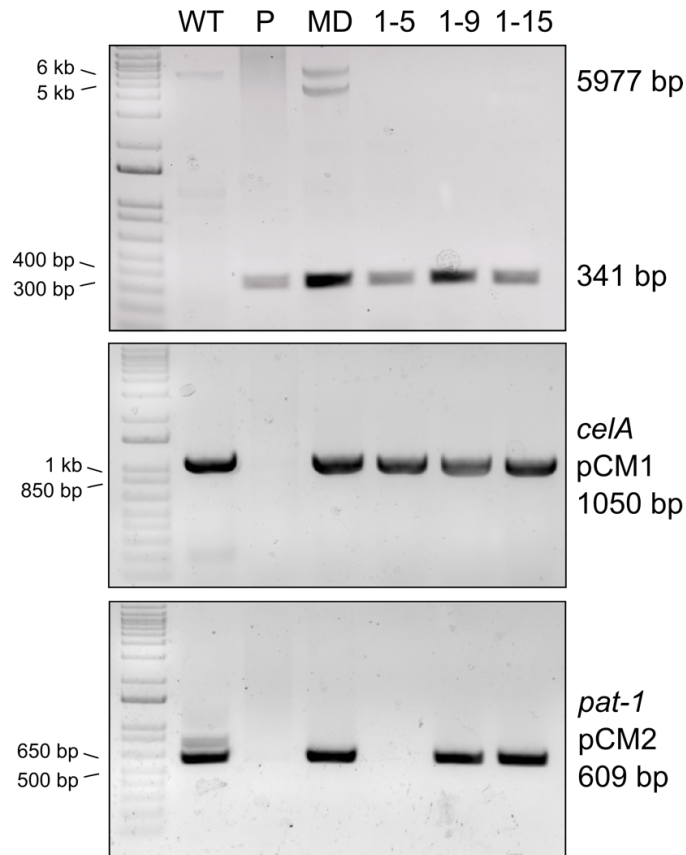
Supplemental Figure 2: Counterselection can lead to loss of native plasmids.
PCR-based verification of native plasmids pCM1 and pCM2 on a 1% TAE agarose gel.
Top: Primers which bind to the flanking regions and span the deleted region. Middle:
Gene specific primers for *celA* (indicates presence of pCM1). Bottom: Gene specific
primers for *pat-1* (indicates presence of pCM2). Lanes are products from the reactions
including: WT; Genomic DNA from CASJ002, P: Plasmid DNA from *E. coli* carrying the
pSelAct_CASJ002_chpE-ppaC_KO; MD: Genomic DNA from the merodiploid; 1-5, 1-9,
and 1-15: Genomic DNA from three CASJ002Δ*chpE-ppaC* recombinant colonies. The
estimated sizes of the PCR products are depicted.
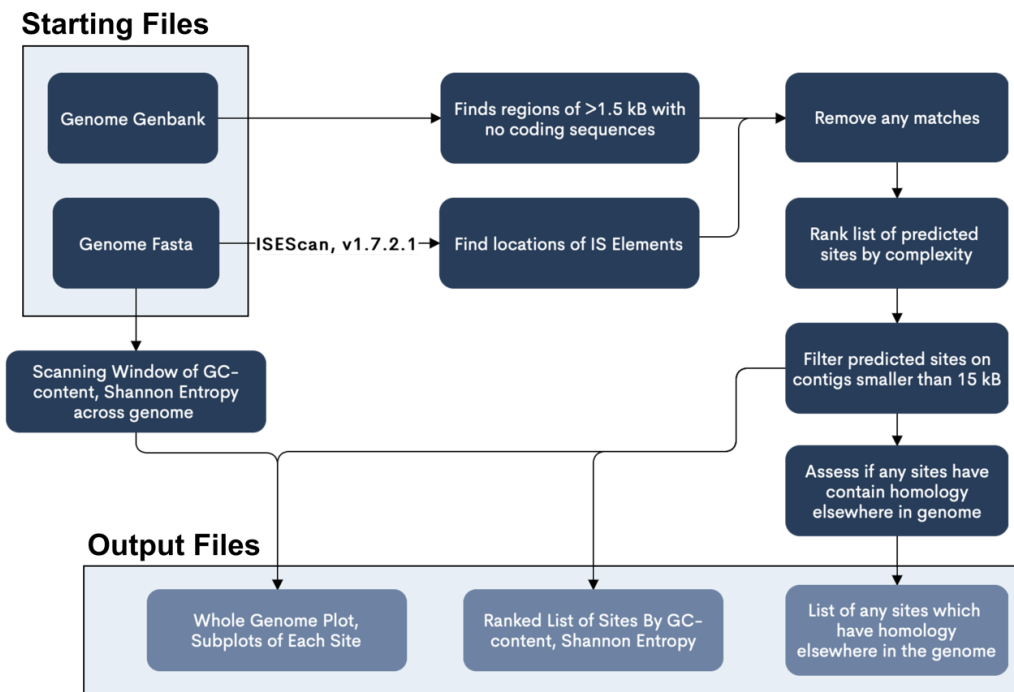
Supplemental Figure 3: Pipeline for the R package permissR to predict permissive sites in bacterial genomes. The pipeline uses whole genome Genbank and Fasta files to identify sites which are at least 1.5 kB in length (for specific recombination), that contain no coding elements, no mobile elements (which may trigger structural changes), and are ranked by complexity to make cloning reasonable (i.e. too high GC-content).

CHAPTER 2

Natural variation of immune epitopes reveals intrabacterial antagonism

Danielle M. Stevens, Alba Moreno-Pérez, Alexandra J. Weisberg, Charis Ramsing, Judith Fliegmann, Ning Zhang, Melanie Madrigal, Gregory Martin, Adam Steinbrenner, Georg Felix, and Gitta Coaker

Abstract

Plants and animals detect biomolecules termed Microbe-Associated Molecular Patterns (MAMPs) and induce immunity. Agricultural production is severely impacted by pathogens which can be controlled by transferring immune receptors. However, the vast majority of studies use a single MAMP epitope and the impact of diverse multi-copy MAMPs on immune induction is unknown. Here we characterized the epitope landscape from five proteinaceous MAMPs across 4,228 plant-associated bacterial genomes. Despite the diversity sampled, natural variation was constrained and experimentally testable. Immune perception in both Arabidopsis and tomato depended on both epitope sequence and copy number variation. For example, Elongation Factor Tu is predominantly single copy and 92% of its epitopes are immunogenic. Conversely, 99.9% of bacterial genomes contain multiple Cold Shock Proteins and 46% carry a non-immunogenic form. We uncovered a new mechanism for immune evasion, intrabacterial antagonism, where a non-immunogenic Cold Shock Protein blocks perception of immunogenic forms encoded in the same genome. These data will lay the foundation for immune receptor deployment and engineering based on natural variation.

Author Contributions

I wrote the initial drafts of the manuscript excluding some methodology and addressed subsequent edits as suggested by my co-authors. I developed all figures in the manuscript except Fig. S5A,C and Fig. S9A. Co-authors AMP, CR, JF, and NZ collected the underlying data for Figures 2B-D, Figures 4A-B, Fig. S4B-E, Fig. S5D, Fig. S7, and Fig. S12B. I collected all other data in the manuscript and developed all other computational pipelines and code.

Natural variation of immune epitopes reveals intrabacterial antagonism

Danielle M. Stevens[a,b], Alba Moreno-Pérez[b], Alexandra J. Weisberg[c], Charis Ramsing[b], Judith Fliegmann[d], Ning Zhang[e,f], Melanie Madrigal[b], Gregory Martin[e,f], Adam Steinbrenner[g], Georg Felix[d], and Gitta Coaker[b,1]

Plants and animals detect biomolecules termed microbe-associated molecular patterns (MAMPs) and induce immunity. Agricultural production is severely impacted by pathogens which can be controlled by transferring immune receptors. However, most studies use a single MAMP epitope and the impact of diverse multicopy MAMPs on immune induction is unknown. Here, we characterized the epitope landscape from five proteinaceous MAMPs across 4,228 plant-associated bacterial genomes. Despite the diversity sampled, natural variation was constrained and experimentally testable. Immune perception in both *Arabidopsis* and tomato depended on both epitope sequence and copy number variation. For example, Elongation Factor Tu is predominantly single copy, and 92% of its epitopes are immunogenic. Conversely, 99.9% of bacterial genomes contain multiple cold shock proteins, and 46% carry a nonimmunogenic form. We uncovered a mechanism for immune evasion, intrabacterial antagonism, where a nonimmunogenic cold shock protein blocks perception of immunogenic forms encoded in the same genome. These data will lay the foundation for immune receptor deployment and engineering based on natural variation.

plant innate immunity | natural variation | pattern recognition receptor | microbe-associated molecular patterns

Plants contain hundreds of innate immune receptors capable of recognizing diverse pathogens. Biotic organisms carry microbe-associated molecular patterns (MAMPs), fragments of larger biomolecules such as proteins, carbohydrates, and lipids, that are recognized by surface-localized pattern recognition receptors (PRRs) (1–3). PRRs include receptor-like kinases (RLKs) and receptor-like proteins (RLPs) that recognize conserved MAMPs or damage-associated molecular patterns, resulting in PRR-triggered immunity (PTI) (3, 4). Activation of PTI induces multiple defense responses including the production of reactive oxygen species (ROS) and ethylene, apoplast alkalization, activation of mitogen-activated protein kinase (MAPK) cascades, transcriptional reprogramming, and callose deposition at the cell wall culminating in disease resistance (4, 5).

The most well-studied PRR is FLAGELLIN-SENSING 2 (FLS2), which recognizes the 22-amino-acid epitope, flg22, from bacterial flagellin FliC (6–9). Unlike most PRRs, *FLS2* is conserved throughout the plant kingdom (7, 9, 10). Since the flg22-FLS2 discovery, researchers have uncovered other epitope-receptor pairs restricted to certain plant families. For instance, the Elongation Factor Tu (EF-Tu)-derived epitope elf18 interacts with the EFR receptor, which is present in the Brassicaceae family (11, 12). The 22-amino-acid epitope of cold shock protein (CSP), csp22, interacts with CORE, and a second flagellin epitope flgII-28 interacts with FLAGELLIN-SENSING 3 (FLS3), both solanaceous RLKs (13–16). Finally, the necrosis-and-ethylene inducing peptide 1 (Nep1)-like protein (NLP) epitope nlp20 interacts with RLP23 found in Brassicaceae (17–19). These receptors recognize certain bacterial MAMPs, including those present in pathogens and commensals.

Proteins that carry MAMP sequences are thought to be important for microbial survival and fitness. For example, flagellin enables bacterial swimming and swarming and is critical for the colonization of certain hosts (20). Different regions of flagellin monomer, FliC, are recognized by different receptors in plants and mammals despite the flagellin apparatus requiring multiple components (21). CSPs are an ancient protein family first described in *Escherichia coli* with roles in RNA chaperoning in cold environments (22). This can include unwinding RNA, maintaining RNA stability, limiting internal cleavage, and enhancing expression via antitermination. Recent genetic work on other bacterial systems has shown additional posttranslational regulatory roles related to virulence, stress tolerance, pili formation, and biofilm development (23–28). CSPs are composed of two critical motifs, RNP-1 and RNP-2, near the N terminus of the small beta-barrel like structure

Significance

Plants recognize pathogens as non-self using innate immune receptors. Receptors on the cell surface can recognize amino acid epitopes present in pathogen proteins. Despite many papers investigating receptor signaling, the vast majority use a single epitope. Here, we analyzed the natural variation across five different epitopes and experimentally characterized their perception in plants. We highlight the importance of analyzing all epitope copies within a pathogen genome. Through genetic and biochemical analyses, we revealed a mechanism for immune evasion, intrabacterial antagonism, where a nonimmunogenic epitope blocks perception of immunogenic forms encoded in a single genome. These data can directly inform disease control strategies by enabling prediction of receptor utility and deployment for current and emerging pathogens.

that binds broadly to nucleic acids (22). While the regulatory function of CSPs appears conserved across bacteria, their targets and sequences are highly variable. NLPs are cross-kingdom phytotoxic virulence factors (17, 19). Flagellin, CSPs, and NLPs are considered expendable, although their loss is predicted to confer fitness costs. Conversely, EF-Tu is a highly abundant, essential protein that transports aminoacyl-tRNAs to the ribosome during translation (29). MAMP-encoded gene function and abundance likely have an impact on their evolution in the context of plant immune interactions.

Using flg22 perception as a model, three general epitope outcomes have been described: immune activation, evasion, and antagonism. Many gram-negative plant pathogens carry immunogenic epitopes, though some can evade perception (7, 30–32). MAMP evasion, also known as masking, occurs by accumulating sequence variation that prevents epitope binding (33, 34). MAMPs can act antagonistically and block subsequent perception of immunogenic epitopes from other bacteria by binding to the primary receptor and inhibiting proper signaling complex formation (34, 35). Sequence variation within the perceived epitope is thought to be constrained by protein functionality, though this has only been studied in the flg22 epitope and FliC protein (8). How polymorphism affects protein function versus immune perception has not been investigated for other MAMP-encoded proteins. In addition, how MAMP copy number variation (CNV) encoded within a single genome impacts immune outcomes remains elusive. Therefore, many questions remain regarding how natural epitope variation interplays with protein function and plant immune perception.

To broadly understand how natural evolution impacts immune outcomes across five different MAMPs, we mined 34,262 MAMP epitopes from 4,228 whole bacterial genomes. Each MAMP displayed substantial copy number and sequence variation. While the theoretical number of MAMP variants is astronomically large, natural variation is constricted, making it experimentally testable to characterize MAMP evolution. We focused on characterizing the immunogenic outcomes of the EF-Tu (elf18) and CSP (csp22) MAMPs for sequence and CNV. Elf18 displays minimal sequence variation and gene expansion, with most variants inducing strong immune responses. In contrast, csp22 displays considerable variation in epitope sequence, CNV, and immune outcomes. Using a combination of phylogenetics, genetics, and biochemistry, we revealed conserved nonimmunogenic CSPs in a subset of bacterial genera, some of which antagonize perception of immunogenic forms encoded in the same genome. We then characterized an actinobacterial CSP that acts as an intrabacterial antagonist of the CORE receptor, which enables immune evasion.

Results

Evolutionary Trajectories Depend on the Immunogenic Feature and Bacterial Genera.
Here, we sought to assess the presence and diversity of bacterial proteinaceous MAMPs. Across 13 genera, we identified plant-associated bacterial genomes with representing members from alpha-, beta-, and gamma-proteobacteria as well as gram-positive actinobacteria (Dataset S1). We extracted features from the following proteins and their corresponding MAMP epitopes: bacterial flagellin (flg22 and flgII-28), cold-shock protein (csp22), EF-Tu (elf18), and Nep1-like protein (nlp20). A computational pipeline was developed that extracts peptides using a modified BlastP protocol and local protein alignment, polymorphic ends and off-target correction, and clonality filtering (Fig. 1A). Mining bacterial epitopes based on gene annotation and local alignment is a convenient approach; however, genome annotations may not be consistent or completely accurate. Conversely, BlastP was not always consistent at detecting polymorphic epitopes. Therefore, we combined both approaches to identify MAMPs including those in genes with unique domain architecture. This enabled mining of epitopes in a gene description-dependent and -independent manner allowing for comprehensive genome-derived epitope comparisons (SI Appendix, Table S1).

Differences in sequence and CNV were detected in a MAMP-dependent manner. From the 4,228 genomes, 34,262 epitopes were extracted, and their abundance was plotted on a 74 gene maximum-likelihood tree (Fig. 1B). Notably, there are different patterns of epitope CNV in a lineage-specific manner (Fig. 1B and SI Appendix, Fig. S1A). The MAMPs elf18, flg22, flgII-28, and nlp20 are primarily encoded by single-copy genes, whereas csp22 variants displayed expansion and are encoded by 1 to 15 genes (Fig. 1C and SI Appendix, Fig. S1A). Next, we investigated CNV across 13 genera (SI Appendix, Fig. S1A). Strains had an average of four CSPs, with variation in a genera and species-dependent manner. For example, CSP CNV was increased in Agrobacterium and Streptomyces (average of eight and seven paralogs, respectively) compared to Dickeya, Pectobacterium, Xanthomonas, Ralstonia, Curtobacterium, Rathayibacter, and Clavibacter (average of three to five copies) (SI Appendix, Fig. S1A). Although fliC is primarily a single-copy gene in 80% (2,921/3,631) of the genomes, additional copies can be observed in Erwinia, Pectobacterium, and Agrobacterium (SI Appendix, Fig. S1A). Similarly, EF-Tu is primarily a single-copy gene in 84% (2,825/3,346) of analyzed genomes, with additional copies predominantly found in gram-negative bacteria in addition to Streptomyces (Fig. 1 B and C and SI Appendix, Fig. S1A). Nlp20 was only found in 13% (565/4,228) of the genomes and predominantly in necrotic bacteria such as Streptomyces, Pectobacterium, and Dickeya (SI Appendix, Fig. S3 A and B).

We investigated MAMP conservation compared to the "consensus" sequence commonly used in the literature for immune assays. When all genera were analyzed together, each MAMP exhibited different distributions compared to the consensus epitope (Fig. 1D). The cumulative distributions of flg22 and flgII-28 variants closely mirror each other and are derived from different regions of the same gene, FliC (Fig. 1D and SI Appendix, Figs. S1B and S3C). However, a lower number of flgII-28 epitopes were detected (3,661) compared to flg22 (5,042), likely due to higher sequence diversity in the flgII-28 region (SI Appendix, Fig S2). Elf18 and nlp20 variants are the most and least conserved, respectively (Fig. 1D and SI Appendix, Fig. S3C).

When analyzing MAMP distributions in individual genera, interesting trends are identified (SI Appendix, Fig. S1A). Diversification of each FliC epitope in gram-negative bacteria does not always mirror each other. Dickeya and Pectobacterium exhibit similar amino acid similarity and CNV for both flg22 and flgII-28 (SI Appendix, Fig S1A). However, Agrobacterium has a multimodal distribution for flg22 and unimodal distribution for flgII-28 (SI Appendix, Fig S1A). Many gram-positives exhibit similar csp22 diversification, indicating ancient CSP paralog emergence (SI Appendix, Fig. S1A). We also compared epitope variation independent of a consensus sequence by calculating amino acid similarity in an all-by-all manner (SI Appendix, Fig. S1B). These results were complementary to the comparison to the consensus epitope, with different patterns of diversity for each MAMP. Overall, we observed that epitope abundance and sequence diversification evolved in both a MAMP-based and genera-derived manner.

Genes that encode MAMP epitopes are postulated to be either essential or conditionally essential for bacterial survival. Considering their ancient origin, we assessed the total degree of epitope variation, which we found was constrained. Except for nlp20, the total number of MAMP epitopes detected ranged from 3,661 to 20,542 (Fig. 1E). However, the total number of unique epitope variants was much lower with csp22 exhibiting the highest number (622)
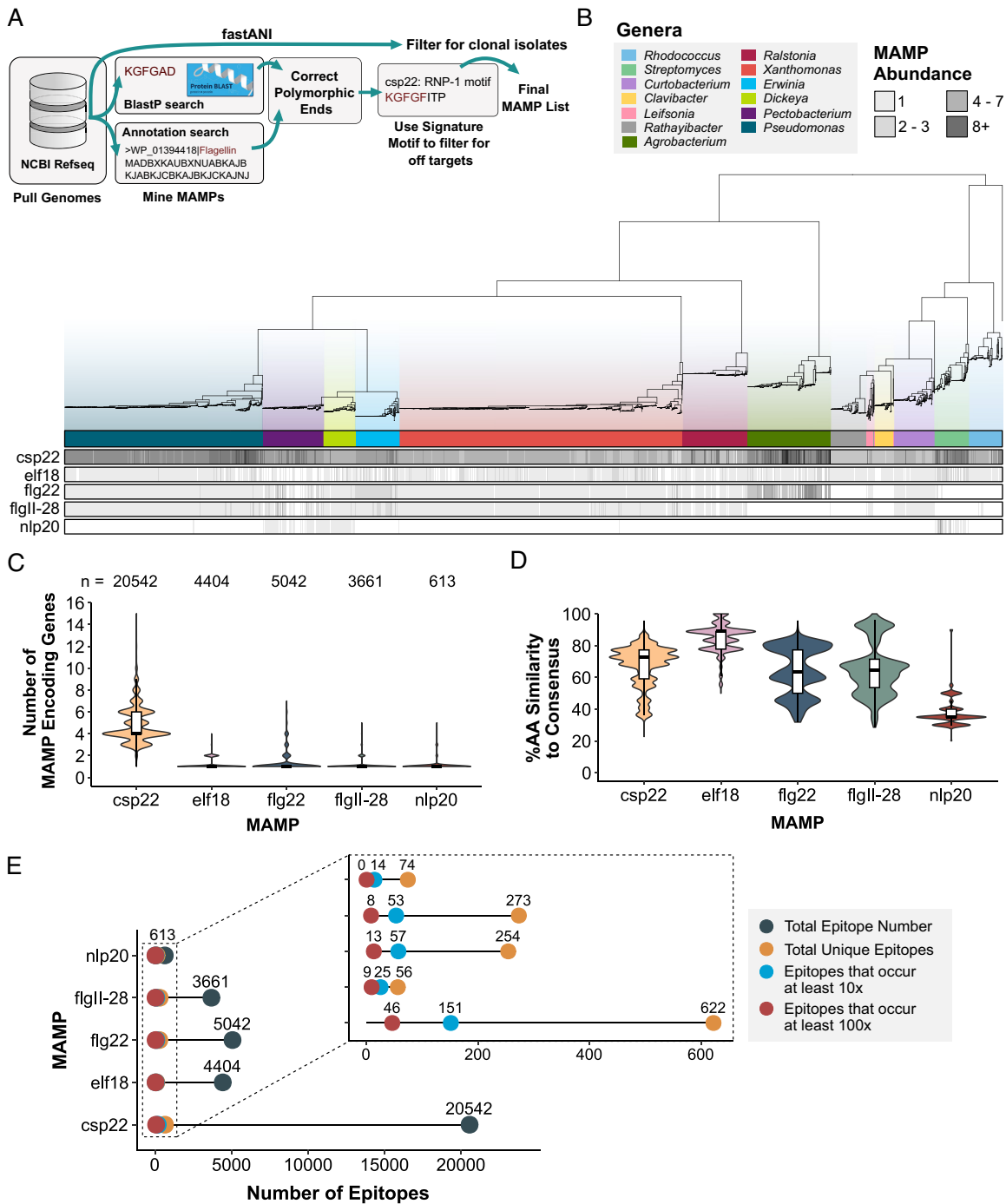
Fig. 1. Epitopes from diverse plant-associated bacteria exhibit different evolutionary trajectories. (*A*) Pipeline to mine for MAMP epitopes from 4,228 plant-associated bacterial genomes (see the *Materials and Methods* section for details). (*B*) Maximum-likelihood phylogeny built from 74 housekeeping genes with tips labeled by genera. Each bar below presents the number of epitopes (csp22, elf18, flg22, flgII-28, and nlp20) present in each genome. (*C*) A violin plot showing the number of MAMP-encoded genes from each genome separated by MAMP type. Tukey's boxplots are plotted on top. The number of epitopes assessed is listed at the *Top*. (*D*) A violin plot of percent amino acid (AA) similarity of each epitope variant in comparison to each respective consensus sequence across all bacteria sampled. Tukey's boxplots are plotted on top. (*E*) Lollipop plot displaying the number of unique MAMP epitopes and their occurring frequencies.

and elf18 displaying the lowest number (56, Fig. 1*E*). When considering the frequency of epitope variants, there was substantially fewer that occurred at least 10 or 100 times (between 8 and 151,

Fig. 1*E*). Collectively, these data indicate it is possible to experimentally test the impact of natural variation on immune outcomes across thousands of bacteria.

Most elf18 Variants Are Immunogenic. To characterize the functional diversity of elf18, 25 variants were synthesized and assessed for their immunogenicity using ROS, an early, quantitative output of immune induction (5, 35). ROS production by NADPH oxidases is required for biological processes, stress responses, and the induction of systematic acquired resistance, providing broad-spectrum resistance (36). The 25 elf18 epitopes (100 nM) were screened for their ability to induce ROS on *Arabidopsis thaliana* Col-0 and the *EFR* receptor mutant line. Water and the *E. coli*–derived consensus elf18 epitope were used as controls (*SI Appendix,* Fig. S4*A*). Of the 25 variants tested, 76% (19/25) displayed equal ROS output in comparison to the consensus, 16% were weakly immunogenic (4/25), and 8% (2/25) were nonimmunogenic (*SI Appendix,* Fig. S4*A*). No measurable ROS was produced in the *efr* mutant line (*SI Appendix,* Fig. S4*A*). To confirm our ROS screen was effective in classifying immunogenicity, a blinded, independent evaluation displayed the same conclusions as the initial screen (*SI Appendix,* Fig. S4*B*). Interestingly, all nonimmunogenic variants and half of the weakly immunogenic variants were derived from a second EF-Tu locus in *Streptomyces*, that clustered separately (Fig. 2*A*). While all epitopes were derived from annotated EF-Tu loci, the second *Streptomyces* EF-Tu copy displayed the most divergent elf18 sequence (Fig. 2*A* and *SI Appendix,* Fig. S4*A*).

Next, we characterized mid- to late-stage immune outputs including MAPK induction (100 nM), callose deposition (1 μM), and seedling growth inhibition (100 nM) for a subset of epitopes that exhibited different ROS immunogenicity. All secondary immune outputs were congruent with each other. All immune outputs by any elf18 epitope variant were abolished in the *efr* mutant (*SI Appendix,* Fig. S4 *C–E*). Epitopes EF-77 and EF-81, immunogenic by ROS, inhibited seedling growth, induced MAPK activation, and displayed high amounts of callose deposition (Fig. 2 *B–D*). In contrast, the weakly immunogenic and nonimmunogenic variants EF-91, 92, 96, and 97 failed to induce mid- to late-stage immune responses, mirroring water or mock controls (Fig. 2 *B–D*). Elf18 variants EF-91 and EF-96 can uncouple their immune outputs, potentially enabling pathogen evasion of strong immune response. Similarly, EF-80 is weakly immunogenic by ROS and has been previously shown to not induce robust callose deposition, a late-stage immune output (5, 37).

To understand the relationship between immunogenic outcomes and protein evolution, we developed a maximum-likelihood tree of EF-Tu and plotted genera, the percent amino acid similarity of each epitope to the consensus, and immunogenicity by ROS (Fig. 2*A*). The EF-Tu tree structure was indicative of taxonomic origin (Fig. 2*A*). The 25 experimentally tested epitope variants enabled the determination of immunogenicity of 98.8% (3,757/3,801) of elf18 variants across 3346 plant-associated bacteria (Fig. 2*E*). Most elf18 encoded variants (92.2%, 3,504/3,801) induce strong immunity when *EFR* was present (Fig. 2*E*). Furthermore, of the genomes that encode for more than one EF-Tu locus, only 35 have one immunogenic copy and second nonimmunogenic copy (Fig. 2*F*). While immune outcomes of weakly immunogenic variants from *Streptomyces, Xanthomonas,* and *Pseudomonas* are comparable, their polymorphisms are unique and their respective EF-Tu proteins do not cluster phylogenetically, indicating convergent evolution (Fig. 2*A* and *SI Appendix,* Fig. S4*A*) (37). Across all elf18 variants, polymorphisms were position dependent, showcasing that epitope variation may be constrained by EF-Tu function (*SI Appendix,* Fig. S4*A*).

Diversification of CSPs Contributes to Differential Immunogenic Responses. Unlike EF-Tu and elf18, CSP and csp22 exhibited CNV and epitope diversification (Fig. 1 *C* and *D* and *SI Appendix,*

Fig. S2*A*). This provided an opportunity to characterize the functional diversity of a second MAMP with a distinct evolutionary trajectory. Therefore, 65 csp22 epitope variants (200 nM) were screened for their ability to induce ROS on Rio Grande tomatoes. Water and the *Micrococcus luetus*–derived consensus csp22 epitope were used as controls (*SI Appendix,* Fig. S6*A*) (13). We tested all variants with one of two independent *core*-deficient lines developed in the Rio Grande cultivar via CRISPR-cas9 (*SI Appendix,* Figs. S5 and S6*A*). Each *core* line failed to perceive csp22 but could still perceive flg22 and flgII-28 (*SI Appendix,* Fig. S5*B*). Across all 65 variants, no ROS was produced in the *core* mutant line, demonstrating *CORE* specificity (*SI Appendix,* Fig. S6*A*). Only 54% (35/65) of csp22 epitope variants displayed equal ROS output in comparison to the consensus, while 11% (7/65) were weakly immunogenic and 25% (23/65) were nonimmunogenic (*SI Appendix,* Fig. S6*A*). A subset of the csp22 variants (25) were also assessed for their ability to induce ethylene production (1 μM). Although the resolution was qualitative, most epitope variants were consistent with our ROS screen (*SI Appendix,* Fig. S7).

Csp22 variation was explored across diverse CSP-domain loci. The CSP domain ranges from 65 to 75 amino acids in length and carries RPN-1 and RPN-2 nucleic acid binding motifs (22). While the CSP domain was conserved, considerable diversity in gene sequence length was found (*SI Appendix,* Fig. S6*B*). We identified CSPs carrying extra domains including calcium binding and DUF domains (*SI Appendix,* Fig. S6*B*). We identified an *Agrobacterium* CSP containing two cold shock protein domains fused together with a linker (*SI Appendix,* Fig. S6*B*). Classical CSPs (<75 amino acids, no additional domains) predominantly encoded immunogenic epitopes. However, several weakly or nonimmunogenic epitopes were encoded within CSPs with unique domain architecture (*SI Appendix,* Fig. S6). Some CSPs were conserved across multiple genera, while others were restricted to a single genus (*SI Appendix,* Fig. S6).

To understand the relationship between immunogenic outcomes and protein evolution, we developed a maximum-likelihood tree of the cold-shock domain and plotted genera, the percent amino acid similarity of each epitope to the consensus, and immunogenicity by ROS (Fig. 3*A*). Unlike EF-Tu, CSP domains display intricate clade structure (Fig. 3*A*). While over 19,000 cold-shock domain-containing proteins were extracted, the 65 csp22 variants tested still managed to capture 75% (14,587/19,423) of the immunogenetic landscape (Fig. 3 *A* and *B*). Three conserved clades contained divergent CSPs carrying nonimmunogenic csp22 epitopes. Nonimmunogenic CSPs within clade 1 were conserved across 62% (781/1,261) of *Xanthomonas,* clade 2 was conserved across 83% (394/475) of *Pectobacterium* and *Erwinia* with a distant relative in *Agrobacterium*, and clade 3 composed of 95% (580/612) of actinobacteria (Fig. 3 *C*, *Right* panel). Unlike EF-Tu, almost all genomes carried more than one CSP and of those, 45.7% had at least one nonimmunogenic form (Fig. 3 *C*, *Left* panel).

Observing the differences in immunogenic outcomes and sequence diversity, we then assessed conservation and selection of specific residues from the conserved nonimmunogenic CSP. Orthologous CSPs were classified using a combination of phylogeny, protein clustering, and motif classification (*SI Appendix, Materials and Methods* and Fig. S8 *A* and *B*). We then used an all-by-all BlastP approach to confirm orthology (*SI Appendix,* Fig. S8*C*). Next, d_N/d_S was calculated for nonimmunogenic and immunogenic CSPs from the same bacterial genera. For nonimmunogenic and immunogenic CSP loci, selection was assessed and codon sites that were considered significantly negative or positive were based on a set posterior probability (*SI Appendix,* Fig. S9*A*). Among representative immunogenic
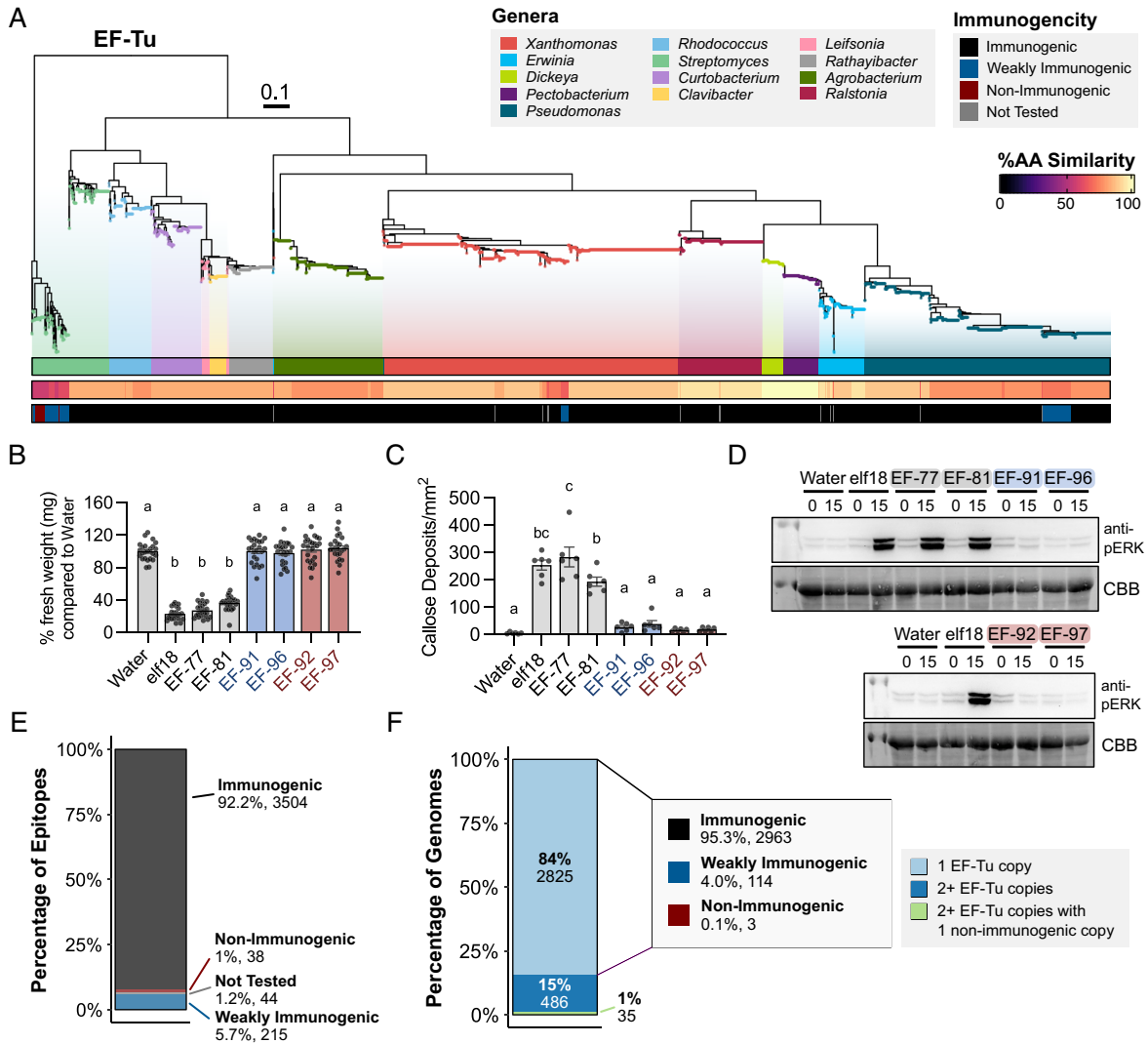
47

Fig. 2. elf18 exhibits minimal diversification, and most variants are immunogenic. (*A*) Maximum-likelihood phylogenetic tree of EF-Tu. Tips were colored by genera, the second bar represents percent amino acid similarity to consensus, and the third bar indicates immunogenicity by ROS from *SI Appendix,* Fig. S4. (*B*) *Arabidopsis* seedling growth inhibition after treatment with water or elf18 peptides (100 nM). One point = one plant. Eight plants per biological replicate. A one-way ANOVA and Tukey's mean comparison were used to determine significance, $P < 0.0001$. (*C*) Quantified callose deposition in Col-0 after treatment with water and elf18 variants (1 μM). Values are from one representative experiment which includes an average of at least two images per leaf from two leaves per plant, three plants per treatment. Statistical analyses were performed as described in *B*, $P < 0.0001$. (*D*) Induction of MAPK by elf18 variants (100 nM) in Col-0 at zero and 15 min postinfiltration. CBB = protein loading. All experiments were repeated at least three times with similar results. (*E*) Summary of immunogenic outcomes of EF-Tu-encoded elf18 epitopes from *SI Appendix,* Fig. S4. (*F*) Summary of immunogenic outcomes with respect to EF-Tu copy number.

and nonimmunogenic CSP members from the three clades, most sites are either negatively selected (purifying) or neutral with only a very small number of codons displaying positive (diversifying) selection (*SI Appendix,* Fig. S9A). For all tested immunogenic, weakly immunogenic, and nonimmunogenic epitopes, the RNP-1 motif is highly conserved and under purifying selection (*SI Appendix,* Fig. S9 *A* and *B*). Near the N terminus (highlighted in gray), a subset of residues between positions four and nine may be critical for strong immunity based on the residue changes, R-group chemistry, and immunogenic outcomes (*SI Appendix,* Fig. S9A). The epitope signature for each clade is unique and derived from distinct paralogs (Fig. 3A and *SI Appendix,* Fig. S9B).

Conserved Nonimmunogenic cps22 Variants Reduce CORE Immune Perception. We observed three nonimmunogenic CSP clades across different bacterial genera (Fig. 3 *A* and *C*). These nonimmunogenic CSPs may be maintained for their role in inhibiting immune perception. To test this, we modified our high-throughput ROS assay to assess epitope antagonism. Briefly, leaf disks were floated on one of two solutions: water or increasing concentrations of the candidate antagonist. After overnight incubation, all liquid was removed, 200 nM of agonist was used to elicit immunity, and ROS was measured (*SI Appendix,* Fig. S10A). We selected at least one member from each nonimmunogenic clade and used consensus csp22 as the agonist. Most nonimmunogenic csp22 epitope variants reduced ROS produced by consensus csp22. The variant conserved in the *Clavibacter* genus, *Cm*
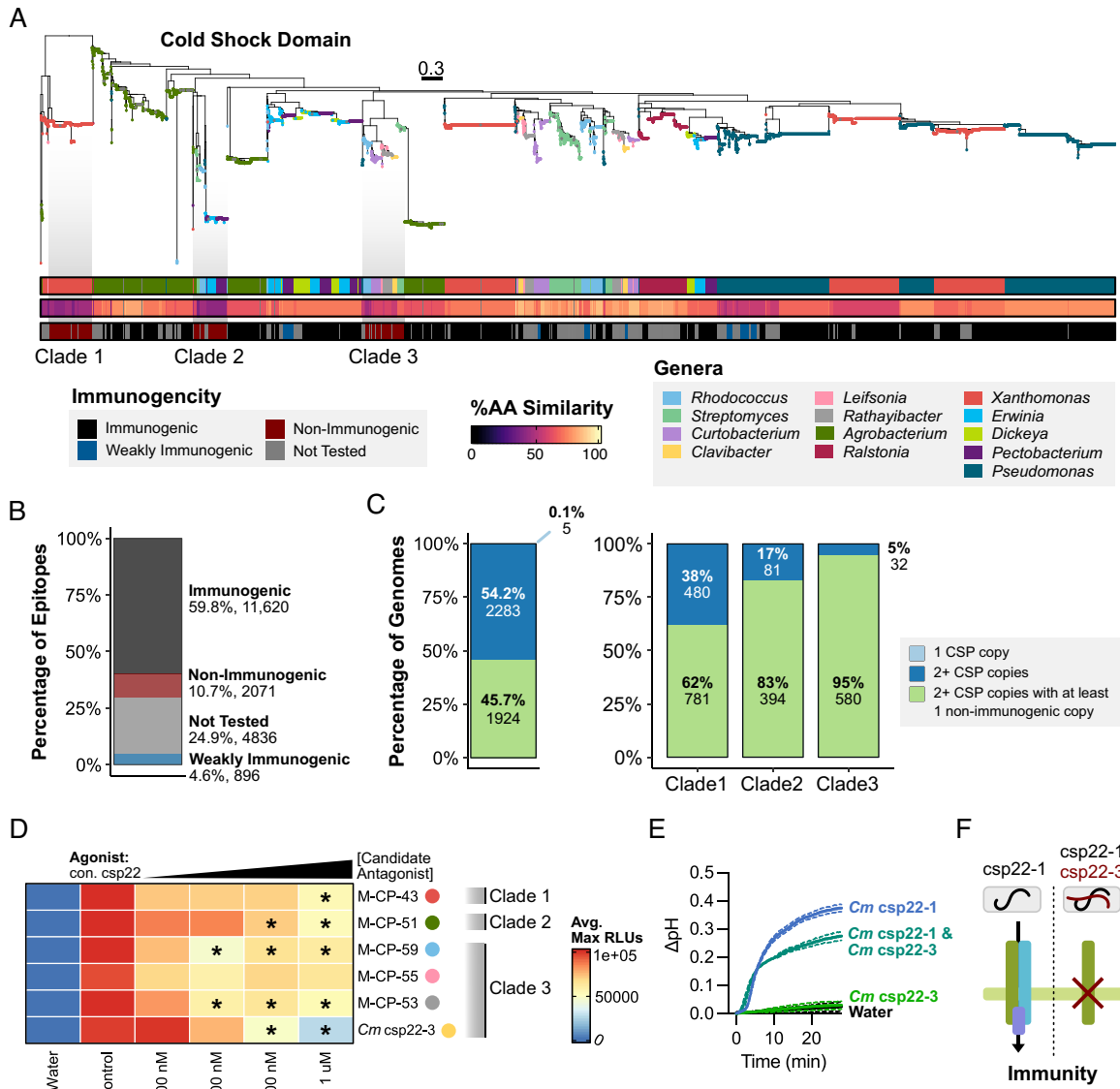
Fig. 3. Convergently evolved nonimmunogenic CSPs reduce CORE perception. (*A*) Maximum-likelihood phylogenetic tree of CSP domains. Tips were colored by genera, the second bar represents percent amino acid similarity to consensus, and the third bar indicates immunogenicity based on ROS experiments in *SI Appendix*, Fig. S6 (200 nM). (*B*) Summary of immunogenic outcomes of csp22 epitopes tested by ROS. (*C*) Summary of immunogenic outcomes with respect to CSP copy number (*Left*). Summary of immunogenic outcomes from bacteria present across each nonimmunogenic clade in A and their relatives in the same genera (*Right*). (*D*) ROS screen for immune antagonism in Rio Grande tomato. Leaf disks were incubated overnight with candidate antagonist followed by elicitation with 200 nM agonist peptide. Control denotes positive control (untreated immunogenic agonist), and concentrations listed are of the candidate nonimmunogenic antagonist. Maximum RLU averages were adjusted to a scale of 0 to 100,000 based on the controls (water and untreated agonist). Candidate antagonists, labeled by genera, were selected across the three conserved non-immunogenic-derived CSPs from panel (A). A one-way ANOVA and Tukey's mean comparison was used to determine significance in respect to the untreated agonist control (denoted *$P < 0.05$). (*E*) Alkalization of *Nicotiana tabacum* cv. Bright Yellow (BY-2) suspension cells after MAMP treatment (10 nM *Cm* csp22-1; 10 nM *Cm* csp22-3). Antagonism was assessed by first treating with 50 nM *Cm* csp22-3 for three minutes and then treating with 10 nM agonist *Cm* csp22-1. Two cell aliquots were tested per treatment, and the experiment was repeated at least two times. (*F*) Diagram of *Clavibacter* CSP intrabacterial antagonism.

csp22-3, displayed the strongest ROS reduction at concentrations which were 2.5× and 5× the respected agonist concentration (Fig. 3*D*).

Clavibacter contains three CSPs with different immunogenic outcomes, immunogenic *cspA1* (*Cm* csp22-1), weakly immunogenic *cspA2* (*Cm* csp22-2), and nonimmunogenic *cspB* (*Cm* csp22-3) (Fig. 3*A* and *SI Appendix*, Fig. S6*A*). We repeated the antagonism assay with agonist *Cm* csp22-1 alongside two negative controls: elf18 and a scrambled version of *Cm* csp22-3 (designated

s-csp22-3) (*SI Appendix*, Fig. S10). Antagonism was observed via decreased ROS production after incubation with *Cm* csp22-3 but not for the other negative control peptides (*SI Appendix*, Fig. S10). We additionally tested *Cm* csp22-3 antagonism using BY-2 tobacco cell cultures by measuring alkalization, an output of membrane depolarization through ion fluxes and inhibition of H+-ATPases at the plasma membrane (Fig. 3*E*) (38, 39). As expected, *Cm* csp22-1 (10 nM) was immunogenic and able to
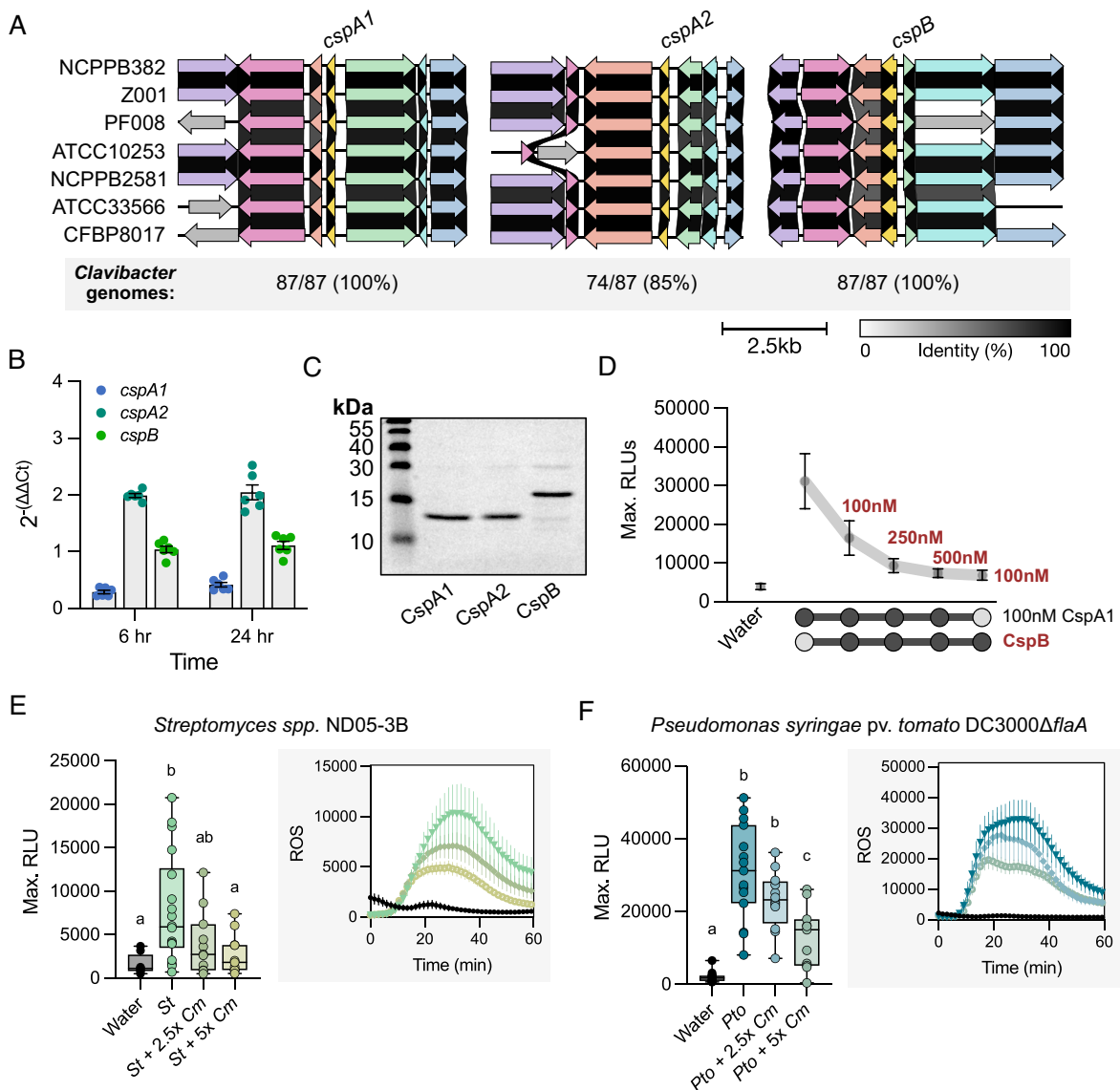
Fig. 4. *Clavibacter* CspB is conserved, expressed, and enables immune evasion of CORE. (*A*) Conservation and syntenic gene structure of CSPs in representative *Clavibacter* genomes (*C. michiganensis:* NCPPB382 and Z001; *C. capsici:* PF008; *C. insidiosus:* ATCC10253; *C. nebraskensis:* NCPPB2581; *C. tessellarius:* ATCC33566 and CFBP8017). (*B*) Expression of CSPs from *Clavibacter michiganensis* NCPBB382 in xylem-mimicking media compared to expression of the housekeeping gene, *bipA,* in rich media TBY. Three technical replicates for two independent biological cultures are plotted. The experiment was repeated three times with similar results. Error bars = SEM. (*C*) Coomassie-stained SDS-PAGE gel of CSP recombinant proteins purified from *E. coli*. CspA1 is 13 kDa and CspB is 17kDa. (*D*) Maximum ROS production in tomato Rio Grande. Recombinant CSP proteins were mixed in the indicated concentrations at the same time and applied to leaf tissue. Concentrations were 100 nM for *Cm* CspA1 and different concentrations of recombinant nonimmunogenic *Cm* CspB (denoted in red font). (*E*) ROS induction of *Streptomyces* sp. ND05-3B (1 μg/mL, *St*) and *C. michiganensis* NCPPB382 bacterial lysates (2.5 and 5 μg/mL, *Cm*) in Rio Grande tomato. (*F*) ROS induction of *Pseudomonas syringae* pv. *tomato* DC3000Δ*flaA* (1 μg/mL) and *C. michiganensis* NCPPB382 bacterial lysates (2.5 and 5 μg/mL) in Rio Grande tomato. Lysates were mixed in the indicated concentrations at the same time and applied to leaf tissue in *E* and *F*. In *D–F* the Max. RLUs include an average of four punches per plant, 12 plants per treatment. Error bars = SEM. A one-way ANOVA and Tukey's mean comparison was used to determine significance, $P < 0.05$.

induce a pH shift, while *Cm* csp22-3 (10 nM) was not. When 5× *Cm* csp22-3 was added 3 min before *Cm* csp22-1, it was able to decrease the pH shift, showcasing its antagonistic nature (Fig. 3*E*). Collectively, these data indicate that nonimmunogenic CSP epitope variants can antagonize perception of immunogenic forms encoded in the same genome, which we term intrabacterial antagonism (Fig. 3*F*).

We also observed that *Streptomyces* carried two copies of EF-Tu with distinct epitope variants (Fig. 2*A* and *SI Appendix,* Fig. S12*A*).

One nonimmunogenic elf18 variant was conserved across multiple *Streptomyces* genomes (EF-92), with a second nonimmunogenic variant found in one genome (EF-97). Using the same ROS antagonism assay, we tested both nonimmunogenic epitopes against two different agonists, consensus elf18 and the immunogenic *Streptomyces* variant, EF-77. We tested elf12, a truncated version of consensus elf18, which has been previously reported as a weak antagonist (11). As expected, elf12 was able to weakly reduce ROS induction for either agonist. However, neither EF-92 nor EF-97

50

displayed consistent ROS reduction at any concentration (*SI Appendix*, Fig. S12*B*). Therefore, the capacity for intrabacterial antagonism is MAMP-dependent.

Intrabacterial Antagonism of Conserved Actinobacterial CSP. Antagonism has been predominantly characterized using peptide assays. Therefore, we investigated intrabacterial antagonism with full-length CSPs. The *Clavibacter* genus comprises eight pathogenic species (40, 41). Pathogens in the *Clavibacter* genus predominantly colonize the xylem vasculature and upon systemic infection, colonize additional tissues (40). First, conservation and synteny of each *Clavibacter* CSP was analyzed across multiple species. Both immunogenic *cspA1* and nonimmunogenic *cspB* were found in all *Clavibacter* genomes, whereas the weakly immunogenic *cspA2* was found in 85% (74/87) (Fig. 4*A*). Across several species, each gene was highly syntenic (Fig. 4*A*). Expression of each *CSP* was assessed via qPCR for the tomato pathogen *C. michiganensis* in xylem-mimicking media and compared to a previously published housekeeping gene, *bipA,* in TBY rich medium. While all CSPs are expressed, *cspA2* and *cspB* exhibited higher expression than *cspA1* at both 6 and 24 h with *cspB* displaying between 2.65 and 3.63× fold change with *cspA1* (Fig. 4*B*). Furthermore, when recombinant protein CspA1 and CspB were mixed in varying concentrations, we observed a strong reduction in ROS at 2.5× concentrations differences, a realistic difference as expression and protein abundance are correlated in bacteria (Fig. 4 *C* and *D*) (42).

Next, we wanted to investigate the ability of *C. michiganensis* (*Cm*) cell lysates to suppress *CORE*-dependent immunity from another actinobacteria *Streptomyces spp.* ND05-3B (*St*) as well as the proteobacteria *Pseudomonas syringae* pv. *tomato* DC3000Δ*flaA* (*Pto*). Both *St* and *Pto* carry immunogenic CSPs and lack *cspB* (*SI Appendix*, Figs. S6*A* and S11). Both *Clavibacter* and *Streptomyces* are nonflagellated and lack *fliC*; we used a *Pto fliC* deletion mutant (Δ*flaA*). On *Arabidopsis*, *Cm* lysates strongly induce ROS while *St* and *Pto* weakly induce ROS in an *EFR*-dependent manner (*SI Appendix*, Fig. S11). *Cm* lysates were unable to induce ROS in tomato cv. Rio Grande or the *core* mutant line (*SI Appendix*, Fig. S11). However, *St* and *Pto* were able to induce ROS in a *CORE*-dependent manner. Furthermore, we were able to observe reduced ROS production in Rio Grande when *Cm* lysates were mixed with either immunogenic *St* or *Pto* lysates at a 2.5 to 5× concentration (Fig. 4 *E* and *F*). Collectively, these data demonstrate *Cm* lysates can antagonize *CORE*-mediated perception of *St* and *Pto* protein extracts.

We were unable to generate a *Clavibacter* knockout of *cspB* using a variety of approaches likely due to the region's high GC content, ranging from 73 to 78% (43, 44). In order to functionally assess intrabacterial CspB antagonism *in planta*, we expressed codon-optimized *cspB* in the tomato pathogen *Pto* DC3000 (Fig. 5). The five CSPs found in the DC3000 genome are immunogenic and expressed *in planta* (Fig. 5*A* and *SI Appendix*, Figs. S5*A* and S11) (42). CSPs are known to act as translational chaperones and anti-terminators during environmental conditions such as cold and stress (22). In vitro expression of *cspB* was confirmed via western blot
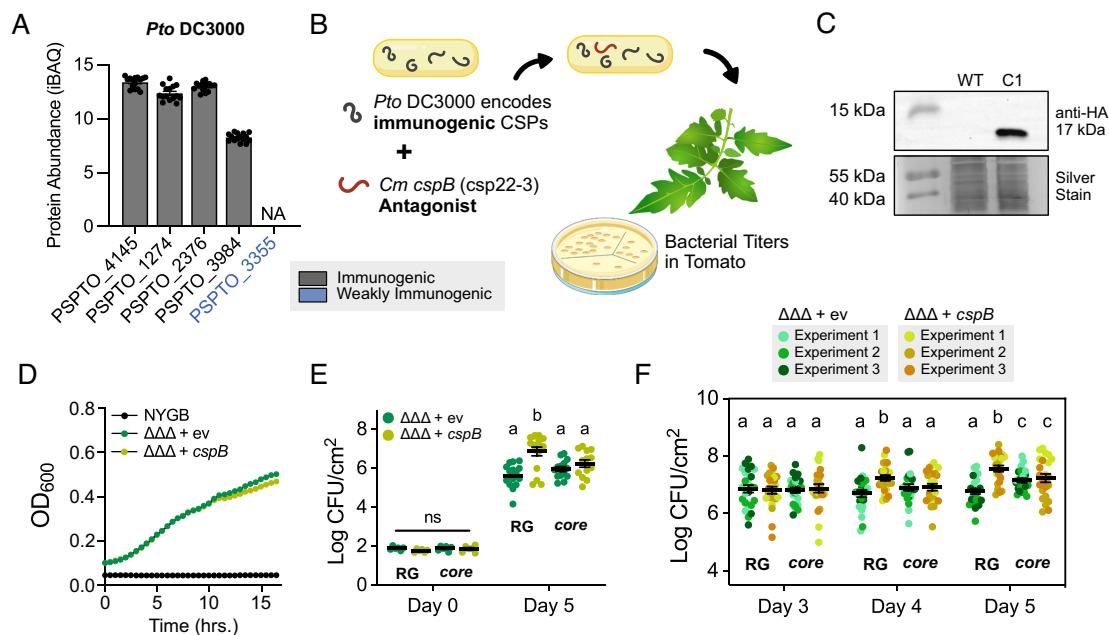

Fig. 5. Intrabacterial transfer of CspB antagonizes CORE perception in a gram-negative tomato pathogen. (*A*) Induction of *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000 CSPs based on raw count proteome data from Nobori et al. (42). Variants are labeled by their immunogenicity based on Fig. 3 and *SI Appendix*, Fig. S6. (*B*) Diagram of *Pto* DC3000 transformed with pDSK519 carrying codon-optimized *cspB* where bacterial titers were tested in tomato. (*C*) In vitro expression of pDSK519 carrying codon-optimized *cspB* (hereafter pDSK519-*cspB*) in *Pto* DC3000Δ*avrPto*Δ*avrPtoB*Δ*flaA* (ΔΔΔ) (OD$_{600}$ = 1). WT = wild-type strain DC3000ΔΔΔ. (*D*) Representative plot of DC3000ΔΔΔ in vitro growth with pDSK519-ngfp (ev) and with pDSK519-*cspB* in NYGB medium. (*E* and *F*) Bacterial titers of DC3000ΔΔΔ carrying pDSK519-ngfp (ev) or pDSK519-*cspB* in Rio Grande (RG) and *core* mutants. Plant genotypes are indicated in pairs. In *E*, leaves were sampled at zero and five days postinfiltration. For *F*, leaves were sampled from 3 to 5 d postinfiltration. At least two leaves were sampled per plant with at least three plants per strain. Experimental repeats in *F* are labeled by color. Error bars = SEM. A one-way ANOVA and Tukey's mean comparison was used to determine significance, *P* < 0.01.

(Fig. 5*C*). In vitro growth was not significantly different between *P. syringae* expressing *cspB* or empty vector, indicating that *cspB* expression does not grossly impact fitness (Fig. 5*D*).

In order to assess CspB antagonism, we used DC3000Δ*avrPto* Δ*avrPtoB*Δ*flaA* (referred to as DC3000ΔΔΔ) to prevent AvrPto/ AvrPtoB effector and flg22 recognition (45). After inoculation with DC3000ΔΔΔ, both *core* mutant lines exhibited increased disease symptoms and bacterial titers compared to the wild-type Rio Grande control (*SI Appendix*, Fig. S5 *C* and *D*). These data demonstrate that tomato *CORE* impacts *P. syringae* colonization. We quantified bacterial titers at zero and five days postinfiltration to assess the effect of expressing *cspB* in *Pto*. In Rio Grande, DC3000ΔΔΔ expressing *cspB* exhibited higher bacterial titers at five days postinoculation compared to DC3000ΔΔΔ carrying empty vector (Fig. 5*E*). We then performed a time course analyses between three and five days postinfiltration for DC3000ΔΔΔ expressing *cspB*. In Rio Grande, DC3000ΔΔΔ expressing *cspB* exhibited higher bacterial titers at four and five days postinoculation compared to DC3000ΔΔΔ carrying empty vector (Fig. 5*F*). It is likely that we detected a late increase in bacterial titer because cells require lysis to release CSPs and a minimum concentration of CspB is required to overcome immunogenic CSPs from *P. syringae*. Importantly, when either strain was inoculated on the *core* mutant line, there were no significant differences in bacterial titers between DC3000ΔΔΔ expressing *cspB* and DC3000ΔΔΔ carrying empty vector over the course of 5 d (Fig. 5*F*). Thus, intrabacterial expression of CspB is sufficient to inhibit CORE perception. Collectively, these data demonstrate that *Clavibacter* CspB is an intrabacterial antagonist.

Discussion

Natural epitope variation showcases outcomes which can serve to inform evolution, mechanistic interactions, and improve engineering approaches. We surveyed thousands of plant-associated bacteria and found variability in epitope sequence and CNV in a MAMP- and genera-dependent manner. Since natural variation was relatively low, characterizing the immune landscape for MAMP perception across thousands of bacteria is possible. Our work showcases the predictable nature of epitope evolution, enabling future rapid prediction of MAMP immunogenic outcomes. We uncovered a mechanism of immune evasion, intrabacterial antagonism, which demonstrates that all genome-encoded MAMPs need to be considered when characterizing bacterial-plant outcomes.

Millions of years of evolution led to bacterial divergence into different phyla. Recent work has assessed sequence variation for the flagellin epitope, flg22, revealing diverse yet constrained variation (32, 35). Colaianni et al. identified 268 unique variants from 627 *Arabidopsis*-associated genomes. We also identified a similar number of flg22 epitope variants (254) from 4,228 genomes, highlighting the constrained nature of epitope variation across plant-associated bacteria (35). In contrast, 1,059 flg22 variants were identified from 1,414 genomes from representative proteobacterial genomes across diverse lifestyles (32). These studies revealed differential host immunogenic outcomes based on specific residue positions. In flg22, sites 14 and 15 can have a direct impact on immune perception but consequently impair motility (8, 32, 46). Therefore, the prevalence of some mutations may be constrained due to negative fitness effects.

Epitope diversification may shift when multiple copies are present (35, 47). Gene expansion can impact both protein function and epitope variation. For some *Pseudomonas* strains, FliC-1 is critical for both plant immune perception and motility, while FliC-2 is poorly expressed and only weakly induces immune

perception (47). For FliC, EF-Tu and CSPs, certain residues changes in their respective epitope sequences were specific to additional copies (Figs. 2 and 3) (47). With the exception of *Streptomyces,* most other bacterial genera carry a single EF-Tu locus and polymorphisms were highly constrained and predictable in position, likely reflective of the protein's critical function to bacterial life (*SI Appendix*, Fig. S4*A*) (48). In contrast, CSPs are predominantly multicopy genes, which likely influences their expanded variation (Figs. 1 and 3). Individual CSPs can have minimal effects on bacterial survival when knocked out, although they can affect pathogen virulence (22, 27, 28, 49).

Some MAMP-encoded genes have multiple epitopes which can be recognized by different receptors (50–52). Different epitopes of FliC can be detected from distinct PRRs in *Arabidopsis*, tomato, rice, and vertebrates (15, 21, 53). Plant genomes contain hundreds of candidate PRRs; thus, it is possible that lineage-specific, convergent evolved receptors are common and the broad conservation of FLS2 is the exception (54). While *CORE* is restricted to the Solanaceae, other genotypes in the Vitaceae and Rutaceae families respond to csp22 variants (10, 55). Considering CSP diversity, it is likely that other convergently evolved PRRs recognize and respond to csp22 variants.

MAMP-derived genes have been rarely assessed beyond the first copy. One study assessed how a small number of *Pseudomonas* strains carrying two *FliC* genes exhibited differential immune outputs when individually assessed (47). For bacteria carrying one *FliC*, antagonism has been demonstrated against other genera (35, 47). Here, we found that actinobacteria contain multiple encoded CSPs where one copy blocks perception of additional immunogenic forms, thus functioning as intrabacterial antagonist (Figs. 3–5). Gram-positive pathogens lack the type three secretion system and cannot deliver immune-suppressing effectors directly into host cells (40). Therefore, intrabacterial CSP antagonism provides an additional mechanism of immune suppression. Intrabacterial MAMP antagonism may also influence community dynamics or mixed infections.

While total variation differs between MAMPs, csp22 epitopes exhibited the most variation (*SI Appendix,* Fig. S2). MAMP epitopes are proposed to undergo an arms race and are presumed to be under diversifying selection (56, 57). However, EF-Tu is under purifying selection (57). Furthermore, across multiple CSP orthologs present in genera with both immunogenic and nonimmunogenic forms, most codons displayed either neutral or purifying selection (*SI Appendix*, Fig. S9*A*). There are many CSP copies per genome with potentially different functions, frequently ranging from three to nine, thus the selection pressure on any one CSP may be less compared to other loci (Fig. 3*A* and *SI Appendix,* Fig. S2). Additionally, perception of CSPs is thought to require cell lysis for perception, which may impact diversifying selection.

By characterizing the consequence of natural epitope variation on immune perception, alternate immune outcomes have been revealed. While reduction in ROS production by some epitope variants was measurable, we failed to characterize any elf18 or csp22 epitopes which displayed statistically significantly higher ROS production compared to the consensus controls (*SI Appendix,* Figs. S4*A* and S6*A*). This may reveal that a minimal number of residues at specific positions are sufficient for strong complex formation and immune induction. Some flg22 variants, called deviant peptides, can uncouple immune outputs; they are able to induce early-stage ROS production but fail to induce late-stage immune responses such as SGI (35). This phenomenon of uncoupled immune outputs was also found for some elf18 variants (Fig. 2). Two weakly immunogenic variants by ROS, one from *Xanthomonas* and one from *Streptomyces,* were unable to induce

SGI, callose deposition, or MAPK phosphorylation. It is possible that MAMPs beyond flg22 and elf18 may also act as deviants, representing another strategy to reduce strong immune outputs. We envision two possible explanations for deviant peptide evolution. One, we may be capturing a snapshot of peptide evolution toward either maskers or antagonists. Two, these deviant peptides may occupy the space of the receptor complex, limiting robust immune outputs by other immunogenic peptides. These strategies are not unique to pathogens alone; beneficial microbial communities have been shown to be enriched in antagonistic FliC variants, which block perception of FLS2 (35).

A common strategy to confer disease resistance is to transfer receptors between plant genotypes. *EFR* has been transferred to tomato, citrus, wheat, and apple (58–61). In each case, *EFR* transgenic plants are able to significantly restrict pathogen titers and disease progression. This conclusion is congruent with our elf18 screen since most EF-Tu epitopes (92%) induce strong immune responses in *A. thaliana* (Fig. 2). In contrast, prevalent nonimmunogenic epitopes exist from FliC and CSPs, of which some block receptor recognition (Fig. 3 and *SI Appendix*, Fig. S2) (30, 32, 35, 62). Therefore, careful consideration of pathogen epitope variation should be considered for receptor transfer and subsequent engineering (63). We encourage future work using this dataset to inform which receptors may be optimal for pathogen control. In concert, large epitope variant databases may enable new synthetic receptor engineering via protein modeling of receptor structure. Rational receptor design focused on contact with low-polymorphic ligand residues may delay evolution of pathogen evasion.

Materials and Methods

Bacterial genomes were pulled from NCBI's Refseq, and bacterial epitopes were mined via BlastP, local alignment, and custom R scripts for filtering (Fig. 1*A* and Dataset S1) (6, 11, 13, 16, 17). Phylogenetic trees for bacteria relatedness were built using GToTree, and protein trees were built using MAFFT for sequence alignment and IQ-TREE tree building. Epitopes of interest that were assessed in planta were chosen based on abundance, epitope sequence, and gene annotation. ROS, MAPK, seedling growth inhibition, callose deposition, and alkalinization were measured similarly as previously described with some minor modifications (13, 31, 35).

Gene structure was assessed in *C. michiganensis* using BlastN and Clinker. Measurement of *C. michiganensis* CSP expression was conducted similarly as previously described (64). Bacteria were cultured in liquid media or on plates and processed in lysis buffer via an Emulsiflex-C3 High-Pressure Homogenizer to collect lysates. Recombinant protein was expressed in *E. coli* and collected using standard techniques (7). Bacterial titers of *P. syringae* in tomato were conducted similarly as described before (14). Detailed Materials and Methods are found in *SI Appendix*.

Author affiliations: [a]Integrative Genetics and Genomics Graduate Group, University of California, Davis, CA 95616; [b]Department of Plant Pathology, University of California, Davis, CA 95616; [c]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331; [d]Center for Plant Molecular Biology, University of Tübingen, Tübingen 72074, Germany; [e]Boyce Thompson Institute for Plant Research, Ithaca, NY 14853; [f]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853; and [g]Department of Biology, University of Washington, Seattle, WA 98195

1. D. C. Lewis, D. M. Stevens, H. Little, G. L. Coaker, R. M. Bostock, Overlapping local and systemic defense induced by an oomycete fatty acid MAMP and brown seaweed extract in Tomato. *Mol. Plant Microbe Interact.* **36**, 359–371 (2023).
2. B. P. M. Ngou, P. Ding, J. D. G. Jones, Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34**, 1447–1478 (2022).
3. D. Ge, I.-C. Yeo, L. Shan, Knowing me, knowing you: Self and non-self recognition in plant immunity. *Essays Biochem.* **66**, 447–458 (2022).
4. S. Lolle, D. Stevens, G. Coaker, Plant NLR-triggered immunity: From receptor activation to downstream signaling. *Curr. Opin. Immunol.* **62**, 99–105 (2020).
5. X. Yu, B. Feng, P. He, L. Shan, From chaos to harmony: Responses and signaling upon microbial pattern recognition. *Annu. Rev. Phytopathol.* **55**, 1–29 (2016).
6. G. Felix, J. D. Duran, S. Volko, T. Boller, Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
7. Y. Wei *et al.*, An immune receptor complex evolved in soybean to perceive a polymorphic bacterial flagellin. *Nat. Commun.* **11**, 3763 (2020).
8. K. Parys *et al.*, Signatures of antagonistic pleiotropy in a bacterial flagellin epitope. *Cell Host Microbe* **29**, 620–634.e9 (2021).
9. L. Wu, H. Xiao, L. Zhao, Q. Cheng, CRISPR/Cas9-mediated generation of fls2 mutant in Nicotiana benthamiana for investigating the flagellin recognition spectrum of diverse FLS2 receptors. *Plant Biotechnol. J.* **20**, 1853–1855 (2022).
10. J. Trinh *et al.*, Variation in microbial feature perception in the Rutaceae family with immune receptor conservation in citrus. *Plant Physiol.* **193**, 689–707 (2023).
11. G. Kunze *et al.*, The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. *Plant Cell Online* **16**, 3496–3507 (2004).
12. C. Zipfel *et al.*, Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts agrobacterium-mediated transformation. *Cell* **125**, 749–760 (2006).
13. G. Felix, T. Boller, Molecular Sensing of Bacteria in Plants. The highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. *J. Biol. Chem.* **278**, 6201–6208 (2003).
14. L. Wang *et al.*, The pattern-recognition receptor CORE of Solanaceae detects bacterial cold-shock protein. *Nat. Plants* **2**, 16185 (2016).
15. S. R. Hind *et al.*, Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system. *Nat. Plants* **2**, 16128 (2016).
16. R. Cai *et al.*, The plant pathogen Pseudomonas syringae pv. tomato is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathog.* **7**, e1002130 (2011).
17. H. Böhm *et al.*, A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in arabidopsis. *Plos Pathog.* **10**, e1004491 (2014).
18. I. Albert *et al.*, An RLP23-SOBIR1-BAK1 complex mediates NLP-triggered immunity. *Nat. Plants* **1**, nplants2015140 (2015).
19. M. F. Seidl, G. V. den., Ackerveken, activity and phylogenetics of the broadly occurring family of microbial Nep1-like proteins. *Annu. Rev. Phytopathol.* **57**, 367–386 (2019).
20. F. F. V. Chevance, K. T. Hughes, Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6**, 455–465 (2008).
21. Y. Rossez, E. B. Wolfson, A. Holmes, D. L. Gally, N. J. Holden, Bacterial Flagella: Twist and stick, or dodge across the kingdoms. *PLoS Pathog.* **11**, e1004483 (2015).
22. Y. Zhang, C. A. Gross, Cold shock response in bacteria. *Annu. Rev. Genet.* **55**, 1–24 (2021).
23. B. Schmid *et al.*, Role of cold shock proteins in growth of listeria monocytogenes under cold and osmotic stress conditions. *Appl. Environ. Microbiol.* **75**, 1621–1627 (2009).
24. L. P. Burbank, D. C. Stenger, A temperature-independent cold-shock protein homolog acts as a virulence factor in Xylella fastidiosa. *Mol. Plant Microbe Interact.* **29**, 335–344 (2016).
25. A. K. Eshwar, C. Guldimann, A. Oevermann, T. Tasara, Cold-shock domain family proteins (Csps) are involved in regulation of virulence, cellular aggregation, and flagella-based motility in Listeria monocytogenes. *Front. Cell. Infect. Microbiol.* **7**, 453 (2017).
26. A. Catalan-Moreno *et al.*, One evolutionarily selected amino acid variation is sufficient to provide functional specificity in the cold shock protein paralogs of Staphylococcus aureus. *Mol. Microbiol.* **113**, 826–840 (2020).
27. W. Wei, T. Sawyer, L. Burbank, Csp1, a cold shock protein homolog in Xylella fastidiosa influences cell attachment, pili formation, and gene expression. *Microbiol. Spectr.* **9**, e01591-21 (2021).
28. L. Wu, L. Ma, X. Li, Z. Huang, X. Gao, Contribution of the cold shock protein CspA to virulence in Xanthomonas oryzae pv. oryzae. *Mol. Plant Pathol.* **20**, 382–391 (2019).
29. K. L. Harvey, V. M. Jarocki, I. G. Charles, S. P. Djordjevic, The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. *Front. Microbiol.* **10**, 2351 (2019).

30. S. Wang *et al.*, Rice OsFLS2-mediated perception of bacterial flagellins is evaded by Xanthomonas oryzae pvs. oryzae and oryzicola. *Mol. Plant* **8**, 1024–1037 (2015).
31. Y. Wei *et al.*, The Ralstonia solanacearum csp peptide, but not flagellin-derived peptides, is perceived by plants from the Solanaceae family. *Plant Biotechnol. J.* **16**, 1349–1362 (2018).
32. J. H. T. Cheng, M. Bredow, J. Monaghan, G. C. diCenzo, Proteobacteria contain diverse flg22 epitopes that elicit varying immune responses in Arabidopsis thaliana. *Mol. Plant Microbe Interact.* **34**, 504–510 (2021).
33. U. Fürst *et al.*, Perception of Agrobacterium tumefaciens flagellin by FLS2XL confers resistance to crown gall disease. *Nat. Plants* **6**, 22–27 (2020).
34. P. Buscaill, R. A. L. van der Hoorn, Defeated by the nines: nine extracellular strategies to avoid microbe-associated molecular patterns recognition in plants. *Plant Cell* **33**, koab109 (2021).
35. N. R. Colaianni *et al.*, A complex immune response to flagellin epitope variation in commensal communities. *Cell Host Microbe* **29**, 635–649.e9 (2021).
36. B. Castro *et al.*, Stress-induced reactive oxygen species compartmentalization, perception and signalling. *Nat. Plants* **7**, 403–412 (2021).
37. F. Yang, G. Li, G. Felix, M. Albert, M. Guo, Engineered Agrobacterium improves transformation by mitigating plant immunity detection. *New Phytol.* **237**, 2493–2504 (2023).
38. M. M. Atkinson, J.-S. Huang, K. A. Knopp, The Hypersensitive reaction of tobacco to Pseudomonas syringae pv. Pisi: Activation of a plasmalemma K+/H+ exchange mechanism. *Plant Physiol.* **79**, 843–847 (1985).
39. J. M. Elmore, G. Coaker, The role of the plasma membrane H+-ATPase in plant-microbe interactions. *Mol. Plant* **4**, 416–427 (2011).
40. S. P. Thapa *et al.*, The evolution, ecology, and mechanisms of infection by gram-positive, plant-associated bacteria. *Annu. Rev. Phytopathol.* **57**, 341–365 (2019).
41. D. Arizala, S. Dobhal, A. M. Alvarez, M. Arif, Elevation of Clavibacter michiganensis subsp. californiensis to species level as Clavibacter californiensis sp. nov., merging and re-classification of Clavibacter michiganensis subsp. chilensis and Clavibacter michiganensis subsp. phaseoli as Clavibacter phaseoli sp. nov. based on complete genome in silico analyses. *Int. J. Syst. Evol. Microbiol.* **72**, e1-e14 (2022).
42. T. Nobori *et al.*, Multidimensional gene regulatory landscape of a bacterial pathogen in plants. *Nat. Plants* **6**, 883–896 (2020).
43. L. Chalupowicz *et al.*, Differential contribution of Clavibacter michiganensis ssp. michiganensis virulence factors to systemic and local infection in tomato. *Mol. Plant Pathol.* **18**, 336–346 (2017).
44. D. M. Stevens, A. Tang, G. Coaker, A genetic toolkit for investigating Clavibacter species: Markerless deletion, permissive site identification, and an integrative plasmid. *Mol. Plant Microbe Interact.* **34**, 1336–1345 (2021).
45. B. H. Kvitko *et al.*, Deletions in the repertoire of Pseudomonas syringae pv. tomato DC3000 type III secretion effector genes reveal functional overlap among effectors. *Plos Pathog.* **5**, e1000388 (2009).
46. K. Naito *et al.*, Amino acid sequence of bacterial microbe-associated molecular pattern flg22 is required for virulence. *Mol. Plant Microbe Interact.* **21**, 1165–1174 (2008).
47. Y. Luo, J. Wang, Y.-L. Gu, L.-Q. Zhang, H.-L. Wei, Duplicated flagellins in pseudomonas divergently contribute to motility and plant immune elicitation. *Microbiol. Spectr.* **11**, e03621-22 (2023).
48. P. J. P. Teixeira, N. R. Colaianni, C. R. Fitzpatrick, J. L. Dangl, Beyond pathogens: Microbiota interactions with the plant immune system. *Curr. Opin. Microbiol.* **49**, 7–17 (2019).
49. Y. Liu *et al.*, The cold shock family gene cspD3 is involved in the pathogenicity of Ralstonia solanacearum CQPS-1 to tobacco. *Microb. Pathog.* **142**, 104091 (2020).
50. T. Furukawa, H. Inagaki, R. Takai, H. Hirai, F.-S. Che, Two distinct EF-Tu epitopes induce immune responses in rice and arabidopsis. *Mol. Plant Microbe Interact.* **27**, 113–124 (2014).
51. Y. Katsuragi *et al.*, CD2-1, the C-terminal region of flagellin, modulates the induction of immune responses in rice. *Mol. Plant Microbe Interact.* **28**, 648–658 (2015).
52. M. de Andrade *et al.*, Suppression of citrus canker disease mediated by flagellin perception. *Mol. Plant Pathol.* **24**, 331–345 (2023).
53. J. Fliegmann, G. Felix, Immunity: Flagellin seen from all sides. *Nat. Plants* **2**, 16136 (2016).
54. B. P. M. Ngou, R. Heal, M. Wyler, M. W. Schmid, J. D. G. Jones, Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nat. Plants* **8**, 1146–1152 (2022).
55. L. P. Burbank, J. Ochoa, Evidence for elicitation of an oxidative burst in Vitis vinifera by Xylella fastidiosa cold shock protein peptide csp20. *PhytoFrontiers* **2**, 339–341 (2022).
56. H. C. McCann, H. Nahal, S. Thakur, D. S. Guttman, Identification of innate immunity elicitors using molecular signatures of natural selection. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4215–4220 (2012).
57. N. Eckshtain-Levi, A. J. Weisberg, B. A. Vinatzer, The population genetic test Tajima's D identifies genes encoding pathogen-associated molecular patterns and other virulence-related genes in Ralstonia solanacearum. *Mol. Plant Pathol.* **19,** 2187–2192 (2018).
58. S. Lacombe *et al.*, Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat. Biotechnol.* **28**, 365 (2010).
59. H. Schoonbeek *et al.*, Arabidopsis EF-Tu receptor enhances bacterial disease resistance in transgenic wheat. *New Phytol.* **206**, 606–613 (2015).
60. L. K. Mitre *et al.*, The Arabidopsis immune receptor EFR increases resistance to the bacterial pathogens Xanthomonas and Xylella in transgenic sweet orange. *Plant Biotechnol. J.* **19**, 1294–1296 (2021).
61. S. Piazza *et al.*, The Arabidopsis pattern recognition receptor EFR enhances fire blight resistance in apple. *Hortic. Res.* **8**, 204 (2021).
62. C. Pfund *et al.*, Flagellin is not a major defense elicitor in Ralstonia solanacearum cells or extracts applied to Arabidopsis thaliana. *Mol. Plant Microbe Interact.* **17**, 696–706 (2004).
63. A. Schultink, A. D. Steinbrenner, A playbook for developing disease-resistant crops through immune receptor identification and transfer. *Curr. Opin. Plant Biol.* **62**, 102089 (2021).
64. N. Jiang *et al.*, Evaluation of suitable reference genes for normalization of quantitative reverse transcription PCR analyses in Clavibacter michiganensis. *Microbiol. Open* **8**, e928 (2019).
65. D. M. Stevens *et al.*, Evolutionary dynamics of proteinaceous MAMPs reveals intrabacterial antagonism of plant immune perception. Zenodo. https://zenodo.org/records/10724865. Deposited 28 February 2024.
66. D. M. Stevens, G. Coaker, DanielleMStevens/Mining_Known_MAMPs. Github. https://github.com/DanielleMStevens/Mining_Known_MAMPs. Deposited 4 September 2023.

PNAS

Supporting Information for
Natural variation of immune epitopes reveals intrabacterial antagonism.

Danielle M. Stevens, Alba Moreno-Pérez, Alexandra J. Weisberg, Charis Ramsing, Judith Fliegmann, Ning Zhang, Melanie Madrigal, Gregory Martin, Adam Steinbrenner, Georg Felix, and Gitta Coaker

Correspondence to: glcoaker@ucdavis.edu

This PDF file includes:

SI Materials and Methods
Figures S1 to S12
Tables S1 to S3
SI References

Other supporting materials for this manuscript include the following:

Datasets S1

1

SI Materials and Methods

Collection of Genomes and MAMPs

Bacterial genomes were pulled from NCBI's RefSeq data repository. Briefly, using the –dry-run command for the ncbi-genome-download package (v0.3.0), candidate genomes were queried and manually assessed for associations with plants and/or agriculture. Once a list for each major genre in this study, *Clavibacter, Leifsonia, Curtobacterium, Streptomyces, Rathayibacter, Rhodococcus, Agrobacterium, Ralstonia, Xanthomonas, Pseudomonas, Pectobacterium, Dickeya, and Erwinia,* was collected, all Genbank, protein fasta and whole genome fasta files were downloaded for each Refseq accession number (Dataset S1) using ncbi-genome-download. Consensus MAMPs csp22 (AVGTVKWFNAEKGFGFITPDDG), elf18 (SKEKFERTKPHVNVGTIG), flg22 (QRLSTGSRINSAKDDAAGLQIA), flgII-28 (ESTNILQRMRELAVQSRNDSNSATDREA), and nlp20 (AIMYSWYFPKDSPVTGLGHR) were used to build a custom blast database via makeblastdb. Each protein fasta file was queried against the database using Blastp with the following modifications: -task blastp-short -xdrop_gap_final 1000 -soft_masking false -evalue 1e-4 (1). Secondly, MAMPs were found by local alignment to proteins with typical gene annotations (flagellin, cold shock protein, elongation factor, and necrosis-and-ethylene inducing peptide). Custom R scripts were built process blast results, filter for missing MAMP hits based on protein annotation, correction for gaps on polymorphic ends, filtering for off-targets based on a signature motif for each MAMP and removing hits from partial protein annotations (Fig. 1A). Motifs used to filter for off-targets include: csp22: "KGFGF", elf18: "NXGTXG", flg22: "SXGXXXXXXXXXAA", and flgII-28: "LQRXRXL". Table S1 includes the number of MAMPs detected using this pipeline.

Comparing genome similarity and MAMP hits to filter for clonality

To assess intra-genera genome diversity, average nucleotide identity via fastANI (v1.32) was calculated in an all-by-all manner within each genus using default parameters (Extended Data Fig. 1; Doi: 10.5281/zenodo.10724865) (2,3). Custom R scripts were used to parse ANI values. Genomes which shared an ANI value of 99.999 percent or higher and shared identical MAMP sequences were considered clonal. Of those that were categorized as clonal, only one representative genome was used. Full description of the methods of ANI analysis can be found in the Github, DanielleMStevens/Mining_Known_MAMPs, particularly under methods and R scripts #10 through 13. Genome accessions which were found to be clonal and removed from subsequent analysis can be found in Supplemental Table 1. The final MAMP list was outputted into an html file using R package DT (v0.20) and can be found in Zenodo (Doi: 10.5281/zenodo.10724865) (3).

Determining epitope diversity

After all epitopes were extracted, we assessed epitope diversity across a variety of scales. First, we determined how similar the epitopes are in comparison to the consensus peptide were by calculating protein similarity via BlastP as listed in Collection of Genomes and MAMPs section or by local-alignment and calculated via R package Biostrings (v2.58.0), which was used when the BlastP results dropped ends which were polymorphic in sequence compared to the consensus query used (R script #18). Two, to understand epitope variation which did not rely on consensus, a custom R script (#17) was developed to determine all the unique combinations found and performed all-by-all local alignment and amino acid similarity calculations. All-by-all similarity values were plotted as a heatmap using ComplexHeatmap (v2.5.1) and circlize (v0.4.8) (4,5). Three, epitope weblogos were made using ggseqlogo (v0.1) (R script #26) (6). Four, the number of epitope variants (described combination) was determined using a custom script (R script #19) and plotted using ggplot and ggbreak (v0.1.1) (7).

Epitope Selection Criteria

To determine which epitopes to evaluate, we first selected epitope variants based on their abundance in the dataset (those which occurred at least 10 to 100x). Additionally, epitope variants which were encoded in additional paralogs with unique evolutionary history, domain structure, and/or annotation were selected. In total, we selected 25 elf18 variants and 65 csp22 variants for synthesis and subsequent testing in plants (Table S2).

Construction of Phylogenetic Trees

To assess the relatedness of plant-associated bacteria in respect to their MAMP abundance, a phylogenomic tree was constructed using GToTree (v1.6.31) with the MAMPs mined plotted (8). GToTree is an automated workflow for developing phylogenetic trees based on a pre-set list of validated conserved genes (74 for bacteria). First, a list of paths of all GenBank files was outputted into a text file. Genomes which were indicated to be clonal by ANI and MAMPs detected were manually removed. GToTree was ran on the text file with the GenBank file paths using default parameters with the following modifications: -H Bacteria -G 0.2 -k.

Phylogenetic trees were built from the CSP domain and full-length sequences of EF-Tu. Proteins sequences were collected into a fasta file, and a multiple sequence alignment was built via MAFFT (v7.310) using the following parameters: --reorder --thread 12 --maxiterate 1000 –localpair (9). For CSP phylogeny, some CSPs that were longer than 180 amino acids in length were discovered to have additional domains. Therefore, the conserved cold shock protein domain was extracted across all CSPs using via HMMER (v3.3.2) using hmmsearch with the following parameters, -E 1 --domE 1 --incE 0.01 --incdomE 0.04, and the hmm model specific for cold shock domain (3,10). These hits were reformatted using the esl-reformat command in HMMER under default settings before a multiple sequence alignment was build using MAFFT as described above. For the EF-Tu phylogeny, no additional modifications before multiple sequence alignment were necessary. Maximum-likelihood phylogenetic trees were built from the multiple sequence alignments without trimming gaps using IQ-TREE2 (v2.1.2) with the following parameters: -st AA -bb 1000 -mtree -nt 12 -safe (11).

Classification of cold shock proteins and subsequent selection tests

To assess selection of orthologous CSP, all loci were grouped into equivalent types. First, all epitope containing hits with annotations including the key words 'cold, shock, and/or antiterminator' were filtered into a fasta file in a genera dependent manner using a custom R script (R script #21, 23). The collected CSPs for each genus were that ran through three pipelines. One, to determine the number of CSP clusters, full length sequences were run through mmseq2 (v13.45111) using the easy-cluster command withing the following parameters: clusterRes tmp –min-seq-id 0.5 -c 0.8 –cov-mode 1 (12). Two, full length sequences were passed through MEME Suite Motif Discovery (v5.5.1) under the following parameters: Classic mode, any number of repetitions (anr), and the number of motifs selected based on the number of clusters outputted by mmseq2 and manual inspection (13). The XML output was downloaded for each genera's CSP loci and processed via the R package ggmotif (v0.2.1) (14,15). Third, the CSP domain was extracted via HMMER (v3.3.2) using hmmsearch with the following parameters, -E 1 --domE 1 --incE 0.01 --incdomE 0.04, and the hmm model specific for cold shock domain can be download (3,10). The CSP domains from each genus were aligned using MAFFT using –auto and a phylogenetic tree was built using FastTree2 (9,16). Information extracted from MEME and mmseq2 analysis was plotted onto the phylogenetic tree to cross reference the number of CSP types identified (R script #24). The R packages phangorn (v2.7.1), treeio (v1.14.4), ggtree (v3.1.2.991), ggnewscale (v0.4.5), and ggtreeExtra (v1.0.4) were used to parse and plot the treefile or newick file (17-20). Details on parameters for plotting can be found in the custom R script #14 and 22.

To confirm within-genera grouped CSPs were the same ortholog, an all-by-all BlastP search was performed using the orthologr R package (v0.4.0) (R script #25) (21). The coding sequences of associated orthologs CSP within a genus were extracted for selection analyses via dn/ds ratio. MACSE (v. 2.07) with the parameters "-prog alignSequences" was used to generate codon alignments from nucleotide gene sequences (22). IQ-TREE (v.1.6.12) with the default parameters was used to generate phylogenies for each dataset (23). HYPHY (v.2.5.48) with the parameter "cln" was used to remove stop codons from alignments (24). AliView (v.1.26) was used to manually inspect alignments and remove gaps present in >75% of sequences (25). The HYPHY algorithm FUBAR, with the codon alignment and gene phylogeny as input and a posterior probability threshold of 0.9, was used to identify sites under purifying or positive selection (24).

Plant growth conditions

Arabidopsis Col-0 and *efr*-1 transgenics were grown in Sunshine Mix soil in chambers at 10/12-h light/dark cycles, 23°C, about 100 µmol/m^2/s light intensity, and at 70% humidity. For reactive oxygen species (ROS) assays, plants were grown to four to five-weeks (26). For MAPK induction, callose deposition and seedling growth inhibition assays, *Arabidopsis thaliana* Col-0 and *efr* seeds were surface sterilized with

3

bleach and then stratified for 2-3 days at 4ºC. Seeds were then germinated on one-half-strength concentration Murashige and Skoog medium (½ MS) plates containing 0.8% plant agar and 1% sucrose in a growth chamber with 16-hrs light, 8-hrs dark at 22ºC. Tomato *Solanum lycopersicum* cultivar Rio Grande PtoR (*Pto/Pto, Prf/Prf*) and CRISPR *core* lines were sown in agronomy mix in growth chambers under 16/8 hr light/dark cycles, 24°C, about 200 µmol/m^2/s light intensity, 50 to 65% humidity, and grown to six to seven-weeks old.

Measuring ROS Production and Antagonism

For measuring ROS, leaf punches were taken using cork borer no. 1 (diameter of 4 mm) from equivalent leaves or leaflets from the cotyledon leaves across biological replicates (for csp22, the 5th leaflet). Leaf punches (4 per plant) were floated on their abaxial side in 190 µl of sterile water in a 96-well white microtiter plate for 18 to 24 hours in the dark. To test ROS elicitation, sterile water was removed and a 100 µl solution of sterile water, L-012 (Wako Chemicals, Cat. No. 143556-24-5), horseradish peroxidase (HRP) (Sigma Aldridge, Cat. No. P6782), and 100 nM to 1 µM of MAMP peptide dissolved in sterile water or 100% DMSO was added (see SI Appendix Table S2 for solvent used). For elf18 (acyl-MSKEKFERTKPHVNVGTI) and flg22 (QRLSTGSRINSAKDDAAGLQIA) tested on *Arabidopsis,* 20 µM L-012 and 20 ug/mL HPR was used (27). For csp22 tested on tomato, 40 µg/mL HPR was used. For MAMPs which were dissolved in 100% DMSO, an equivalent volume of 100% DMSO was added to the controls. Assay measurements were taken every 1.5 minutes for a minimum 60 minutes. Luminescence was measured using a TriStar LB 941 plate reader (Berthold Technologies). Across all ROS experiments, the maximum RLU was determined for each leaf punch and averaged between punches per plant.

Peptide antagonism: For antagonist assays, the leaf disks were incubated in a solution of water and candidate antagonist MAMP peptide (100 nM to 1 µM) for 18 to 24 hours in the dark. The candidate antagonist MAMP peptide was then removed and the remainder of the assay including elicitation and luminescence measurements are the same as above. Recombinant protein antagonism: Recombinant protein concentrations were quantified and between 100 nM and 500 nM of each protein was added in solution with HRP and L-012. The solution was immediately used for elicitation. Lysate antagonism: Initial protein concentrations were quantified and between 1 and 2.5 µg/mL was added to HPR and L-012 solution and immediately used for elicitation. For all ROS assays, a minimum 12 plants were used. To compare ROS plates, plants and punches per treatment were averaged and scaled based on the appropriate controls.

Measuring alkalinization and antagonism in Tobacco BY-2 Cells

The tobacco cell line BY-2 (*Nicotiana tabacum* L. cv. Bright Yellow 2) was culture on a modified MS medium supplemented with 100 µg/mL timentin (32). Callus tissue was passaged every three to four weeks on MS medium with 2% phytoagar. Transplanted calli in petri dishes were wrapped in parafilm and incubated at 25-26°C in the dark. To establish suspension cultures, a 1.5x1.5 cm slice of calli was transferred to 100 mL of modified MS media in a 250 mL flask, covered with autoclavable paper, sealed with a rubber band and wrapped in foil. The flask was incubated for 10 to 14 days in a 26°C shaker rotating at 130 rpm in the dark. Cultures were passed by transferring 10 mL of cells via a stereological pipette into ~90 mL of media. After three passages, multiple cultures were grown in parallel and were grown for experimentation.

To assess alkalinization changes due immune induction, cells were passage from the main line and grown in similar conditions for 5 to 7 days until cell density was about 3.0x10^6, which was quantified using a hemi-cytometer. Cells were then aliquoted into 20 mL glass flat bottom vials (Thermo Fisher Scientific, Cat. No. 6PCV20-1F) and sealed the tops with Millipore tap (Micropore, Cat. No. 1530-1). The cell vials were then placed on a shaker rotating at 100 rpm at room temperature at ~30° angle such that cells were in continuous motion. After two to three hours of incubation, the pH probe and meter were prepared for measurements similarly to previously established protocols (33). Between 10 nM of peptides were used to measure induced alkalinization and antagonism was measured by first treating with 50 nM candidate antagonist for three minutes, then treating with 10 nM agonist. The pH change was sampled every 10 seconds for up to 30 minutes.

58

Plant tissue-based immune assays

For measuring MAPK induction, *Arabidopsis thaliana* Col-0 and *efr* seeds were germinated on ½ MS plates for five days. The seedlings were then transferred to a 48-well plate (Costar, Cat. No. CLS3548) containing 600 µL of liquid ½ MS medium. Plates were sealed with 3M Micropore Surgical Tape (Micropore, Cat. No. 1530-1) and transferred back to the growth chamber. After nine days of growth, the 14-days-old seedlings were incubated with water or 100 nM of elf18 variants for 0 and 15 min before being pooled for harvest. Three seedlings per treatment were pooled. Seedlings were ground to fine powder in liquid nitrogen and resuspended in 200 µL of extraction buffer 50 mm HEPES (pH 7.5), 50 mm NaCl, 10 mm EDTA, 0.2% Triton X-100, 1× Pierce protease inhibitor mini tablets (Thermo Scientific, Cat. No. A32955), and 1x Pierce phosphatase inhibitor mini tablets (Thermo Scientific, Cat. No. A32957). Total proteins were isolated by centrifugation at 21,000 × g at 4 °C for 10 min and Laemmli buffer was added for subsequent SDS–PAGE. To measure the protein concentration, we used Pierce 660 nm Protein Assay (Thermo Scientific, Cat. No. 22660) with Ionic Detergent Compatibility Reagent (Thermo Scientific, Cat. No. 22663). Equal concentration of protein was loaded per lane and separated by SDS–PAGE. Activated MAPK3 and MAPK6 were detected by immunoblotting using p44/42 MAPK (pERK) antibodies at a concentration of 1:2000 (Cell Signaling Technology, Cat. No. 4370), followed by secondary rabbit antibodies at a concentration of 1:3000 (Biorad, Cat. No. 1705046). Both primary and secondary antibody were suspended in 1x TBS-T with 5% BSA. The experiment was repeated three times with similar results.

For analysis of callose deposition, *Arabidopsis thaliana* Col-0 and *efr* seeds were germinated on ½ MS plates for four days. The seedlings were then transferred to a 48-well plate (Costar, Cat. No. CLS3548) containing 500 µL of liquid ½ MS medium, sealed with 3M Micropore Surgical Tape (Micropore, Cat. No. 1530-1) and transferred back to the growth chamber. Three seedlings were transferred per treatment. After four days of growth, the ½ MS medium was replaced with water or 1 µM of elf18 variants and the seedlings were incubated in a shaker for 18 to 20 h in the growth chamber. Seedlings were fixed in ethanol/acetic acid (3:1, v/v) at 37ºC for about 2 to 3 hours or overnight at room temperature (until cotyledons were completely cleared), with one change of fixing solution. Seedlings were subsequently rehydrated in 70%, 50% and 30% ethanol for 30 min each step. Then the seedlings were washed one time with 150mM K_2HPO_4, pH 9.5 and stained with 0.01% (w/v) aniline blue in 150mM K_2HPO_4, pH 9.5 for 1 h in dark. The cotyledons were loaded onto the slides with 50% glycerol and callose deposition was imaged by fluorescence microscopy (Leica CMS) using a DAPI filter. Callose quantification was performed with Fiji as previously described (28,29). Experiments were repeated three times with similar results.

To assess seedling growth inhibition (SGI) we followed Gómez-Gómez et al., 1999 with modifications (30). *Arabidopsis thaliana* Col-0 and *efr* seeds were germinated on ½ MS plates for four days. The seedlings were then transferred to a 48-well plate (Costar, Cat. No. CLS3548) containing 500 µL of liquid ½ MS medium supplemented with 100 nM of elf18 variants. Plates were sealed with 3M Micropore Surgical Tape (Micropore, Cat. No. 1530-1) and transferred back to the growth chamber. After eight days of growth, the seedlings were patted dry and weighed. Each experiment included eight seedlings per treatment. A mock control (½ MS medium) and canonical active elf18 sequence (acyl-SKEKFERTKPHVNVGTIG) were assayed on each plate. Experiments were repeated three times with similar results.

Ethylene induction was measured as previously described (31). For measurements of ethylene, each dot represents one measurement (from one tube with water and three leaf pieces); four replicates per treatment. Leaf pieces (squares of approx. 2 x 2-3 mm) were from two individual plants, four- to five-weeks-old, cut the evening before, and incubated on water overnight and induced using 1 µM csp22 variants. Ethylene induction was measured using a Shimadzu Gas Chromatograph GC-14.

Development of tomato core mutants

To generate the *core* mutants in the tomato (*Solanum lycopersicum*) cultivar Rio Grande (RG)-PtoR, we designed two guide RNAs (gRNA1: 5'- GTAGCATTTGACAATGTCCC-3'; gRNA2: 5'-GACTGGCCTGGGGTCTCATG-3') that target the first exon of *CORE* using the software Geneious R11. The gRNA cassette was cloned into the p201N:Cas9 binary vector as described previously (34,35). Tomato transformation was performed at the Biotechnology Center at the Boyce Thompson Institute as described previously (36). Mutations were confirmed by Sanger sequencing at the Biotechnology Resource Center (BRC) at Cornell University.

Bacterial Strains and Growth conditions

Clavibacter michiganensis strains were grown in tryptone broth with yeast (TBY) medium at 28°C, *Escherichia coli* was grown in Luria broth (LB) medium at 37°C, *Streptomyces spp.* ND05-3B was grown on yeast malt extract (YME) at 28°C, and *Pseudomonas syringae* pv. *tomato* strains were grown in (NYGB) medium at 28°C. The following concentrations were used for antibiotic selection: 50 μg/mL gentamycin, 25 μg/mL kanamycin, 25 μg/mL chloramphenicol, 50 μg/mL spectinomycin, and 100 μg/mL rifamycin.

Assessing syntenic gene structure in *Clavibacter*

Seven genomes across five species in the genome were assessed for their gene structure surrounding the CSP loci identified in the MAMP mining pipeline. Those selected included *Clavibacter michiganensis* NCPPB382 (Accession: GCF_000063485.1), *C. michiganensis* Z001 (Accession: GCF_002931335.1), *C. capsici* PF008 (Accession: GCF_001280205.1), *C. insidiosus* ATCC10253 (Accession: GCF_003076355.1), *C. nebraskensis* NCPPB2581 (Accession: GCF_000355695.1), and *C. tessellarius* ATCC33566 and CFBP8017, respectively (Accession: GCF_002240635.1 and GCF_002151185).[1] Briefly, the region within each genome was identified via a blast search and regions of interest were extracted via a custom Biopython script. The extracted regions in Genbank format were fed into clinker, a gene clustering program (v0.0.27) (37). The outputs clustered gene structures were output as an SVG file and loci color was manually edited in Inkscape (v1.0.2).

Measuring expression of CSPs in Clavibacter michiganensis

To assess if all CSP loci are expressed, we cultured the type-strain Gram-positive pathogen *C. michiganensis* NCPPB382 *in vitro* in both rich TBY medium and minimal m9 medium supplemented with xylem sap, which mimics the xylem vasculature (38). To collect xylem sap, tomatoes were grown for 6 weeks in chamber conditions described above and a similar procedure was followed (39). Briefly, using scissors sterilized with 70% ethanol, the stem was cut about 10-12 cm above the cotyledons and additional leaflets were removed. Wounds were sealed with parafilm. After re-sterilizing the scissors, a P1000 tip was cut such that the diameter of the tip can go on top the stem. The tips were parafilmed to the stem and a small bag placed over the tip. Over the next two hours, xylem sap was collected. The collected sap was filter sterilized after passing through a 0.22 μM filter.

C. michiganensis NCPPB382 was grown on TBY agar plates. After three days, the plates were scraped, and the bacteria transferred to 25 mL flasks of TBY broth for a starter culture. The starter cultures were grown at 28°C for ~20 hours and used to inoculate growth curves. Two flasks of m9 broth with xylem sap were inoculated with *C. michiganensis* NCPPB382 to a starting OD_{600} of 1.0 and two flasks of TBY broth were inoculated with *C. michiganensis* NCPPB382 to a starting OD_{600} of 0.5. One mL of culture was removed from each flask at 6 hours and 24 hours of growth. The flasks were grown shaking at 28°C.

One mL samples were removed from cultures pelleted via centrifugation. 500 μL of supernatant was then removed and two volumes of RNA Protect Bacteria Reagent (QIAGEN, California, USA) was added. The cultures were resuspended, incubated at room temperature for five minutes, and pelleted again. They were then frozen with liquid nitrogen and transferred to a -80°C freezer until RNA extraction. RNA was extracted with the SV Total RNA isolation kit (Promega, Wisconsin, USA). Growth curves and RNA extraction were repeated three times. All PCR primers are listed in Table S3.

RNA was quantified by Qubit via a high sensitivity RNA kit (Thermo Fisher Scientific) and the concentration was standardized for each repeat. Random primers were annealed to the RNA and first-strand synthesis performed with M-MLV reverse transcriptase (Promega, Wisconsin, USA). For qPCR, cDNA was diluted 1:5 and SsoFast EvaGreen Supermix was used (BioRad, California, USA). *BipA* was used as a housekeeping gene (40). Each repeat was performed in a single qPCR plate. Except for rich media TBY broth at 6 hours, where only one sample was used due to space constraints, two samples were used for each broth and time point. qPCR measurements were detected on a Bio-Rad CFX96 Real-Time PCR Detection System.

Collection of bacterial lysates

An initial starter 20 mL culture of *C. michiganensis* NCPPB382 was incubated at 28°C at 200 rpm until the optical density (OD_{600}) equals one. A subculture was made in 300 mL TBY medium and cultured in similar conditions until the optical density (OD_{600}) equaled two. *Pseudomonas syringae* pv. *tomato* DC3000Δ*flaA* was cultured similarly in NYGB. For *Streptomyces spp.* ND05-3B, spores were populated on YME medium over four to five days at 28°C (41). Using a cell scraper and sterile water, spores were

6

collected in 25 mL and quantified to a concentration of 9.05 x10^6 spores/mL via a hemacytometer. Cells were spun down at 5000 rpm and resuspended in lysis buffer (300 mM NaCl, 50 mM Hepes, pH 7.0) with 100 mM phenylmethylsulfonyl fluoride (PMSF) (Thermo Scientific) and 0.2 U/mL DnaseI (Thermo Scientific). The chilled cells were sheered via an Emusiveflex-C3 High-Pressure Homogenizer (Avestin, Ottawa, Ontario, Canada) until the solution turned clear. The collected lysates were passed through a 0.22-micron filter, froze with liquid nitrogen and store in the -80°C until used. To test induction of immunity by ROS production, protein lysates were quantified using a Pierces 660 assay using BSA as a standard control and tested using the same procedure as describe in Plant tissue-based immune assays at 1 µg/mL concentration.

Expression and Purification of Clavibacter CSPs

The coding sequence of each CSP from NCPPB382 was codon-optimized for *E. coli* and synthesized into entry vector, pENTRY, from either ThermoFisher Scientific or Twist Biosciences (Waltham, Massachusetts or San Francisco, California, USA). Each vector was transformed into *E. coli* DH5a and after subsequent colony isolation and plasmid miniprep, each vector sequence was confirmed via whole plasmid sequencing (Plasmidsaurus, Eugene, Oregon, USA). Each CSP sequence was then transferred into expression vector pRSET via LR clonase using manufacturer's instructions (Thermo Scientific). The subsequent reactions were transformed into *E. coli* DH5a and putative transformants were checked via PCR for the insert and the sequence was confirmed via whole plasmid sequencing. Subsequent CSP expression vectors were transformed into *E. coli* expression genotype BL21(DE3).

For protein expression, confirmed vector-carrying transformants were grown in 50 mL LB media at 37°C until OD$_{600}$ at least 1. The initial culture was used to seed a larger culture of LB medium, which was grown until OD$_{600}$ 0.4 and 0.6, at which a final concentration of 0.5 mM IPTG was added to each culture and shaken at 200 rpms at 28°C for four hours. Bacterial cells were pelleted by centrifugation at 4°C and resuspended in buffer A (20 mM Tris, 10 mM Imidazole, 500 mM NaCl, pH 8.0) with 1 mM PMSF, 10 uM leupeptin serine protease inhibitor, and 10 ug/mL lysozyme. Cell suspension was incubated on ice for 30 minutes, then sonicated in four 10 second cycles at 30% output and 30% duty cycle with 15 second breaks between each cycle. Cells were centrifuged at 5,000 rpm for 20 minutes at 4°C. The supernatant was incubated with prepped Ni-NTA Resin (Thermo Fisher Scientific Cat. No. 88222), previously washed 3x with buffer A, for one hour at 4°C on a rotating shaker. The protein-bound beads were centrifuged at 5,000 rpm for 10 minutes at 4°C, washed 3x with buffer A, then incubated with elution buffer B (20 mM Tris, 500 mM NaCl, 5% Glycerol, pH 8.0). Each incubating elution set had increasing concentrations of Imidazole from 50 to 250 mM. The eluted protein was collected after the beads were centrifuged at 5,000 rpm for 10 minutes at 4°C. At each step from cell fraction to elution, sample was collected and prepped to run on a 15% SDS-Page gel. The gel was either stained with Coomassie SimpleBlue SafeStain (Thermo Fisher Scientific, Cat. No. LC6060) or transferred to a PVDF membrane and blotted with 1:2000 anti-6xHis-HRP antibody in 5% BSA-TBST (Thermo Fisher Scientific, Cat. No. MA1-21315-HRP). Purified recombinant protein was desalted via 3.5K MWCO dialysis cassettes (Thermo Fisher Scientific, Cat. No. A52966) in buffer (20 mM Tris, 50 mM NaCl, 5% glycerol, pH 8.0) overnight at 4°C and concentration quantified via Qubit protein assay kit (Thermo Fisher Scientific, Cat. No. Q33211).

Overexpression of cspB and transformation into Pst via Tri-parental mating

A *Pst* DC3000 codon-optimized gBlock of *Cm* NCPPB382 *cspB* was synthesized (Azenta, USA) and amplified via PCR with primers carrying restriction sites NdeI and EcoRI (Thermo Fisher Scientific). Both board host range plasmid pDSK519-ngfp and amplified fragment were digested with *Nde*I and *Eco*RI and ligated using T4 ligase (Thermo Fisher Scientific) (42). Standard DH5a *E. coli* transformation was performed and plasmid was confirmed via whole plasmid sequencing (Plasmidsaurus). For pDSK519-*cspB*, an additional restriction with *Bsi*WI and ligation was performed to move the chloramphenicol antibiotic cassette from a Gateway cassette into pDSK519.

Cultures of donor vectors, helper plasmids (pRK2013/pRK600), and recipient strains were grown overnight in no antibiotics (43,44). Cultures were mixed in the following ratio (1 Donor: 1 Helper: 5 Recipient) and spot plated onto a NYGB plate with no antibiotics and incubated in 28°C. After two days, spots were re-streaked onto NYGB plates which contained antibiotics for selection of recipient strain with donor plasmid and grown overnight. Transformants were confirmed for the donor plasmid via standard gDNA preparation and PCR screening.

Overexpression of *cspB* was confirmed via in vitro growth in NYGB media, protein lysates were separated using standard 15% SDS-Page Gel, probed using an anti-HA-HRP conjugated antibody (Cat. No. 26183-HRP, Thermo Fisher Scientific). In detail, cultures were grown to an OD of 1.0, cells were spun down and resuspended in water and 3x Laemmli buffer supplemented with DDT. Cells incubated at 95°C for five minutes, briefly put on ice and spun at max speed. Proteins samples were run on an 15% SDS-PAGE Gel, transferred to a nitrocellulose membrane via semi-dry, and blotted with 1:1000 anti-HA-HRP conjugated antibody in blocking solution (Thermo Fisher Scientific). Chemiluminescence was detected using the Bio-Rad Chemidoc system. Protein loading for *in vitro* bacterial expression was visualized using the SilverQuest Silver Straining Kit (Thermo Fisher Scientific).

Disease assays and bacterial titers of *P. syringae* in tomato

Tomato Rio Grande (RG)-PtoR and mutant line *core* were grown to about seven-weeks of age just before flowering. Three days before inoculation, strains *Pto* DC3000ΔΔΔ carrying empty vector pDSK519-ngfp or constitutive expressing vector pDSK519-CO-*cspB*-HA (pDSK519-cspB) were streaked from glycerol stocks and cultured on NYGB plates with appropriate antibiotics. Cells were scraped from the plate, resuspended in plain NYGB and an inoculum was adjusted to an $OD_{600} = 0.00001$ (or $1x10^4$ CFUs/mL) by resuspension in 5 mM $MgCl_2$. Several leaves per plant were infiltrated with a needless syringe and a black sharpie was used to circle the infiltration area. After three days post inoculation, using a #3 cork borer, a single punch was taken from each plant with four plants per genotype per treatment and moved to a 1.5 mL tube containing 200 µl 5 mM $MgCl_2$. The tissues were ground and dilutions were carried out between 10^{-2} and 10^{-5} and plated on NYGB plates containing ½ antibiotics and 50 µg/mL cycloheximide. After 2 days incubation at 28°C, colonies were counted and CFU per mg tissue was determined.
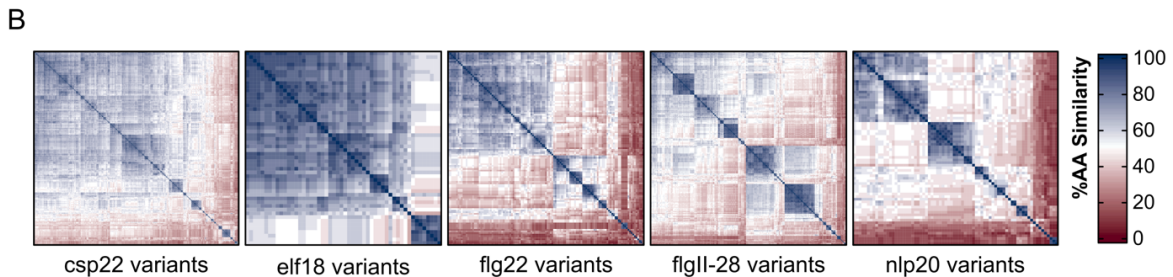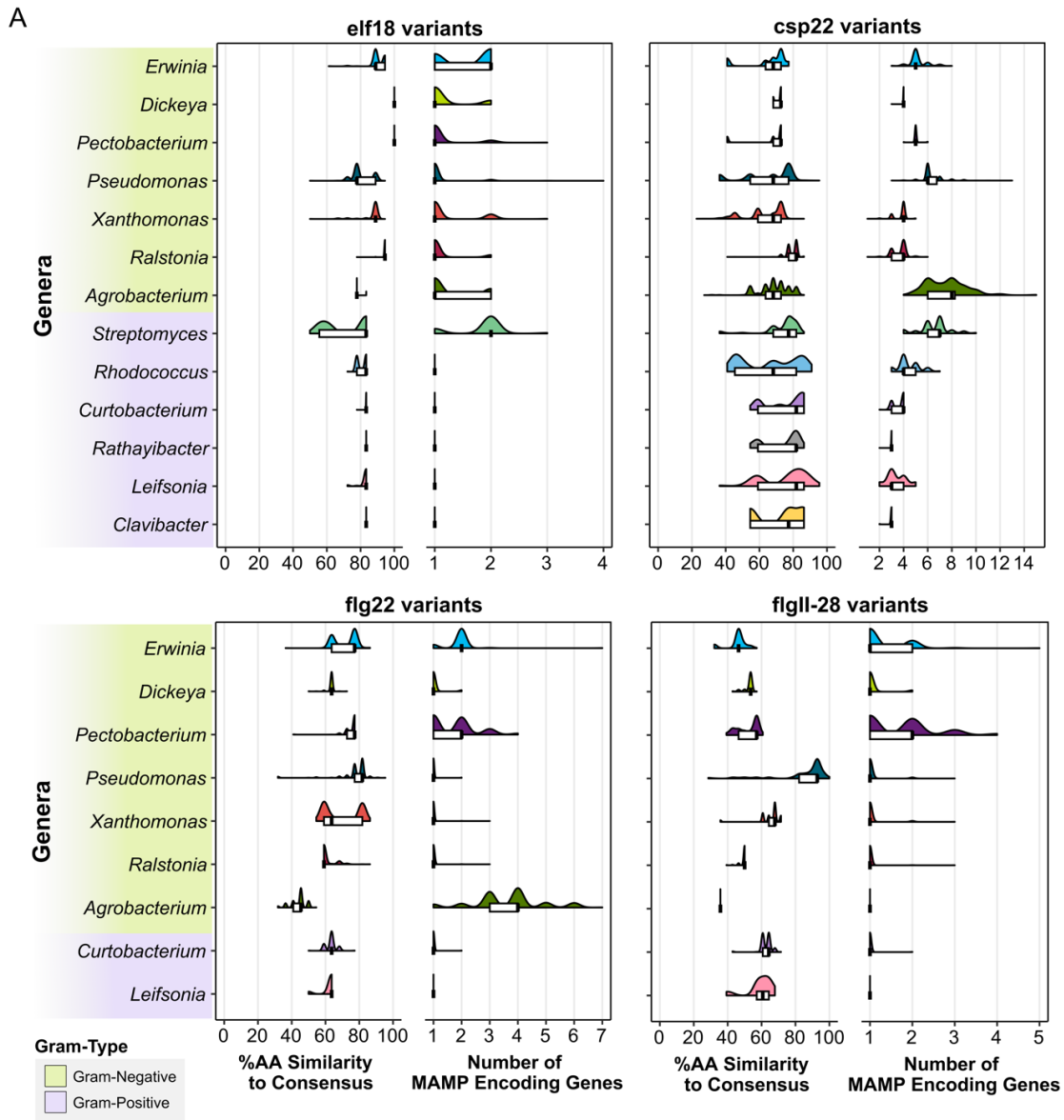
Fig. S1. Copy number and conservation differ across on immunogenic features and bacterial genera. (A) Violin plots of percent amino acid (AA) similarity of epitope variants in comparison to each respective consensus sequence across all genera sampled. Tukey's boxplots are plotted on top. **(B)** All-by-all amino acid (AA) similarity comparisons for each MAMP independent of the consensus sequence.
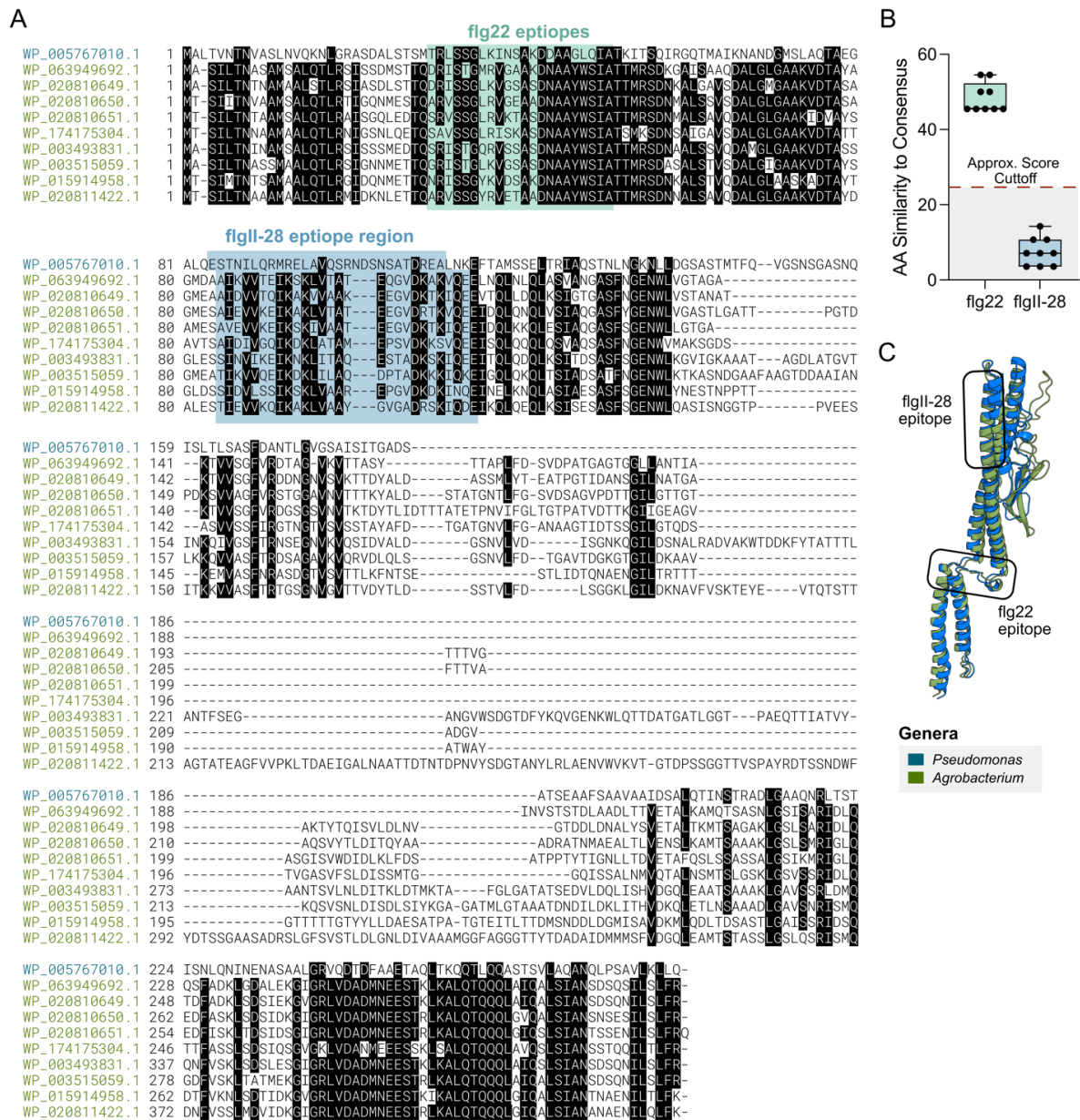
Fig. S2. Examples of divergent FliC proteins with undetectable flgII-28 epitopes. (A) Alignment of flagellin proteins from the reference *Pseudomonas syringae* pv. *tomato* DC3000 and *Agrobacterium,* which carry a flg22 epitope but fail detection for flgII-28. **(B)** Percent amino acid (AA) similarity of epitope variants from (a) in comparison to each respective consensus sequence. **(C)** Structural modeling of flagellin from *Pto* DC3000 (blue = WP_005767010.1) and *Agrobacterium* (green = WP_020810649.1). RMSD = 1.579.

10
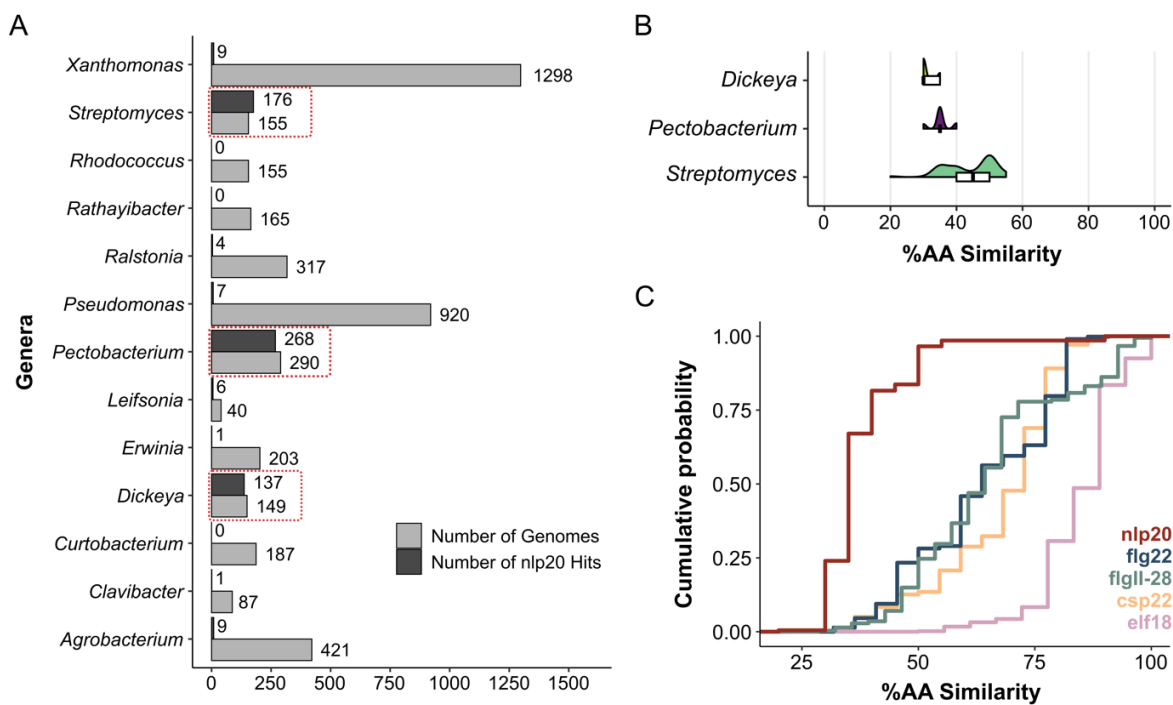
Fig. S3. Nlp20 is present but not conserved across diverse bacteria. (A) The number of nlp20 hits in respect to the number of genomes analyzed. Red boxes highlight genera where nlp20 MAMPs are abundantly conserved. **(B)** Violin plot with Tukey's boxplot on top of percent AA similarity of nlp20 epitopes in comparison to the consensus sequence across the three major genera's nlp20 are found. **(C)** Cumulative probability of Fig. 1C in respect to percent amino acid (AA) similarity.
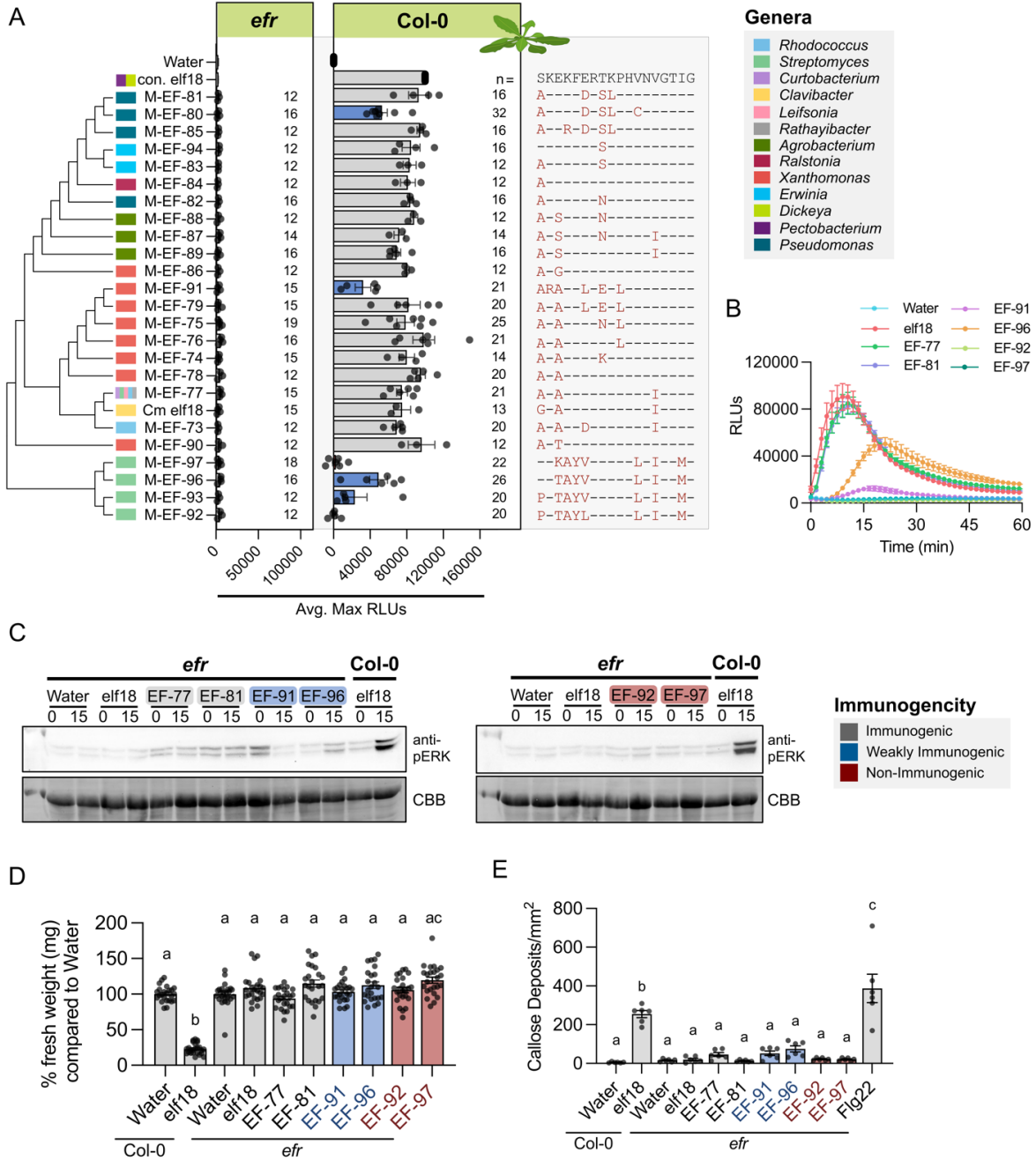
Fig. S4. Perception and specificity of elf18 variants. (A) Left: Cladogram based on an alignment of elf18 epitope sequences with tips labeled by genera. Middle: ROS induction of elf18 variants (100 nM concentration) in *Arabidopsis* Col-0 and the *efr* mutant line. Each data point represents an average max reactive light unit (RLU) from 4 plants with 4 leaf disks per plant. The number of plants sampled are plotted to the right. For induction in Col-0, values of each plate were adjusted to a scale of 0 to 100000 maximum RLUs based on the controls (water and consensus elf18; also referred to as con. elf18 and elf18). A one-way ANOVA and Tukey's mean comparison was used to determine significance on non-scaled data, $p < 0.05$

12

66

(grey = not significantly different from the consensus; blue = significantly less than the consensus but higher than water; red = not significantly different from water). Across all plates, water and consensus elf18 were tested alongside *efr* transgenic plants. Error bars = SEM. Right: Alignment of epitopes tested with changes from the consensus noted in red. **(B)** Independent analysis of ROS induction of a subset of elf18 variants. Error bars = SEM. Two plants per plate across six plates was conducted. All peptides were tested at 100 nM. **(C)** Induction of MAPK by elf18 variants (100 nM) in *Arabidopsis* Col-0 and the *efr* mutant at zero- and 15-minutes post infiltration. CBB = protein loading. **(D)** *Arabidopsis thaliana* seedling growth inhibition by either water (mock) or elf18 variants (100 nM). One point = 1 plant. Eight plants per biological replicate. The experiment was complete three times. A one-way ANOVA and Tukey's mean comparison was used to determine significance. **(E)** Quantified deposition of callose in *Arabidopsis* Col-0 and *efr* mutant by water and elf18 variants (1 μM). Values are from one representative experiment which includes an average of at least two images per leaf from two leaves per plant, three plants per treatment. The experiment was repeated three times with similar results. A one-way ANOVA and Tukey's mean comparison was used to determine significance for panel (D) and (E), $p < 0.0001$.
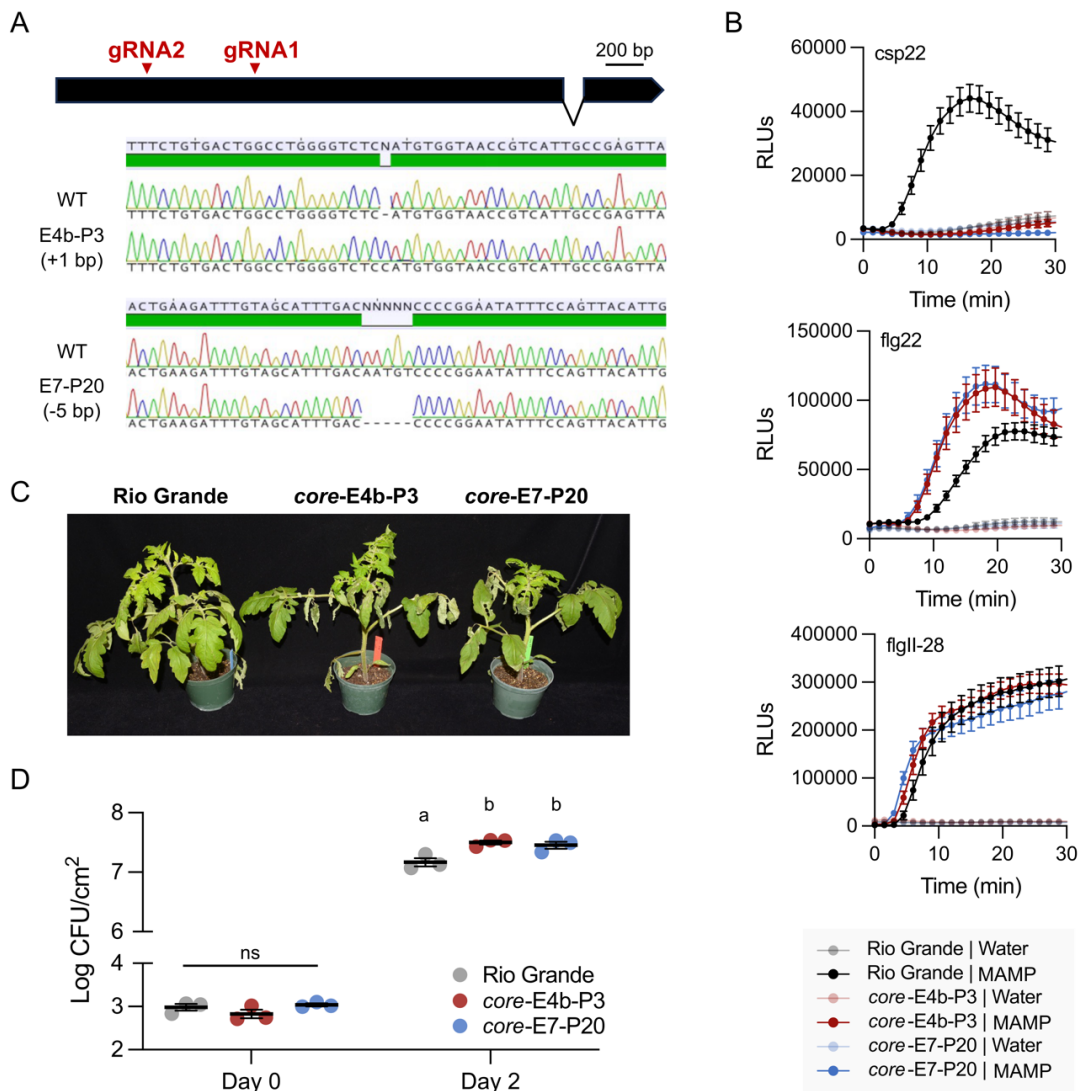
Fig. S5. Development of *core* mutant lines via CRISPR-cas9. (A) Gene structure of *CORE* (Solyc03g096190) with the region targeted for edits denoted in red. Bottom: Sanger sequencing reads of targeted edits in two independent lines compared to wild-type (WT) Rio Grande (RG)-PtoR. **(B)** ROS production in response to consensus MAMP peptides csp22 (200 nM), flg22 (100 nM), and flgII-28 (100 nM) in two mutant lines. **(C)** The *core* mutants were more susceptible to *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPto*Δ*avrPtoB*Δ*flaA* (DC3000ΔΔΔ). Four-week-old core mutants and wild-type RG-PtoR plants were vacuum infiltrated with 5 x 10^4 cfu/mL DC3000ΔΔΔ. Photographs of disease symptoms were taken five DPI. **(D)** Bacterial populations in leaves were measured at three hours (Day 0) and two days (Day 2) after infiltration. Error bars = SEM. Different letters indicate significant differences based on a one-way ANOVA followed by Student's t test ($p < 0.05$). ns = no significant difference. Three plants for each genotype were tested per experiment. The experiment was performed twice with similar results.
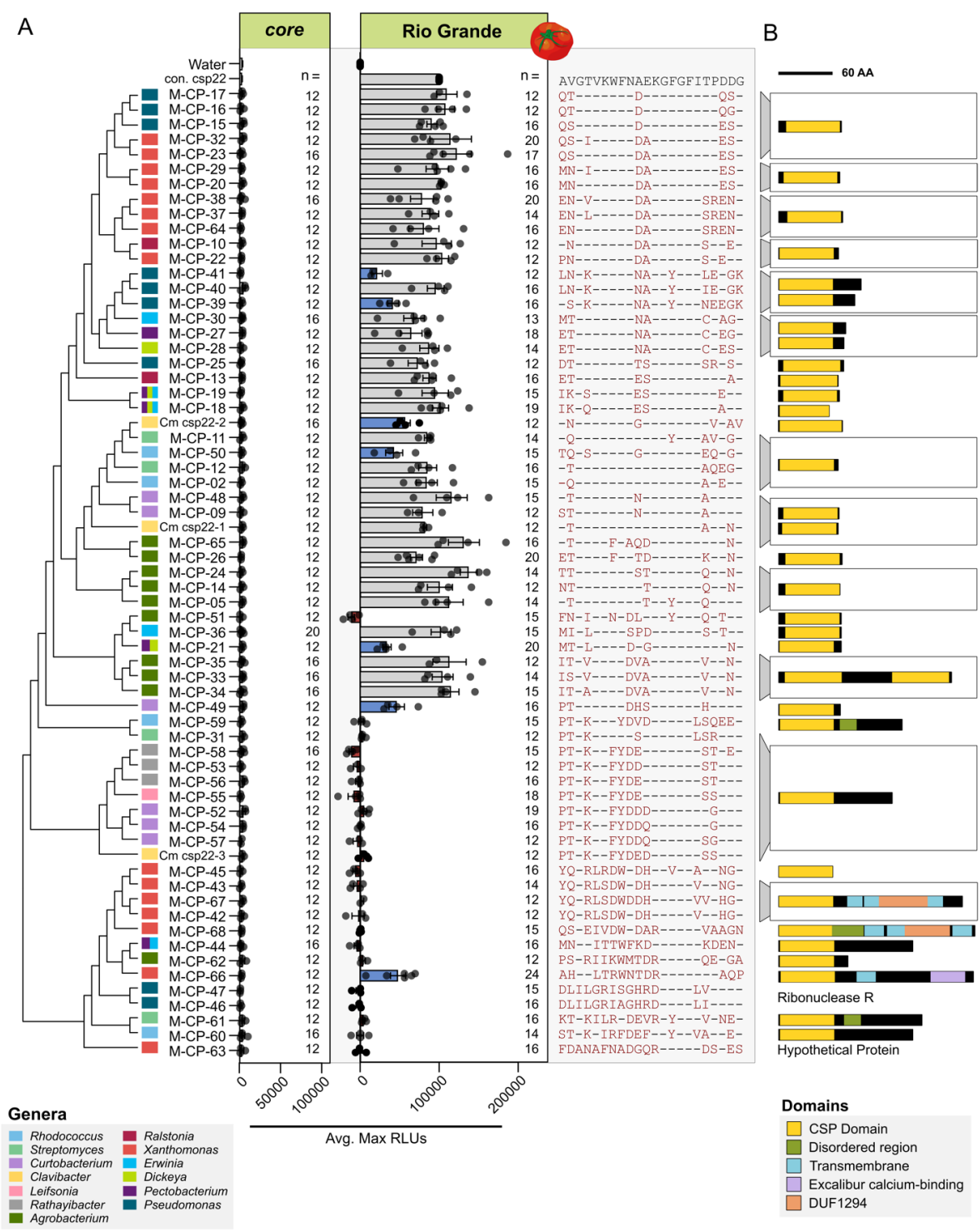
Fig. S6: csp22 variants exhibit differential immune induction in tomato. (A) Left: Cladogram based on alignment of epitope sequences with tips label by genera. Middle: ROS induction of csp22 variants (200 nM concentration) in tomato Rio Grande and a *core* mutant line. Each data point represents an average

max RLU from 4 plants with 4 leaf disks per plant. The number of plants sampled are plotted to the right. For immune induction in Rio Grande, values of each plate were adjusted to a scale of 0 to 100000 maximum RLUs based on the controls (water and consensus csp22/con. csp22). A one-way ANOVA and Tukey's mean comparison was used to determine significance on non-scaled data, $p < 0.05$ (grey = not significantly different from the consensus; blue = significantly less than the consensus but higher than water; red = not significantly different from water). Across all plates, water and con. csp22 were tested. Error bars = SEM. Right: Alignment of epitopes tested with changes from the consensus noted in red. **(B)** Domain structure of diverse CSPs derived from csp22 epitopes in (A) predicted using InterProScan. For CSPs with two domains, the epitope from the first domain was tested.
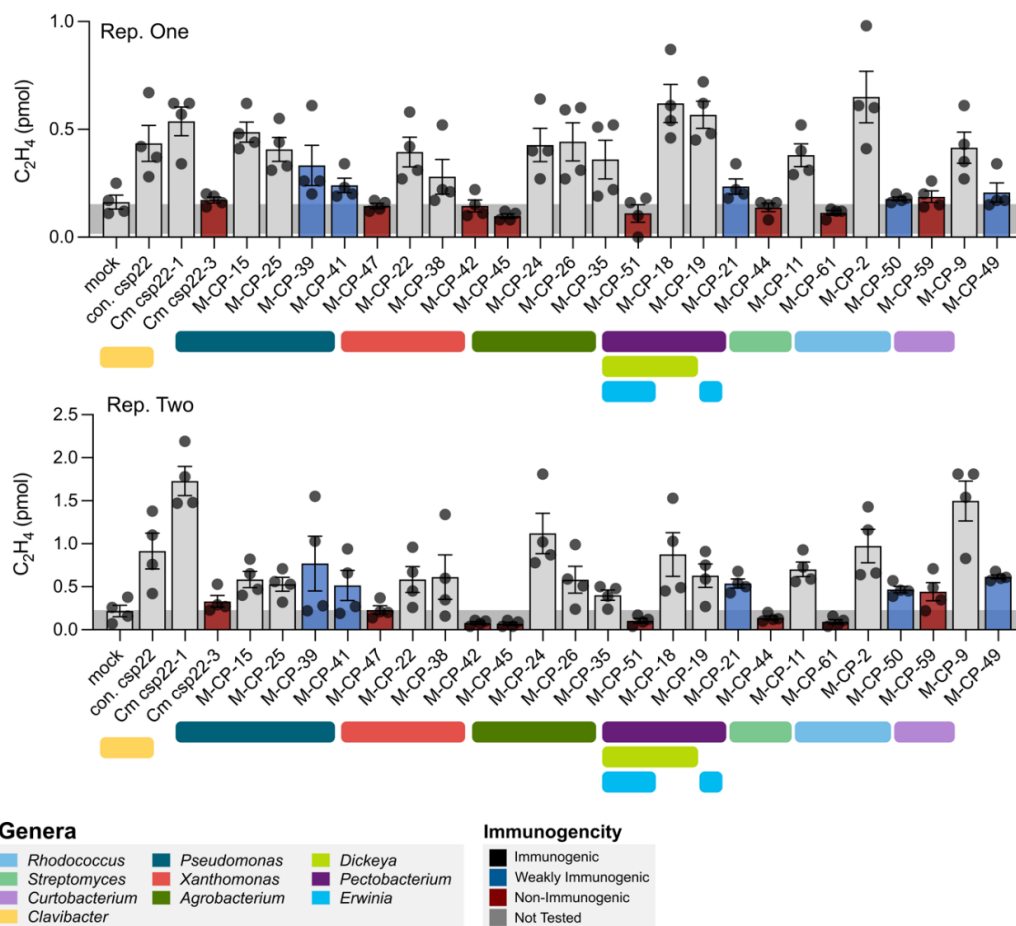
Fig. S7. Ethylene production induced by csp22 variants. Induction of ethylene by csp22 variants (1 µM) in 4- to 5-week-old tomatoes. Each dot represents one measurement (from one tube with water and 3 leaf pieces); four replicates per treatment. Bars are colored to match immunogenicity conclusions from ROS data in Figure 4A (grey = not significantly different from the consensus; blue = significantly less than the consensus but higher than water; red = not significantly different from water). Error bars = SEM.
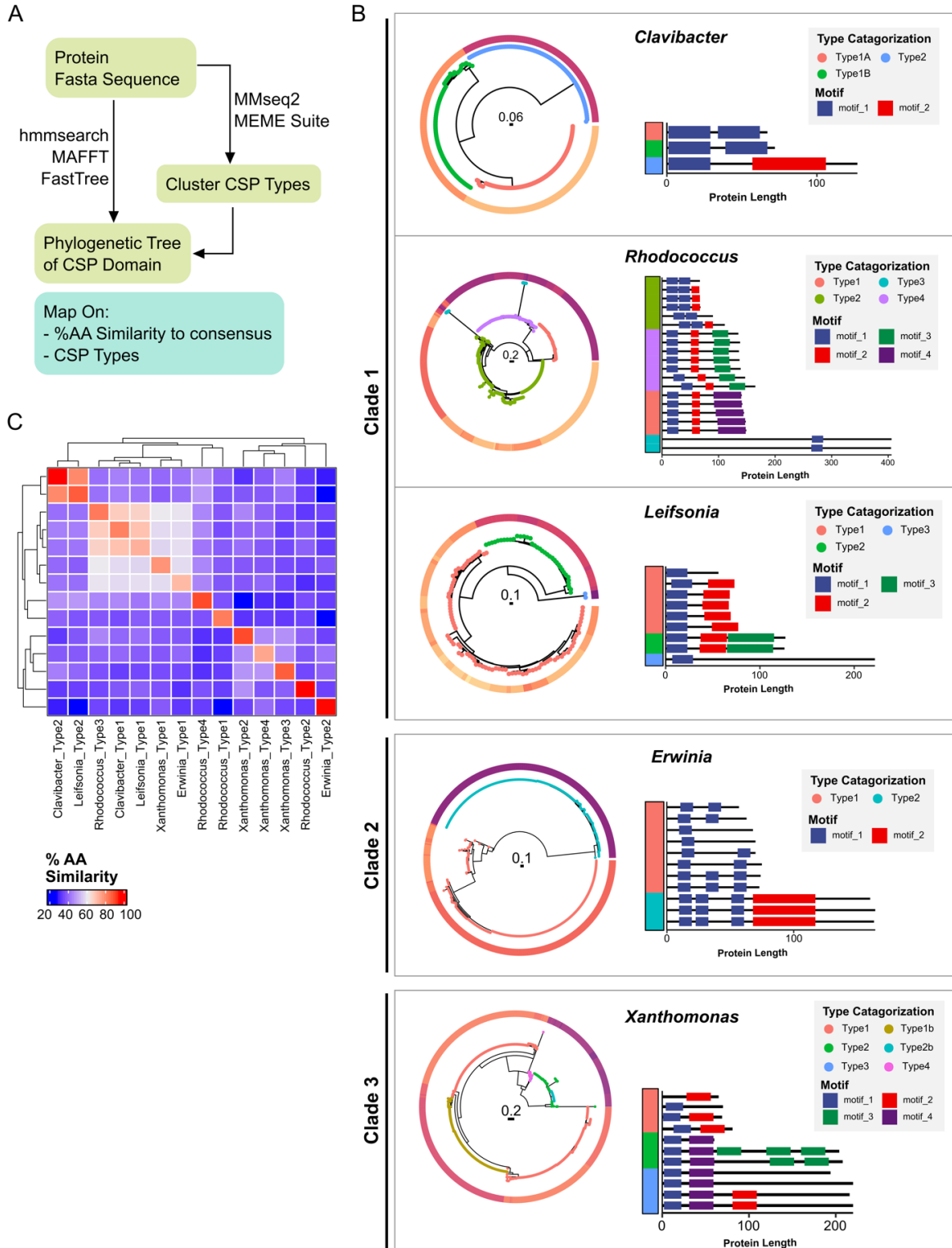
Fig. S8. Classification of CSP Types. (A) Pipeline for CSP classification. **(B)** Phylogenetic trees of cold shock proteins in different genera. Across all trees, tips are labeled by type categorization which is based

on clade structure, mmseq2 cluster classification, and motif structure predicted by MEME. The motif classification of represented cold shock proteins can be seen on the right of each tree and similarly labeled by the same type classification. Genera assessed include those found in clades highlighted in Figure 3. **(C)** BlastP assessment of CSP typing across genera. Those CSP types whose average similarity are above 70% are predicted as orthologs.
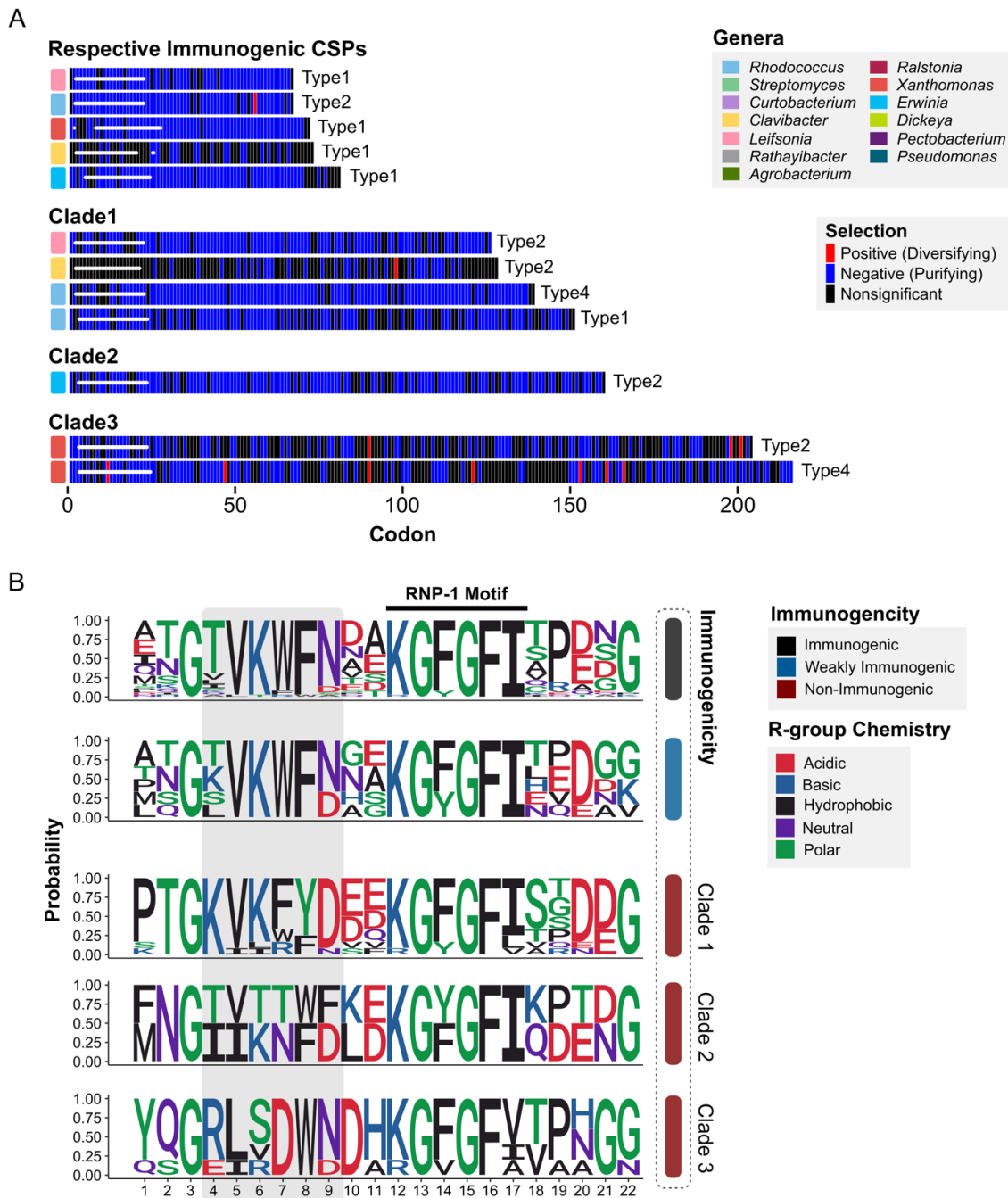
Fig. S9. Convergently evolved non-immunogenic CSPs are under purifying selection. (A) d_N/d_S tests of immunogenic (top) and non-immunogenic CSP types (Clades 1 - 3) across multiple bacterial genera. Sites which are significant by Bayesian approximation for selection are colored in blue (negative, purifying) or red (positive, diversifying). The region of the epitope is represented in the white bar. Typing of CSPs is described in Supplemental Figure 8. **(B)** Weblogos of immunogenic, weakly immunogenic, and non-immunogenic csp22 variants from each major clade tested by ROS and shown in Figure 3.
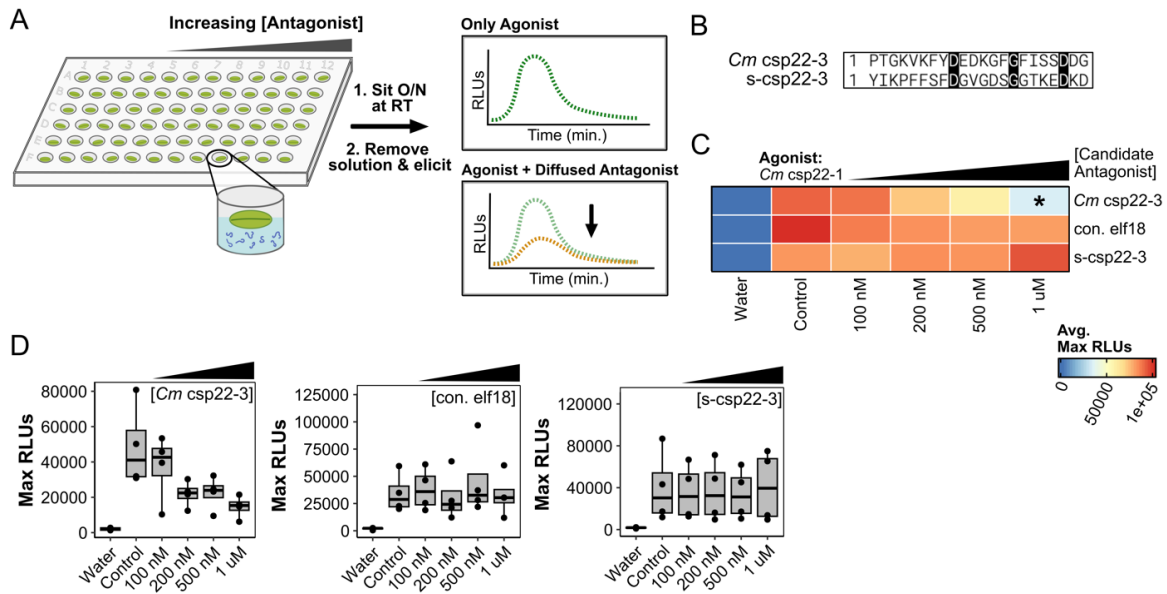
Fig. S10. Non-immunogenic *Cm* csp22-3 antagonizes CORE perception. (A) Diagram of the assay to assess MAMP antagonism. Concentrations for antagonism were 100 nM, 200 nM, 500 nM, and 1 μM (left to right). Immunogenic agonist was tested at 200 nM. All antagonism assays included four punches per plant, four plants for each experiment, and the experiment was repeated at least three times. **(B)** Alignment of *Cm* csp22-3 with a scrambled version of the peptide (s-csp22-3). **(C)** *Cm* csp22-3 assessed for antagonism against immunogenic *Cm* csp22-1. ROS screen for antagonism in Rio Grande tomato. Control denotes positive control (untreated agonist) and concentrations listed are of the candidate antagonist. Maximum RLU averages were adjusted to a scale of 0 to 100,000 based on the controls (water and untreated agonist). The assay was also repeated using con. elf18 and s-csp22-3. A one-way ANOVA and Tukey's mean comparison was used to determine significance in respect to the untreated agonist control (denoted *: $p < 0.05$). **(D)** Representative assays for (C).

21
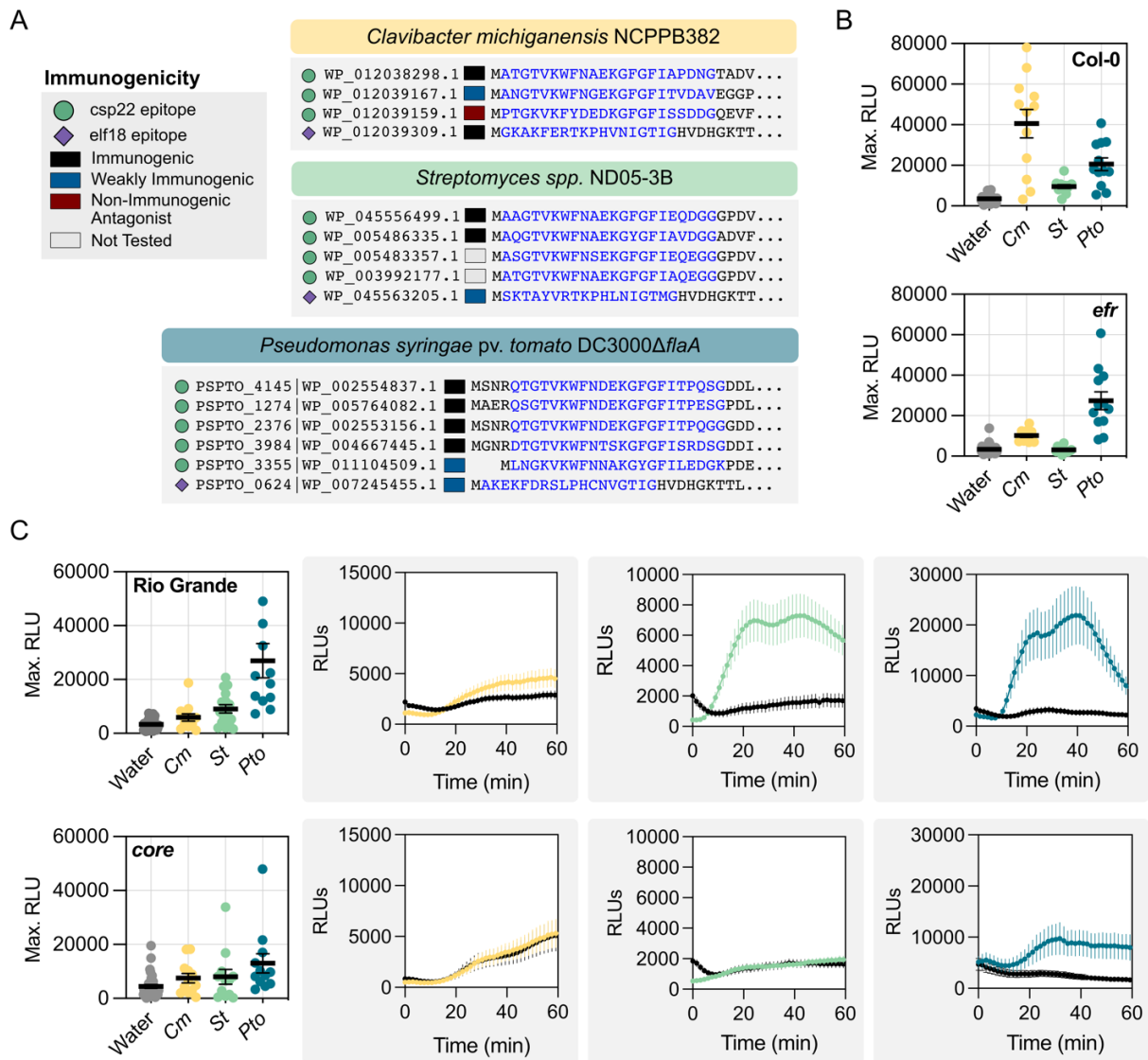
Fig. S11. MAMP perception from bacterial lysates on *Arabidopsis* and tomato. (A) Predicted csp22 and elf18 epitopes (highlighted in blue) encoded in bacterial genomes and their immunogenic outcomes based on the ROS screen. **(B and C)** ROS induction of bacterial lysates (1 μg/mL) in (B) *Arabidopsis* Col-0 and the *efr* mutant and (C) tomato Rio Grande and the *core* mutant. *Cm* = *C. michiganensis* NCPPB382, *St* = *Streptomyces spp.* ND05-3B, and *Pto* = *Pseudomonas syringae* pv. *tomato* DC3000Δ*flaA.* Each dot represents an average of the maximum RLUs of four disks per plant (n = 12). Representative plates shown on the right. Error bars = SEM.

22
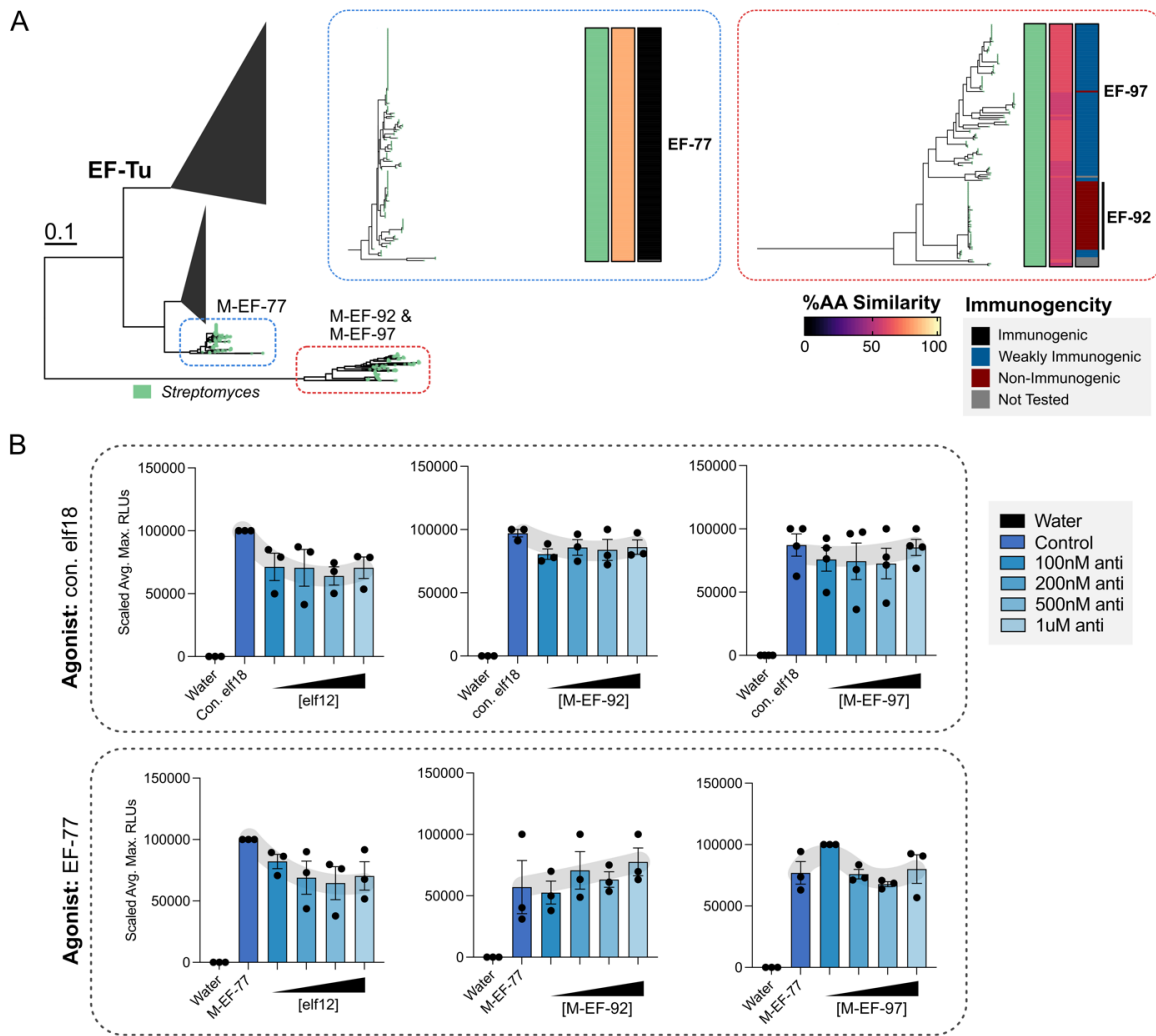
Fig. S12. No antagonism was identified for elf18. (A) Subset of the EF-Tu protein tree from Figure 2 highlighting those from *Streptomyces.* The immunogenic variant (M-EF-77) is found in the clade outlined in blue and the non-immunogenic variants (M-EF-92 and M-EF-97) are found in the clade outlined in red. **(B)** ROS screen for antagonism in *A. thaliana* Col-0. Control denotes positive control (untreated agonist) and concentrations listed are of the candidate antagonist tested. Maximum RLU averages were adjusted to a scale of 0 to 100000. Error bars = SEM. Elf12 has been previously reported as a weak antagonist.

Table S1: MAMPs detected from the computational pipeline in this study. Hits detected via BlastP and Local Alignment were combined. Duplicates were then removed before final filtering for partial proteins and off-targets based on the corresponding conserved motif.

MAMP	MAMPs Detected via BlastP	MAMPs Detected via Local Alignment	Total MAMPs Detected Post Filtering
Elf18	4660	3960	4404
Csp22	20665	19851	20542
Flg22	3217	4958	5042
FlgII-28	3697	3419	3661
Nlp20	124	636	613

Table S2. Custom Synthetic Peptides used in this study. Epitope sequence colored: green = immunogenic, blue = weakly immunogenic (predicted/confirmed deviant), orange = non-immunogenic (bolded, antagonist), not colored = not tested. Solvents colored: pink = water, yellow = 100% DMSO. Frequency = number of occurrences across all mined epitopes.

Reference Name	Sequence	Frequency	Modification	Purity	Solvent	Company	Notes
M-CP-1	ANGTVKWFNAEKGFGFITVDGG	175					Unable to be synthesized
M-CP-2	AQGTVKWFNAEKGFGFIAPEDG	151		95.22	Water		
M-CP-3	ANGTVKWFNAEKGYGFITVDGG	129					Unable to be synthesized
M-CP-4	ATGTVKWFNAEKGFGFIEQDGG	145					Unable to be synthesized
M-CP-5	ATGTVKWFNATKGYGFIQPDDG	240		96.35	Water		
M-CP-6	ATGTVKWFNDAKGFGFITPDEG	251					Unable to be synthesized
M-CP-7	ATGTVKWFNETKGFGFITPDGG	261					Unable to be synthesized
M-CP-8	PTGTVKWFNSEKGFGFIAPDDG	142					Unable to be synthesized
M-CP-9	STGTVKWFNNEKGFGFIAPDDG	109		95.11	Water		
M-CP-10	ANGTVKWFNDAKGFGFISPDEG	153		96.7	Water		
M-CP-11	AQGTVKWFNAEKGYGFIAVDGG	153		95.78	Water		
M-CP-12	ATGTVKWFNAEKGFGFIAQEGG	127		96.17	Water		
M-CP-13	ETGTVKWFNESKGFGFITPDAG	140		95.8	Water		
M-CP-14	NTGTVKWFNATKGFGFIQPDNG	219		96.5	Water	Biomatik, Cambridge, Ontario, Canada	
M-CP-15	QSGTVKWFNDEKGFGFITPESG	860	None	95.33	Water		
M-CP-16	QTGTVKWFNDEKGFGFITPQGG	852		95.69	Water		
M-CP-17	QTGTVKWFNDEKGFGFITPQSG	857		95.99	Water		
M-CP-18	IKGQVKWFNESKGFGFITPADG	578		95.88	Water		
M-CP-19	IKGSVKWFNESKGFGFITPEDG	596		95.51	Water		
M-CP-20	MNGTVKWFNDAKGFGFITPESG	127		95.35	Water		
M-CP-21	MTGLVKWFDAGKGFGFITPDNG	275		96.26	Water		
M-CP-22	PNGTVKWFNDAKGFGFISPEDG	1148		95.32	Water		
M-CP-23	QSGTVKWFNDAKGFGFITPESG	483		95.54	Water		
M-CP-24	TTGTVKWFNSTKGFGFIQPDNG	337		95.15	Water		
M-CP-25	DTGTVKWFNTSKGFGFISRDSG	876		96.21	Water		
M-CP-26	ETGTVKFFNTDKGFGFIKPDNG	245		96.1	Water		
M-CP-27	ETGTVKWFNNAKGFGFICPEGG	275		96.53	Water		
M-CP-28	ETGTVKWFNNAKGFGFICPESG	126		96.97	Water		
M-CP-29	MNGIVKWFNDAKGFGFITPESG	205		97.06	Water		
M-CP-30	MTGTVKWFNNAKGFGFICPAGG	132		96	Water		
M-CP-31	PTGKVKWFNSEKGFGFLSRDDG	144		95.56	Water		
M-CP-32	QSGIVKWFNDAKGFGFITPESG	101		95.01	Water		
M-CP-33	ISGVVKWFDVAKGFGFIVPDNG	130		95.38	DMSO		
M-CP-34	ITGAVKWFDVAKGFGFIVPDNG	124		95.56	DMSO		
M-CP-35	ITGVVKWFDVAKGFGFIVPDNG	444		95.37	DMSO		
M-CP-36	MIGLVKWFSPDKGFGFISPTDG	131		95.91	DMSO		
M-CP-37	ENGLVKWFNDAKGFGFISRENG	249		96.53	Water		
M-CP-38	ENGVVKWFNDAKGFGFISRENG	824		95.22	Water		
M-CP-39	ASGKVKWFNNAKGYGFINEEGK	115		95.39	Water		
M-CP-40	LNGKVKWFNNAKGYGFIIEDGK	118		95.93	Water		

M-CP-41	LNGKVKWFNNAKGYGFILEDGK	270		95.56	Water		
M-CP-42	YQGRLSDWNDHKGFGFVTPHGG	326		96.04	Water		
M-CP-43	**YQGRLSDWNDHKGFGFVTPNGG**	219		97.17	Water		
M-CP-44	MNGTITTWFKDKGFGFIKDENG	438		95.03	Water		
M-CP-45	YQGRLRDWNDHKGVGFATPNGG	198		95.11	Water		
M-CP-46	DLILGRIAGHRDGFGFLIPDDG	109		95.85	DMSO		
M-CP-47	DLILGRISGHRDGFGFLVPDDG	710		95.63	DMSO		
M-CP-48	ATGTVKWFNNEKGFGFIAPDDG	58	None	95.53	Water		
M-CP-49	PTGTVKWFDHSKGFGFIHPDDG	59		96.68	Water		
M-CP-50	TQGSVKWFNGEKGFGFIEQDGG	91		95.41	Water		
M-CP-51	**FNGIVKNFDLEKGYGFIQPTDG**	62		95.47	Water		
M-CP-52	PTGKVKFYDDDKGFGFITGDDG	100		96.1	Water		
M-CP-53	**PTGKVKFYDDEKGFGFISTDDG**	31		95.26	Water		
M-CP-54	PTGKVKFYDDQKGFGFITGDDG	50		95.97	Water		
M-CP-55	PTGKVKFYDEEKGFGFISSDDG	11		96.26	Water		
M-CP-56	PTGKVKFYDEEKGFGFISTDDG	85		95.94	Water		
M-CP-57	PTGKVKFYDDQKGFGFISGDDG	19		95.65	Water		
M-CP-58	PTGKVKFYDEEKGFGFISTDEG	31		95.48	Water		
M-CP-59	**PTGKVKWYDVDKGFGFLSQEEG**	88		97.33	Water		
M-CP-60	STGKVIRFDEFKGYGFVAPDEG	84		97.19	Water		
M-CP-61	KTGKILRFDEVRGYGFIVPNEG	50		97.11	Water		
M-CP-62	PSGRIIKWMTDRGFGFIQEDGA	6		95.34	Water		
M-CP-63	FDANAFNADGQRGFGFIDSDES	5		95.71	DMSO		
M-CP-64	ENGTVKWFNDAKGFGFISRENG	82		97.11	Water		
M-CP-65	ATGTVKFFAQDKGFGFITPDNG	95		95.87	Water		
M-CP-66	AHGTLTRWNTDRGFGFITPAQP	15		95.5	Water		
M-CP-67	YQGRLSDWDDHKGFGFVVPHGG	56		96.49	Water		
M-CP-68	QSGEIVDWNDARGFGFIVAAGN	1		95.57	DMSO		
M-EF-73	AKAKFDRTKPHVNIGTIG	72		97.31	Water	Biomatik, Cambridge, Ontario, Canada	
M-EF-74	AKAKFERKKPHVNVGTIG	37		96.87	Water		
M-EF-75	AKAKFERNKLHVNVGTIG	24		96.19	Water		
M-EF-76	AKAKFERTKLHVNVGTIG	19		95.77	Water		
M-EF-77	AKAKFERTKPHVNIGTIG	582		97.08	Water		
M-EF-78	AKAKFERTKPHVNVGTIG	998		97.45	Water		
M-EF-79	AKAKFLREKLHVNVGTIG	54		96.95	Water		
M-EF-80	AKEKFDRSLPHCNVGTIG	102		96.39	Water		
M-EF-81	AKEKFDRSLPHVNVGTIG	522		97.17	Water		
M-EF-82	AKEKFERNKPHVNVGTIG	253		97.19	Water		
M-EF-83	AKEKFERSKPHVNVGTIG	124		96.6	Water		
M-EF-84	AKEKFERTKPHVNVGTIG	312		97.36	Water		
M-EF-85	AKERFDRSLPHVNVGTIG	13	N-terminal acetylation	96.57	Water		
M-EF-86	AKGKFERTKPHVNVGTIG	83		97.35	Water		
M-EF-87	AKSKFERNKPHVNIGTIG	360		97.18	Water		
M-EF-88	AKSKFERNKPHVNVGTIG	62		95.03	Water		
M-EF-89	AKSKFERTKPHVNIGTIG	18		98.6	Water		
M-EF-90	AKTKFERTKPHVNVGTIG	70		95.65	Water		
M-EF-91	ARAKFLREKLHVNVGTIG	42		95.74	Water		
M-EF-92	PKTAYLRTKPHLNIGTMG	38		95.76	Water		
M-EF-93	PKTAYVRTKPHLNIGTMG	28		95.83	Water		
M-EF-94	SKEKFERSKPHVNVGTIG	78		95.72	Water		
M-EF-95	SKEKFERTKPHVNVGTIG	330		96.41	Water		
M-EF-96	SKTAYVRTKPHLNIGTMG	60		95.6	Water		
M-EF-97	SKKAYVRTKPHLNIGTMG	1		95.85	Water		
Cm elf18	GKAKFERTKPHVNIGTIG	63		96.9	Water	Genscript, Piscataway, New Jersey, USA	
Cm csp22-1	ATGTVKWFNAEKGFGFIAPDNG	87	None	99.2	Water		
Cm csp22-2	ANGTVKWFNGEKGFGFITVDAV	72		96.2	DMSO		

26

Cm csp22-3	PTGKVKFYDEDKGFGFISSDDG	87		95.5	Water		
s-csp22-3	YIKPFFSFDGVGDSGGTKEDKD	N/A		95.3	Water		
elf12	SKEKFERTKPHV	N/A	N-terminal acetylation	96.7	Water		

Table S3: Oligonucleotides used in this study.

Primers for amplification of *Clavibacter CSP*s and screening of tomato *core* mutants.				
Name	Product Size (bp)	Sequence	Tm (°C)	Reference
Amplify_cspB_CO_Pst_F	408	GGTGGTcatatgATGCCCACGG	60	This Study
Amplify_cspB_CO_Pst_R		GGTGGTgaattcTCAAGCATAGTC		
Screen_cspB_pDSK519_F	441	GGTGGTcatatgATGCCCACGG	60	
Screen_cspB_pDSK519_R		GGTGGTgaattcTCAAGCATAGTC		
Amplify_cmx_from_pselact_ko_w_promotor_BsiWI_RS_F	827	GGTGGTCGTACGtatccagtcactatggcggcc	56	
Amplify_cmx_from_pselact_ko_BsiWI_RS_R		GGTGGTCGTACGTTAGTAAGCCGGATCCTCTAGA		
Screen_antibiotic_cassette_pDSK519_F	1995/1064	GGGCTATGTGCAACGGGAAT	60	
Screen_antibiotic_cassette_pDSK519_R		GCGATCTGGCTATCGCGG		
Primers for screening of codon-optimized *Clavibacter* CSPs in *E. coli* for recombinant protein expression				
Name	Product Size (bp)	Sequence	Tm (°C)	Reference
attB1_F	Varies	GGGGACAAGTTTGTACAAAAAAGCAGGCT	57	This Study
attB1_R		GGGGACCACTTTGTACAAGAAAGCTGGGT		
Primers to screen for bp edits within regions of CORE targeted by gRNA				
Name	Product Size (bp)	Sequence	Tm (°C)	Reference
Screen_CORE_CRISPR_edit_F	1184	ATGATTCTCCCAAAGAATTCTCACTT	55	This Study
Screen_CORE_CRISPR_edit_R		ACTTGATTATAAGACAACGAAAGCTC		
Primers for qPCR primer for relative expression in *Cm* NCPPB382				
Name	Product Size (bp)	Sequence	Tm (°C)	Reference
NCPPB382_cspA1_ver1_qPCR_F	125	TTAGAGCGGGCGGATGTTC	59	This Study
NCPPB382_cspA1_ver1_qPCR_R		TGTTCGCTCACTACTCCGC		
NCPPB382_cspA2_qPCR_F	80	GCCGAGTAGTGGACGAAGAC	59	
NCPPB382_cspA2_qPCR_R		GAGAAGGGGTTCGGGTTCAT		
NCPPB382_cspB_qPCR_F	51	CCGAGCTGATGAACCCGAA	59	
NCPPB382_cspB_qPCR_R		GCAAGGTGAAGTTCTACGACG		
NCPPB382_bipA_qPCR_F	63	GGGTGCTGGTCGTCGTA	59	Jiang et al., 2019
NCPPB382_bipA_qPCR_R		CGAGCCGCTGTTCAAG		

SI References

1. C. Camacho, *et al.,* BLAST+: architecture and applications. *BMC Bioinform.* 10, 421 (2009).

2. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9, 5114 (2018).

3. D. Stevens, A. Moreno-Pérez, A. Weisberg, C. Ramsing, J. Fliegmann, N. Zhang, M. Madrigal, G. Martin, A. Steinbrenner, G. Felix, G. Coaker, Data from "**Evolutionary dynamics of proteinaceous MAMPs reveals intrabacterial antagonism of plant immune perception.**" Zenodo. Available at http://dx.doi.org/10.5281/zenodo.10724865 (2024).

4. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849 (2016).

5. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812 (2014).

6. O. Wagih, ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647 (2017).

7. S. Xu, *et al.,* Use ggbreak to Effectively Utilize Plotting Space to Deal With Large Datasets and Outliers. *Front. Genet.* 12, 774846 (2021).

8. M. D. Lee, GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, btz188 (2019).

9. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30, 772–780 (2013).

10. Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinform. *11*, 431. 10.1186/1471-2105-11-431.

11. B. Q. Minh, *et al.,* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* (2020) doi:10.1093/molbev/msaa015.

12. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 35, 1026–1028 (2017).

13. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* 43, W39–W49 (2015).

14. X. Li, L. Ma, X. Mei, Y. Liu, H. Huang, ggmotif: An R Package for the extraction and visualization of motifs from MEME software. *Plos One* 17, e0276979 (2022).

15. S. L. Nystrom, D. J. McKay, Memes: A motif analysis environment in R using tools from the MEME Suite. *PLoS Comput. Biol.* 17, e1008991 (2021).

16. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One* 5, e9490 (2010).

17. K. P. Schliep, phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593 (2011).

29

18. L.-G. Wang, *et al.,* treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* 37, 599–603 (2019).

19. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36 (2017).

20. S. Xu, *et al.,* ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Mol. Biol. Evol.* 38, 4039–4042 (2021).

21. H.-G. Drost, A. Gabel, I. Grosse, M. Quint, Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Mol. Biol. Evol.* 32, 1221–1231 (2015).

22. V. Ranwez, E. J. P. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol. Biol. Evol.* 35, 2582–2584 (2018).

23. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274 (2015).

24. B. Murrell, *et al.,* FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* 30, 1196–1205 (2013).

25. A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278 (2014).

26. C. Zipfel, *et al.,* Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell* 125, 749–760 (2006).

27. S. Lacombe, *et al.,* Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat Biotechnol* 28, 365 (2010).

28. J. Schindelin, *et al.,* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682 (2012).

29. K. N. Mason, G. Ekanayake, A. Heese, Chapter 10 Staining and automated image quantification of callose in Arabidopsis cotyledons and leaves. *Methods Cell Biol.* 160, 181–199 (2020).

30. L. Gómez-Gómez, G. Felix, T. Boller, A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. *Plant J.* 18, 277–284 (1999).

31. M. Albert, *et al.* Arabidopsis thaliana Pattern Recognition Receptors for Bacterial Elongation Factor Tu and Flagellin Can Be Combined to Form Functional Chimeric Receptors*. *J. Biol. Chem.* 285, 19035–19042 (2010).

32. F. Kumagai-Sano, T. Hayashi, T. Sano, S. Hasezawa, Cell cycle synchronization of tobacco BY-2 cells. *Nat. Protoc.* 1, 2621–2627 (2006).

33. G. Fiorin, A. Sánchez-Vallet, B. Thomma, G. Pereira, P. Teixeira, MAMP-triggered Medium Alkalinization of Plant Cell Cultures. *BIO-Protoc.* 10, e3588 (2020).

30

34. T. B. Jacobs, P. R. LaFayette, R. J. Schmitz, W. A. Parrott, Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15, 16 (2015).

35. T. B. Jacobs, N. Zhang, D. Patel, G. B. Martin, Generation of a Collection of Mutant Tomato Lines Using Pooled CRISPR Libraries. *Plant Physiol.* 174, 2023–2037 (2017).

36. N. Zhang, C. Hecht, X. Sun, Z. Fei, G. B. Martin, Loss of function of the bHLH transcription factor Nrd1 in tomato enhances resistance to Pseudomonas syringae. *Plant Physiol.* 190, 1334–1348 (2022).

37. C. L. M. Gilchrist, Y.-H. Chooi, clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475 (2021).

38. K.-H. Gartemann, *et al.,* The Genome Sequence of the Tomato-Pathogenic Actinomycete Clavibacter michiganensis subsp. michiganensis NCPPB382 Reveals a Large Island Involved in Pathogenicity ▿ †. *J Bacteriol* 190, 2138–2149 (2008).

39. M. Alexou, A. D. Peuke, Plant Mineral Nutrients, Methods and Protocols. *Methods Mol. Biol.* 953, 195–207 (2012).

40. N. Jiang, *et al.,* Evaluation of suitable reference genes for normalization of quantitative reverse transcription PCR analyses in Clavibacter michiganensis. *Microbiology open* 8, e928 (2019).

41. A. J. Weisberg, *et al.,* A Novel Species-Level Group of Streptomyces Exhibits Variation in Phytopathogenicity Despite Conservation of Virulence Loci. *Mol. Plant-Microbe Interact.* **34**, 39–48 (2021).

42. A. G. Matthysse, S. Stretton, C. Dandie, N. C. McClure, A. E. Goodman, Construction of GFP vectors for use in Gram-negative bacteria other than Escherichia coli. *FEMS Microbiol. Lett.* 145, 87–94 (1996).

43. D. H. Figurski, D. R. Helinski, Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci.* 76, 1648–1652 (1979).

44. B. Kessler, V. de Lorenzo, K. N. Timmis, A general system to integrate lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy. *Mol. Gen. Genet. MGG* 233, 293–301 (1992).

31

CONCLUSIONS AND FUTURE DIRECTIONS

Plants regulate biotic threats through a variety of physical, chemical, and molecular barriers. This includes a two-layer signaling defense system composed of receptors that reside at the plant cell surface and intracellularly to respond directly or indirectly to pathogen effectors or epitopes (1). As pathogens attempt to overcome plant's defense network, an evolutionary arms race can emerge, leading to novel virulence strategies in coordination with the plant host (1). However, how pathogen diversity affects immune signaling has not been well explored. This work has revealed undescribed diversity through comparative genomics and novel mechanisms of pathogen evasion, though many questions remain and will be discussed in the sections below.

Quantitative immunity versus pathogen evolution

A classic example of understanding plant disease resistance was through the gene-for-gene hypothesis, where a single pathogen virulence gene was controlled by a matching resistance gene in the plant host (2). Alternatively, many traits including disease resistance can exhibit a continuous phenotypic distribution (i.e., do not follow Mendelian segregation ratios) (3). While resistance (R)-genes can display qualitative disease resistance, particularly in controlled, greenhouse conditions, diverse populations of plants and/or pathogens can display diverse responses.

We find through characterizing the natural diversity of pathogen epitopes that plant immune responses tend to fall into three categorical responses: strongly immunogenic, weakly immunogenic, and non-immunogenic (4). Some elf18 variants from *Xanthomonas, Pseudomonas,* and *Streptomyces* displayed a weaker, though detectable, reactive oxygen species (ROS) burse but failed to induce other classical immune responses such as MAPK induction, callose deposition,

and growth inhibition (4). These peptides were termed 'deviants' for their ability to induce early (ex. ROS burse) induction but fail to induce later (ex. MAPK cascade) immune responses. Flg22 deviant peptides have also been described; though, their epitope sequences are more closely shared (5). Conversely, we find elf18 deviants have convergently evolved and the bifurcating responses are not due to a particular sequence polymorphism (4). While not fully characterized, some csp22 variants may also act as deviants due to their reduced ROS production (4).

Since there is a consistent bifurcation of responses by different epitope-receptor combinations, it is curious why this phenotype has evolved. Below I will discuss a few possible explanations using flagellin as an example. One, we may be observing a new evasion strategy, where some bacteria induce minor ROS production but limit other stronger immune responses, enabling colonization. While ROS can act as an anti-microbial compound, it also acts as a signaling molecule to induce priming of immunity in other plant cells distal to the site of infection (6, 7). ROS is also required for the activation of systematic acquired resistance (SAR) (6, 8). While deviants are unable to induce late immune responses locally, it is unclear if ROS induced by deviant peptides are sufficient to activate defense and transcriptional remodeling in distal tissue. A lack of induction of stronger immune outputs locally and SAR in distal tissue may provide enough of an advantage for the bacterium to colonize and cause disease systemically.

Alternatively, the evolution of these intermediate-immune inducing epitopes may be reflective of the fitness costs associated with polymorphism in certain bacterial proteins. It is hypothesized that antagonistic pleiotropy, where a protein experiences different selection and fitness pressures due to variable environments, may contribute to their evolution; in plant-microbe interactions, this could be framed as negative selection on the epitope-encoded protein to maintain core functionality and positive selection to aid in evasion of plant immune perception. For

example, highly diverged forms of the flagellin monomer, even in artificial contexts, have direct consequences on swimming motility (9, 10). Many pathogens require motility via the flagellin to colonize a plant host and flagellin deletion mutants impact pathogen virulence (10, 11). Therefore, the evolution of a functional, albeit slightly immunogenic form could be a selective advantage in a pathogen population.

Finally, many pathogens have other means to suppress recognition and thus, an immune-inducing epitope may not hinder colonization and disease development. Some bacterial pathogens have effectors secreted via the type III-secretion system, which target and suppress signaling components of immune induction (12). Other effectors may downregulate protein expression during *in planta* growth or induce the production of other catalytic enzymes which mask the epitope before recognition (13, 14). Bacteria may also carry extra epitope-encoding proteins which block perception of immunogenic forms (4). In other non-plant hosts, fine tuning the interaction between epitope and receptor has additionally been described; flagellin monomers from *Salmonella* display direct binding with a convergently evolved mammalian PRR, TLR5, but circumvent subsequent TLR5 signaling (15).

For proteins where multiple possible epitopes are encoded and perceived, major questions regarding the epitope evolution, protein function, and epitope processing and perception remain. For example, are some epitope-encoding regions more prone to polymorphisms and/or less likely to abolish protein function (or vice versa)? If one epitope encoding region changes in protein function and/or immune perception, does this influence the evolution of other epitope encoding regions along the protein sequence in a linked-manner or can they evolve independently of each other, reflective of the differences in receptor conservation from the host? Does this affect release of the epitope from the full-length protein? While we elucidate the influence of multi-copy epitopes

on immune outcomes in bacteria, would we see a similar outcome or not in other organisms such as fungi, oomycetes, parasitic plants, and nematodes? Research into these major questions would yield critical insights into the diversity and functional evolution of epitope-receptor biology.

Going beyond the model plant – what is Pattern Triggered Immunity across the kingdom?

Much of our plant immune knowledge is focused primarily on model plants such as *Arabidopsis thaliana* and *Nicotiana benthamiana* as well as a handful of model crops such as wheat, rice, tomato, potato, lettuce, and bean (2). While there has been a recent study assessing the diversity plant immune perception, revealing contraction and expansion of receptor number associated with diverse lifestyles, little has been conducted to functionally characterize this diversity as well as explore immunity extensively in the plant kingdom (16). The plant kingdom encompasses immense diversity across most climates of the world that broadly have not been studied (17). For non-agricultural associated plants which grow in diverse and extreme environments, how much of pattern triggered immunity (PTI) perception and signaling is shared with cultivated crops and their wild relatives remains elusive? Using comparative genomics and a systematic screen be applied to better understand the evolution and impact of natural diversity on PTI recognition and response.

Plants have a careful balance between defense and growth, both of which can be affected by their environment (18, 19). In the model plant *A. thaliana* infected with the model pathogen *Pseudomonas syringae pv. tomato* DC3000*,* elevated temperatures enhance pathogen virulence and reduce plant hormone-mediated defense, leading to increased disease susceptibility (19). Overall, studying immunity in non-agricultural species from diverse and extreme environments may reveal novel mechanisms used by plants for immune activate and/or rely on other barriers

such as physiology and metabolites composition to limit biotic threats. Previous work has shown some extremophytes within the Brassicaceae family display differential response to the stress-associated phytohormone ABA (20). Therefore, plant species that natively grow in extreme environments may have evolved differential signaling networks or other novel immune components that differ from model plants. Understanding how plants have evolved their immune responses to adapt to diverse environments will likely inspire the development of new bioengineering approaches that are less susceptible to pathogen and their adaption, particularly in the wake of climate change.

For a particular receptor, diversity can have a considerable impact on perception abilities and shaping biotic communities. The most well studied PRR, the flagellin receptor FLS2, has shown that diverse homologs have differing recognition profiles (21-23). The csp22 receptor CORE is restricted to Solanaceous family with the few species assessed displaying different recognition capacities (22, 24). In addition, csp22 is a multi-copy epitope and members within the Vitaceae and Rutaceae plant families can respond to csp22 despite lacking a homologous CORE receptor, indicating convergent evolution (4, 25, 26). The diversity and differing recognition profiles may influence the groups of bacteria and other biotic pathogens a plant species may encounter. For other receptors, however, far less is known. Systematically characterizing receptor diversity across the plant kingdom may reveal functional receptors with broader and/or stronger epitope interactions and provide insight into how receptors shape pathogen communities.

A final layer to which has been severely overlooked is the understanding of immune signaling and PTI in different plant niches, organs, and cell types. Much of our understanding of PTI has relied on assessing induction in mature foliar and occasionally, root tissue. For example, some plant receptors are only expressed in the roots or the shoots, certain tissue zones, and are

heavily tied to plant age/development. But much of this knowledge is restricted to a handful of receptors such as the flagellin receptor FLS2 and the elongation factor Tu receptor EFR (27). Many pathogens primarily colonize a particular tissue niche or organ (28). Thus, it would be important to determine if certain tissues and zones are more effective for immune engineering against pathogens. In *A. thaliana,* the first seed-to-seed single cell atlas was conducted which could serve as a good initial dataset to evaluate this central question (29). It would be curious to assess if there are expression differences, different functional isoforms of immune components across cell types, and differences in developmental stages that can influence the response to pathogen perception and disease susceptibility. Research into the expression profiles of receptors in different cell types and plant ages may provide insight.

Leveraging computational structural biology to understand receptor-ligand interactions

Diverse plant species can encode many receptors with diverse recognition capabilities (2, 16). When considering the remarkable diversity of the plant kingdom, most receptor-ligand interactions are unknown. A common strategy to confer resistance to pathogens is to transfer receptors between plant species (30). However, without sufficient understanding of host-pathogen interactions from an evolutionary perspective, these approaches frequently fall short due to rapid evolution of pathogens, particularly when only a single gene transfer is involved. Pathogens encode an array of effectors, frequently undergoing gene gain/loss events and thus in diverse combinations enable them to overcome host defenses (28).

Considering the millions of years of possible interactions, there are likely many uncharacterized epitopes that are recognized by plant immune systems. If characterizing their interactions may unlock a strategy to engineer more effective disease resistance in crops, the

question lies – How to scale for these interactions? Using currently established assays, this goal is nearly impossible. However, advances in computational and structural biology may provide a gateway. Using computational biology and machine learning (ML), we have gained several seminal advancements in prediction and modeling of biological processes and molecules (31-33). For epitope outcomes, predicting their immunogenicity may be possible due to an increase in large epitope screening studies, data which would be required to train any ML model (4, 5, 10). Such models could induce a paradigm shift in studying plant immunity, particularly for emerging pathogens, where immunogenic outcomes could be determined via amplicon or whole genome sequencing and biochemical screening may not be required.

With new advancements in structure prediction both empirically via cryo-EM and computationally, we may build structure-guided receptor-epitope studies that provide insights into the biochemical requirements for strong complex formation. For example, the multi-copy epitope csp22 can be recognized by different receptors from Solanaceous and other plant families (4, 24-26). While experimental screening the hundreds of candidate receptors encoded would likely yield the other cognate receptor(s), such approaches are tedious, time and resource consuming (34, 35). Using a genomics approach may yield homologous receptors rapidly but would likely fall short when a receptor is unknown or has convergently evolved. A ML model of epitope-receptor interactions could be used to predict other receptors which display similar predicted binding interactions, thus reducing the degree of experimental screening required. Additionally, such models could aid in developing synthetic receptors with expanded recognition profiles. Such knowledge of receptor diversity and their ligand recognition capabilities may provide insights into how *de novo* receptors develop and function, which could be potentially used to engineer receptors with custom perception profiles.

The paradigm of Gram-positive actinobacteria and plant immunity

For many bacterial pathogens, a zig-zag model was established to describe the arms race with plant species (1). This model included the initial recognition of pathogens by surface-localized PRRs, the secretion of pathogen effector proteins to silence recognition and remodel the plant host for colonization and disease, and the possible subsequent recognition of pathogen effector proteins and/or their perturbations through intracellular receptors, triggering a strong immune response and restricting disease development. For actinobacteria, central questions remain if this model was applicable. Unlike their Gram-negative peers, actinobacteria lack the type III secretion system and thus, pathogen proteins are unable to be transported into the plant cell (36). For bacteria of the *Clavibacter, Streptomyces,* and *Curteobacterium* genus*,* previous QTLs have been described though none were a dominant resistance gene (36). Additionally, it was previously thought that no purified epitope was demonstrated as a clear elicitor sufficient for inducing immunity (36).

Based on data presented here as well as recent studies, actinobacteria-host interaction seems likely. We have found common bacterial epitopes elf18 and csp22 are perceived by plant species *Arabidopsis thaliana* and tomato and for bacteria of the *Clavibacter* genus, can affect host colonization (4). Like other pathogens, *Clavibacter* effectors are upregulated during *in planta* infection, including some serine proteases, and are recognized by some Solanaceous plant species, potentially limiting pathogen host range (37-39). These data support the model that plant immunity can recognize Gram-positive pathogens and influence disease susceptibility.

However, a classical zig-zag model falls short for actinobacterial pathogens. Unlike other pathogen groups, actinobacteria release effectors into the apoplast primarily though the basic Sec and Tat secretion system (36). While effector triggered-immunity (ETI) is well-known for its

strong induction including the localized cell death (as known as hypersensitive response - HR), HR has been measured in different combinations of plant species and *Clavibacter* pathogens (37, 39). As a result, any R gene would likely be a PRR and any effector evasion or detection would occur at the cell-surface. In some ways, the interaction we observed in *Clavibacter* mirrors certain aspect of the fungal pathogen *Cladosporium fulvum* and tomato, another pathosystem which does not follow the classical zig-zag model (40, 41). In this system, they propose the model of effector-triggered defense, where surface-localized RLPs act as R genes though the associated cell death is more delayed than classical ETI (40). Thus far, the data supports a similar model. However, question remains what receptor(s) recognize these bacteria, if recognition is direct or indirect, and if this mechanism is found beyond *Clavibacter* pathogens remains elusive.

References

1. J. D. G. Jones, J. L. Dangl, The plant immune system. *Nature* **444**, 323 (2006).

2. B. P. M. Ngou, P. Ding, J. D. G. Jones, Thirty years of resistance: Zig-zag through the plant immune system. *The Plant Cell* **34**, 1447–1478 (2022).

3. D. A. St. Clair, Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. *Annu. Rev. Phytopathol.* **48**, 247–268 (2010).

4. D. M. Stevens, A. Moreno-Pérez, A. J. Weisberg, C. Ramsing, J. Fliegmann, N. Zhang, M. Madrigal, G. Martin, A. Steinbrenner, G. Felix, G. Coaker, Natural variation of immune epitopes reveals intrabacterial antagonism. *Proc. Natl. Acad. Sci.* **121**, e2319499121 (2024).

5. N. R. Colaianni, K. Parys, H.-S. Lee, J. M. Conway, N. H. Kim, N. Edelbacher, T. S. Mucyn, M. Madalinski, T. F. Law, C. D. Jones, Y. Belkhadir, J. L. Dangl, A complex immune response to flagellin epitope variation in commensal communities. *Cell Host Microbe* **29**, 635-649.e9 (2021).

6. B. Castro, M. Citterico, S. Kimura, D. M. Stevens, M. Wrzaczek, G. Coaker, Stress-induced reactive oxygen species compartmentalization, perception and signalling. *Nat. Plants* **7**, 403–412 (2021).

7. P. Jacob, J. Hige, J. L. Dangl, Is localized acquired resistance the mechanism for effector-triggered disease resistance in plants? *Nat. Plants* **9**, 1184–1190 (2023).

8. M. Á. Peláez-Vico, Y. Fichman, S. I. Zandalinas, C. H. Foyer, R. Mittler, ROS are universal cell-to-cell stress signals. *Curr. Opin. Plant Biol.* **79**, 102540 (2024).

9. K. D. Smith, E. Andersen-Nissen, F. Hayashi, K. Strobe, M. A. Bergman, S. L. R. Barrett, B. T. Cookson, A. Aderem, Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat. Immunol.* **4**, 1247–1253 (2003).

10. K. Parys, N. R. Colaianni, H.-S. Lee, U. Hohmann, N. Edelbacher, A. Trgovcevic, Z. Blahovska, D. Lee, A. Mechtler, Z. Muhari-Portik, M. Madalinski, N. Schandry, I. Rodríguez-Arévalo, C. Becker, E. Sonnleitner, A. Korte, U. Bläsi, N. Geldner, M. Hothorn, C. D. Jones, J. L. Dangl, Y. Belkhadir, Signatures of antagonistic pleiotropy in a bacterial flagellin epitope. *Cell Host Microbe* **29**, 620-634.e9 (2021).

11. K. Naito, F. Taguchi, T. Suzuki, Y. Inagaki, K. Toyoda, T. Shiraishi, Y. Ichinose, Amino Acid Sequence of Bacterial Microbe-Associated Molecular Pattern flg22 Is Required for Virulence. *Mol. Plant-Microbe Interact.* **21**, 1165–1174 (2008).

12. D. Lu, P. He, L. Shan, Bacterial effectors target BAK1-associated receptor complexes. *Commun. Integr. Biol.* **3**, 80–83 (2010).

13. Y. Deng, H. Chen, C. Li, J. Xu, Q. Qi, Y. Xu, Y. Zhu, J. Zheng, D. Peng, L. Ruan, M. Sun, Endophyte Bacillus subtilis evade plant defense by producing lantibiotic subtilomycin to mask self-produced flagellin. *Commun Biology* **2**, 368 (2019).

14. M. O. Andrade, Z. Pang, D. S. Achor, H. Wang, T. Yao, B. H. Singer, N. Wang, The flagella of 'Candidatus Liberibacter asiaticus' and its movement in planta. *Mol. Plant Pathol.* **21**, 109–123 (2020).

15. S. J. Clasen, M. E. W. Bell, A. Borbón, D.-H. Lee, Z. M. Henseler, J. de la Cuesta-Zuluaga, K. Parys, J. Zou, Y. Wang, V. Altmannova, N. D. Youngblut, J. R. Weir, A. T. Gewirtz, Y. Belkhadir, R. E. Ley, Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Sci. Immunol.* **8**, eabq7001 (2023).

16. B. P. M. Ngou, R. Heal, M. Wyler, M. W. Schmid, J. D. G. Jones, Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nat. plants* **8**, 1146–1152 (2022).

17. A. R. Zuntini, *et al.*, Phylogenomics and the rise of the angiosperms. *Nature*, 1–8 (2024).

18. B. Huot, J. Yao, B. L. Montgomery, S. Y. He, Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. *Mol. Plant* **7**, 1267–1287 (2014).

19. B. Huot, C. D. M. Castroverde, A. C. Velásquez, E. Hubbard, J. A. Pulman, J. Yao, K. L. Childs, K. Tsuda, B. L. Montgomery, S. Y. He, Dual impact of elevated temperature on plant defence and bacterial virulence in Arabidopsis. *Nat Commun* **8**, 1808 (2017).

20. Y. Sun, D.-H. Oh, L. Duan, P. Ramachandran, A. Ramirez, A. Bartlett, K.-N. Tran, G. Wang, M. Dassanayake, J. R. Dinneny, Divergence in the ABA gene regulatory network underlies differential growth control. *Nat. Plants* **8**, 549–560 (2022).

21. K. Mueller, P. Bittel, D. Chinchilla, A. K. Jehle, M. Albert, T. Boller, G. Felix, Chimeric FLS2 Receptors Reveal the Basis for Differential Flagellin Perception in Arabidopsis and Tomato. *Plant Cell Online* **24**, 2213–2224 (2012).

22. Y. Wei, C. Caceres-Moreno, T. Jimenez-Gongora, K. Wang, Y. Sang, R. Lozano-Duran, A. P. Macho, The Ralstonia solanacearum csp22 peptide, but not flagellin-derived peptides, is perceived by plants from the Solanaceae family. *Plant Biotechnol J* **16**, 1349–1362 (2018).

23. U. Fürst, Y. Zeng, M. Albert, A. K. Witte, J. Fliegmann, G. Felix, Perception of Agrobacterium tumefaciens flagellin by FLS2XL confers resistance to crown gall disease. *Nat. Plants* **6**, 22–27 (2020).

24. G. Felix, T. Boller, Molecular Sensing of Bacteria in Plants THE HIGHLY CONSERVED RNA-BINDING MOTIF RNP-1 OF BACTERIAL COLD SHOCK PROTEINS IS RECOGNIZED AS AN ELICITOR SIGNAL IN TOBACCO. *J Biol Chem* **278**, 6201–6208 (2003).

25. L. P. Burbank, J. Ochoa, Evidence for Elicitation of an Oxidative Burst in Vitis vinifera by Xylella fastidiosa Cold Shock Protein Peptide csp20. *PhytoFrontiers^{TM}* **2**, 339–341 (2022).

26. J. Trinh, T. Li, J. Y. Franco, T. Y. Toruño, D. M. Stevens, S. P. Thapa, J. Wong, R. Pineda, E. Á. de Dios, T. L. Kahn, D. K. Seymour, C. Ramadugu, G. L. Coaker, Variation in microbial feature perception in the Rutaceae family with immune receptor conservation in citrus. *Plant Physiol.*, doi: 10.1093/plphys/kiad263 (2023).

27. I. Wyrsch, A. Domínguez-Ferreras, N. Geldner, T. Boller, Tissue-specific FLAGELLIN-SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. *N. Phytol.* **206**, 774–784 (2015).

28. R. B. Abramovitch, J. C. Anderson, G. B. Martin, Bacterial elicitation and evasion of plant innate immunity. *Nat Rev Mol Cell Bio* **7**, 601–611 (2006).

29. T. A. Lee, T. Nobori, N. Illouz-Eliaz, J. Xu, B. Jow, J. R. Nery, J. R. Ecker, A Single-Nucleus Atlas of Seed-to-Seed Development in Arabidopsis. *bioRxiv*, 2023.03.23.533992 (2023).

30. S. Lacombe, A. Rougon-Cardoso, E. Sherwood, N. Peeters, D. Dahlbeck, H. P. van Esse, M. Smoker, G. Rallapalli, B. P. H. J. Thomma, B. Staskawicz, J. D. G. Jones, C. Zipfel, Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat Biotechnol* **28**, 365 (2010).

31. J. L. Watson, *et al.*, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).

32. S. Ohno, N. Manabe, Y. Yamaguchi, Prediction of protein structure and AI. *J. Hum. Genet.*, 1–4 (2024).

33. S. V. Torres, *et al.*, De novo design of high-affinity binders of bioactive helical peptides. *Nature* **626**, 435–442 (2024).

34. A. Schultink, A. D. Steinbrenner, A playbook for developing disease-resistant crops through immune receptor identification and transfer. *Curr. Opin. Plant Biol.* **62**, 102089 (2021).

35. Y. J. Ahn, H. Kim, S. Choi, C. Mazo-Molina, M. Prokchorchik, N. Zhang, B. Kim, H. Mang, N. Koehler, J. Kim, S. Lee, H. Yoon, D. Choi, M. Kim, C. Segonzac, G. B. Martin, A. Schultink, K. H. Sohn, Ptr1 and ZAR1 immune receptors confer overlapping and distinct bacterial pathogen effector specificities. *N. Phytol.* **239**, 1935–1953 (2023).

36. S. P. Thapa, E. W. D. II, Q. Lyu, A. J. Weisberg, D. M. Stevens, C. R. Clarke, G. Coaker, J. H. Chang, The Evolution, Ecology, and Mechanisms of Infection by Gram-Positive, Plant-Associated Bacteria. *Annu Rev Phytopathol* **57**, 341–365 (2019).

37. R. Nissinen, Y. Xia, L. Mattinen, C. A. Ishimaru, D. L. Knudson, S. E. Knudson, M. Metzler, M. Pirhonen, The putative secreted serine protease Chp-7 is required for full virulence and induction of a nonhost hypersensitive response by Clavibacter michiganensis subsp. sepedonicus. *Mol Plant-microbe Interactions Mpmi* **22**, 809–19 (2009).

38. I. S. Hwang, E.-J. Oh, E. Song, I. W. Park, Y. Lee, K. H. Sohn, D. Choi, C.-S. Oh, An Apoplastic Effector Pat-1Cm of the Gram-Positive Bacterium Clavibacter michiganensis Acts as Both a Pathogenicity Factor and an Immunity Elicitor in Plants. *Front. Plant Sci.* **13**, 888290 (2022).

39. R. K. Verma, D. Teper, Immune recognition of the secreted serine protease ChpG restricts the host range of Clavibacter michiganensis from eggplant varieties. *Mol Plant Pathol* **23**, 933–946 (2022).

40. H. U. Stotz, G. K. Mitrousia, P. J. G. M. de Wit, B. D. L. Fitt, Effector-triggered defence against apoplastic fungal pathogens. *Trends Plant Sci.* **19**, 491–500 (2014).

41. P. J. G. M. D. Wit, Cladosporium fulvum Effectors: Weapons in the Arms Race with Tomato. *Annu. Rev. Phytopathol.* **54**, 1–24 (2016).

APPENDEX

Plant NLR-triggered immunity: from receptor activation to downstream signaling

Signe Lolle*, Danielle M. Stevens*, and Gitta Coaker

*These authors contributed equally.

Abstract

Innate immune perception is the first line of inducible defense against invading pathogens. Plants lack specialized circulating immune cells. Therefore, diverse cell types are able to recognize and respond to pathogens. Surface-localized and intracellular plant innate immune receptors are capable of recognizing diverse pathogen components. Intracellular nucleotide-binding leucine-rich repeat (NLR) receptors recognize pathogen effectors delivered inside host cells. Recent advances shed light onto NLR activation, phosphorylation of defense signaling nodes and overlap in transcriptional responses between pathogen perception and abiotic stress.

Author Contributions

I developed the figures of the manuscript and wrote the following sections: Introduction, NLR architecture and diverse modes of effector recognition, and NLR activation and resistosome formation.

Plant NLR-triggered immunity: from receptor activation to downstream signaling

Signe Lolle[1,3], Danielle Stevens[1,2,3] and Gitta Coaker[1]

Innate immune perception is the first line of inducible defense against invading pathogens. Plants lack specialized circulating immune cells. Therefore, diverse cell types are able to recognize and respond to pathogens. Surface-localized and intracellular plant innate immune receptors are capable of recognizing diverse pathogen components. Intracellular nucleotide-binding leucine-rich repeat (NLR) receptors recognize pathogen effectors delivered inside host cells. Recent advances shed light onto NLR activation, phosphorylation of defense signaling nodes and overlap in transcriptional responses between pathogen perception and abiotic stress.

Addresses
[1] Department of Plant Pathology, One Shields Avenue, University of California, Davis, 95616, USA
[2] Integrated Genetics and Genomics, University of California, Davis, 95616, USA

Corresponding author: Coaker, Gitta (glcoaker@ucdavis.edu)
[3] These authors contributed equally.

Introduction

With the exception of viruses and specialized insect-vectored bacteria, most pathogens do not replicate inside plant cells. To cause disease and modify their hosts, pathogens secrete proteins, called effectors, into the extracellular space or directly into host cells [1]. Plant innate immune receptors include surface-localized pattern recognition receptors (PRRs) as well as intracellular NLR receptors [2,3]. PRRs can recognize conserved microbe-associated or pathogen-associated molecular patterns (M/PAMPs), damage-associated molecular patterns (DAMPs) and extracellular effector proteins [1,3]. NLRs detect the presence or activity of effectors delivered into host cells during infection [2]. Both PRR-triggered and NLR-triggered immunity (PTI and NTI) lead to a suite of downstream defense responses including the generation o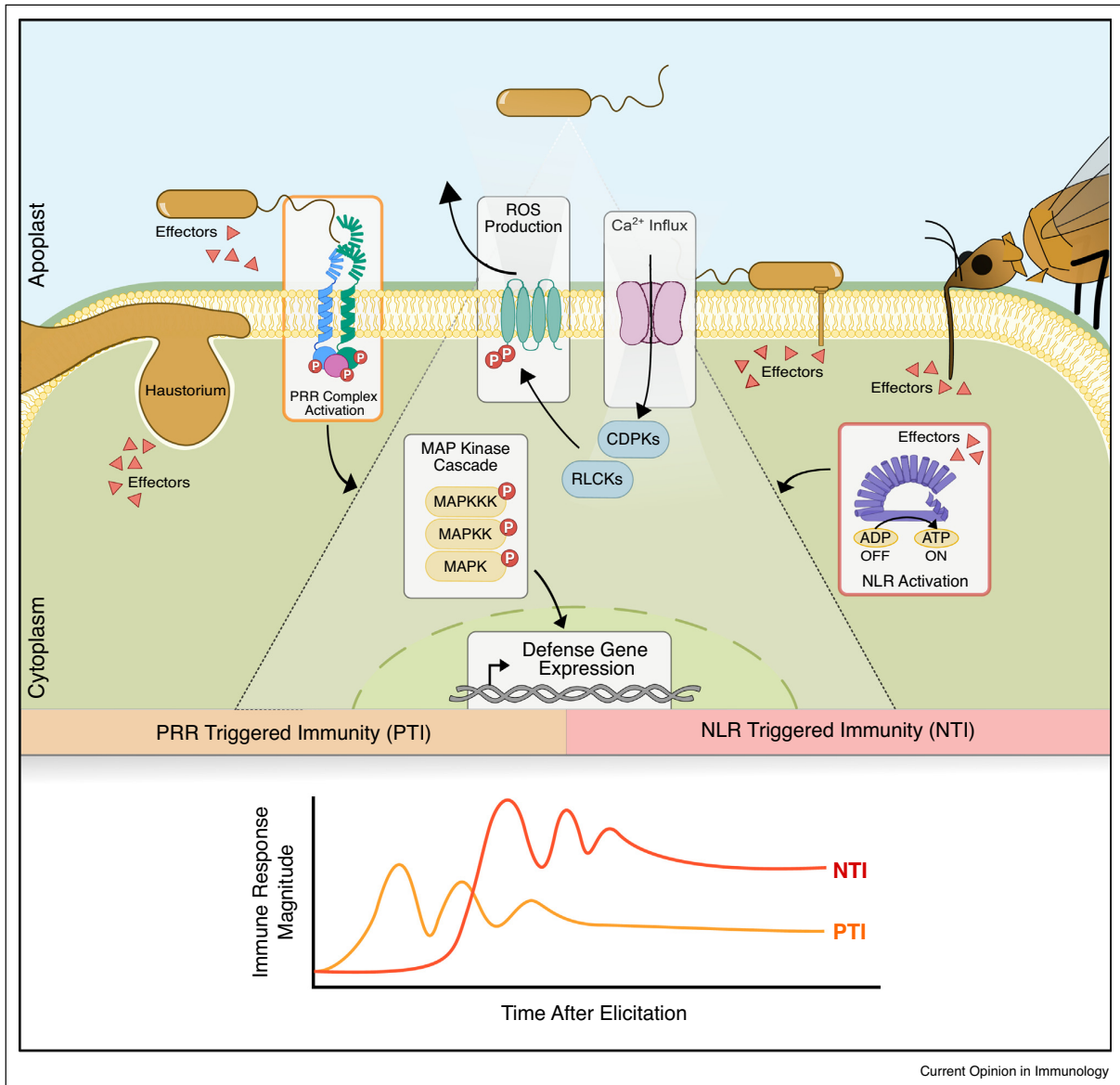f reactive oxygen species (ROS), an influx of extracellular calcium, kinase activation and global transcriptional reprogramming for defense [2,3] (Figure 1). After immune recognition, signals of pathogen perception are propagated from the initial infection site to distal tissues [4]. This systemic immune signaling primes naïve tissue against subsequent attack. Despite similarities, the timing, intensity and duration of defense can differ between PTI and NTI (Figure 1) [5]. NLR activation induces a quantitatively stronger, prolonged and robust response, frequently culminating in programmed cell death at the site of infection [2]. Here we will focus on recent advancements in NLR biology from receptor activation to downstream signaling.

NLR architecture and diverse modes of effector recognition

Plant genomes possess diverse NLR repertoires, with many species possessing hundreds of distinct NLRs that can be used to control pathogen infection in crops [6]. NLRs are composed of a central nucleotide-binding site (NBS) and C-terminal leucine-rich repeats (LRRs). They can be divided into two broad classes based on their N-termini, with CNLs carrying a coiled-coil (CC) domain and TNLs carrying a Toll-interleukin 1 receptor (TIR) domain (Figure 2) [2]. Both pathogen effectors and NLRs can localize to diverse subcellular locations including the cytoplasm, nucleus, plasma membrane, tonoplast and endoplasmic reticulum [2]. The barley CNL MLA10 and the *Arabidopsis* TNL RPS4 reside in the nucleus and cytoplasm, with both locations required for full resistance [7,8]. The *Arabidopsis* CNL RPM1 constitutively associates with the plasma membrane and recognizes a membrane-targeted *Pseudomonas* effector protein [9,10]. How NLRs with diverse subcellular localizations are able to trigger similar defense responses remains unknown.

Not only do NLRs localize to distinct cellular compartments, they also exhibit diversity in mechanisms of pathogen effector recognition. Some receptors can directly bind and recognize cognate pathogen effectors, while others monitor for effector-mediated perturbations of host targets [2]. For example, the *Nicotiana* TNL Roq1 confers resistance against *Xanthomonas* and is able to physically associate with the recognized effector XopQ [11]. In contrast, RPM1 recognizes effector-induced phosphorylation of the host protein RIN4 [12,13]. Animal NLRs recognize PAMPs as well as monitor for pathogen-mediated perturbations, such as the mouse NLR NOD1 that is activated by *Salmonella* SopE effector activity [14]. Plant NLRs can also act as pairs and can exhibit head-to-head chromosomal orientation to

Surface-localized and intracellular plant innate immune receptors recognize diverse pathogens. Plant immune receptors include surface-localized pattern recognition receptors (PRRs) and intracellular nucleotide-binding leucine rich repeat (NLR) receptors. PRRs can recognize microbial features, damage associated molecular patterns, and extracellular receptors from insects, bacteria and filamentous pathogens. NLRs perceive pathogen effectors directly or through effector-mediated perturbations. Both PTI and NTI induce downstream defense responses including an influx of extracellular calcium, reactive oxygen species (ROS) production, kinase activation and transcriptional reprogramming for defense. While downstream defense responses are similar between PTI and NTI, the timing, amplitude and duration of responses differ.

facilitate co-expression (Figure 2) [15,16]. These paired NLRs have been characterized from both monocots and dicots [17,18••]. Most NLR pairs consist of a canonical signaling NLR, such as RPS4, and a sensor NLR carrying an integrated domain that interacts with an effector target, such as RRS1-R with a WRKY domain (Figure 2) [15,16].

Finally, some receptors require downstream helper NLRs to form a functional unit for disease resistance [19••].

NLR activation and resistosome formation

NLR activity undergoes multilayered regulation, including self-inhibition, dimerization or oligomerization, epigenetic

Figure 2
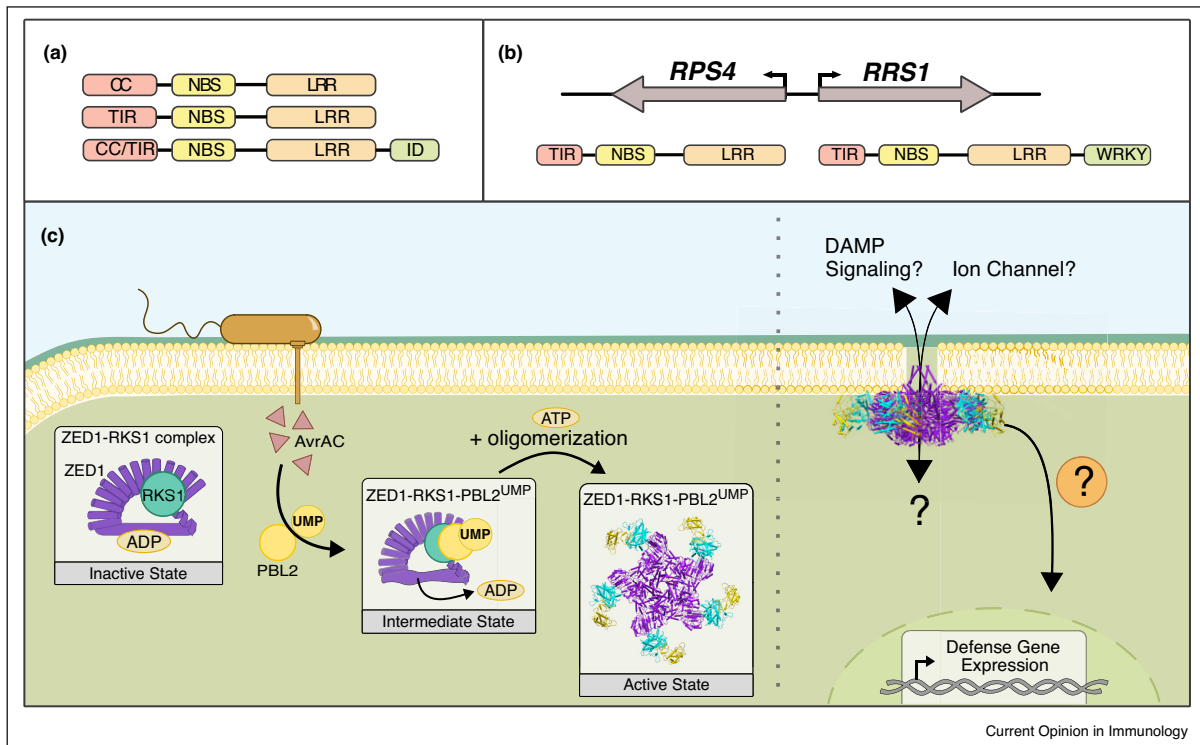
NLRs act as molecular switches to provide robust resistance against pathogens. **(a)** Plant NLR domain architecture includes an N-terminal coiled-coil (CC) or Toll/interleukin-1 receptor (TIR) domain, a central nucleotide-binding site (NBS) and C-terminal leucine-rich repeats (LRRs). Some NLRs carry an integrated domain (ID) that can directly sense pathogen effectors. **(b)** NLRs can act in pairs, such as Arabidopsis *RPS4* and *RRS1*. Top: head-to-head genomic orientation, bottom: domain architecture from N-termini to C-termini. RRS1 is a sensor NLR with a WRKY ID. **(c)** The *Arabidopsis* NLR ZAR1, pseudokinase RSK1, and kinase PBL2 can form a pentameric complex, or resistosome. Upon uridylation by the *Xanthomonas* effector AvrAC, PBL2 is recruited to the ZAR1–RKS1 complex (intermediate state). The active ZAR1 complex exhibits enhanced membrane affinity and the CC-domains of ZAR1 resemble a pore-like structure. Many questions remain regarding the activation of innate immune responses upon resistosome formation.

and transcriptional regulation, alternative splicing and proteasome-mediated degradation [20]. Their similar structure across diverse organisms, especially the presence of the NBS, indicates that nucleotide-binding acts as a switch for receptor activity. Early work in plants identified the first NLRs, demonstrated conservation of NBS domains and determined that the NBS is essential to their functionality [21–23]. If any of the multiple ATP-binding motifs within the NBS are mutated, this renders the NLR either locked in an activated state (ATP bound) or nonfunctional (ADP bound or unbound) in terms of its ability to elicit a defense response [22,24]. A longstanding model within the field of plant immunology posited NLRs are tightly folded and bound to ADP in an inactive state and upon effector perception exhibit conformational changes enabling ATP binding and higher order complex formation (Figure 2) [25,26].

The first structure of a plant NLR complex in inactive, intermediate and activated states was recently elucidated

using cryo–electron microscopy [27**,28**] (Figure 2). This was accomplished with the CNL ZAR1, which recognizes the *Xanthomonas* effector AvrAC indirectly, through effector-mediated uridylation of the host kinase PBL2 [29]. When inactive, ZAR1 self-associates through inter-domain interactions and interacts with the pseudo-kinase RKS1 through its LRR domain [28**]. Upon uridylation, PBL2 recruits and binds to RKS1 [29]. The allosteric binding of PLB2 to RSK1 induces conformational changes in ZAR1's NBS domain, causing release of ADP and formation of a ZAR1–RSK1–PBL2 trimeric complex, likely representing a primed intermediate state [28**]. ZAR1 dATP or ATP binding induces conformational changes within the NBS domain, which in turn mediates oligomerization of the complex into a higher order wheel-like pentamer, called the resistosome [27**]. This multistep activation of ZAR1 may function to ensure appropriate activation of defense. When oligomerized, the N-terminal α-helices of the ZAR1 CC domains form a

protruding funnel-like structure with similarity to pore-forming toxins [27$^{••}$,30] (Figure 2). The N-terminal α-helix is essential for enhanced membrane association and signaling upon ZAR1 activation [27$^{••}$]. Animal NLRs undergo higher order complex formation upon pathogen perception, forming inflammasomes and apoptosomes that trigger cell death [31]. Thus, ATP binding, oligomerization and cell death induction appears to be a common feature of NLR activation.

Resolving the resistosome structure is an important step in understanding mechanisms of NLR activation and opens new avenues for investigating downstream signaling. If activated CNLs are able to form pores in membranes, cell death may occur through disrupting selective membrane permeability in a similar manner to pore forming toxins. Membrane disruption could also induce DAMP signaling and be perceived by PRRs to amplify immune responses [3]. Alternatively, CNL resistosomes could form selective ion channels and transport signaling ions, such as Ca^{2+}, but this would need additional layers of regulation to control ion selectivity (Figure 2). Given the diverse and dynamic NLR sub-cellular localizations, it will be interesting to determine whether other NLRs can form similar structures targeted to various membranes. CNL signaling may be twofold, with pore/ion channel formation coupled to signaling initiated intracellularly through the resistosome complex.

Regulation of downstream signaling
Responses downstream of CNL receptors frequently require the NDR1 locus, while TNL receptors require a set of lipase-like proteins including EDS1 and SAG101 [2,32,33]. NDR1 is anchored to the plasma membrane, mediates plasma-membrane cell wall adhesions and possess similarity to plant proteins involved in abiotic stress responses and mammalian integrins [34]. With the recent discovery that CNLs may form pore-like structures, it will be important to address the role of NDR1 and other immune signaling nodes for effects in plasma membrane integrity. While TNLs lack CC domains, they frequently require downstream helper NLRs of the CNL class, including ADR1, NRG1 and NRCs [19$^{••}$,35,36]. For example, the TNLs Roq1 and RPP1 require the helper NLR NRG1 [37]. Furthermore, multiple CNLs in *Solanaceous* plants that require NRC helpers possess extended N-terminal regions before their CC domain, including the tomato CNLs Prf and Mi [19$^{••}$,38]. Deletion of the N-terminal 13aa from the CC-domain of the NRG1 helper blocks its cell death inducing activity [27$^{••}$]. It is possible that primary NLRs with diverse subcellular localizations or N-terminal domains unable to form pore-like structures partner with helper NLRs to achieve a robust NTI response.

Accumulation of extracellular ROS by NADPH oxidases is a hallmark of both PTI and NTI responses [2,3].

Extracellular ROS also physically strengthens the plant cell wall, induces cell wall depositions and functions as a secondary signal required for both local and systemic innate immune responses [39]. In *Arabidopsis*, the primary NADPH oxidase required for ROS production, RBOHD, is activated by conformational changes induced by Ca^{2+} binding and N-terminal phosphorylation of conserved residues [40–43]. Phosphorylation of RBOHD S343 and S347 occurs during both PTI and NTI, but through distinct kinases [44$^{••}$]. Downstream MAPK cascades are also similarly induced, but it is unknown if the upstream activating kinases are similar for both receptor types [5]. Thus, distinct kinases may converge upon critical signaling nodes with varying intensity to regulate the duration and magnitude of responses during PTI and NTI (Figure 1).

Transcriptional regulation of immunity: overlap with general stress response
NLR recognition of pathogen effectors induces massive transcriptional reprogramming towards defense. Genetic studies have demonstrated that several transcription factor families play critical roles in innate immune and abiotic stress responses, including those in the AP2/ERF, bHLH, MYB, NAC, WRKY, bZIP and CAMTA families [45,46$^{••}$]. Transcriptional profiling after activation of the barley CNL MLA1, the *Arabidopsis* TNL RPS4 and various PRRs recognizing bacterial and fungal PAMPs revealed significant overlap in early response genes [46$^{••}$]. Early response genes are enriched in loci encoding signaling components, such as transcription factors, with CAMTA binding motifs enriched in their promoters [46$^{••}$]. CAMTAs are a group of calmodulin-binding transcription factors involved in both positive and negative regulation of various Ca^{2+}-dependent stress responses [47]. Upon abiotic and biotic stress, CAMTAs rapidly and transiently induce gene expression by binding the Rapid Stress Response Element (RSRE). RSREs are overrepresented in the promoters of general stress response-associated genes [48]. The enrichment of CAMTA-binding motifs in early response genes supports the notion that both innate immune activation and abiotic stress induce a similar and rapid general stress response, with differential transcriptional outputs at later time points depending on pathogen or abiotic stimulus.

Systemic immunity and transgenerational resistance
After pathogen perception, immune signals are subsequently propagated within a tissue, systemically move to distal tissues, and prime the plant for heightened resistance against subsequent attack. Local immune priming can be established by NLR activation as well as crosstalk between PRRs and their co-receptors after MAMP perception [49]. After bacterial challenge, the flagellin co-receptor, BAK1, phosphorylates the receptor-like kinase for chitin perception, CERK1, which

103

primes the plant and enhances defense activation upon subsequent fungal attack [49]. The plant hormone salicylic acid (SA) is required for defense in local and distal tissues. Systemic immunity in distal tissues induces transcriptional and metabolic reprogramming leading to heightened resistance against biotrophic pathogens that typically lasts for several weeks [50,51]. SA-dependent immune priming can be propagated between individual *Arabidopsis* plants through monoterpene emissions [52,53]. This monoterpene-associated response depends on signals associated with systemic resistance, potentially mediating propagation of resistance at a population level [53].

SA application induces chromatin modifications, including acetylation and methylation, on the promoters of defense genes, which correlate with stronger and more robust expression upon pathogen challenge [54–56]. Progeny of *Arabidopsis* infected with *Pseudomonas syringae* or treated with an SA analog displayed stronger induction of SA defense genes and enhanced resistance to *P. syringae* and the oomycete pathogen *Hyaloperonospora arabidopsidis* [54,55]. Furthermore, repeated pathogen challenge within a single generation increased the longevity of transgenerational resistance [54]. A greater mechanistic understanding of the interplay between defense priming, post-translational modifications, epigenetic changes and plant growth can be used to enhance disease resistance and minimize the growth penalty.

Conclusions

Plants represent excellent model systems to study NLR innate immune receptors. Recent evidence has revealed the structure of an NLR complex in various states of activation, demonstrating the formation of the first plant resistosome. Despite differences in defense timing and amplitude between innate immune receptor types, there is overlap in protein phosphorylation and early transcriptional responses. Future research focusing on how diverse NLR receptors induce cell death and resistance upon activation will significantly advance our understanding of this common protein family. Furthermore, given the impact of disease for agricultural production, a comprehensive understanding of NLR biology has significant translational applications for crop improvement.

Conflict of interest statement

Nothing declared.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

•• of outstanding interest

1. Toruno TY, Stergiopoulos I, Coaker G: **Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners**. *Annu Rev Phytopathol* 2016, **54**:419-441.

2. Chiang Y, Coaker G: **Effector triggered immnity: NLR immune perception and downstream defense responses**. *The Arabidopsis Book 11*. The American Society of Plant Biologists; 2015.

3. Couto D, Zipfel C: **Regulation of pattern recognition receptor signalling in plants**. *Nat Rev Immunol* 2016, **16**:537-552.

4. Hartmann M, Zeier J: **N-hydroxypipecolic acid and salicylic acid: a metabolic duo for systemic acquired resistance**. *Curr Opin Plant Biol* 2019, **50**:44-57.

5. Peng Y, van Wersch R, Zhang Y: **Convergent and divergent signaling in PAMP-triggered immunity and effector-triggered immunity**. *Mol Plant Microbe Interact* 2018, **31**:403-409.

6. Giolai M, Paajanen P, Verweij W, Witek K, Jones JDG, Clark MD: **Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire**. *BMC Genomics* 2017, **18**:564.

7. Bai S, Liu J, Chang C, Zhang L, Maekawa T, Wang Q, Xiao W, Liu Y, Chai J, Takken FL *et al.*: **Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance**. *PLoS Pathog* 2012, **8**:e1002752.

8. Shen QH, Saijo Y, Mauch S, Biskup C, Bieri S, Keller B, Seki H, Ulker B, Somssich IE, Schulze-Lefert P: **Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses**. *Science* 2007, **315**:1098-1103.

9. Gao Z, Chung EH, Eitas TK, Dangl JL: **Plant intracellular innate immune receptor resistance to *Pseudomonas syringae* pv. maculicola 1 (RPM1) is activated at, and functions on, the plasma membrane**. *Proc Natl Acad Sci U S A* 2011, **108**:7619-7624.

10. Nimchuk Z, Marois E, Kjemtrup S, Leister RT, Katagiri F, Dangl JL: **Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from *Pseudomonas syringae***. *Cell* 2000, **101**:353-363.

11. Schultink A, Qi T, Lee A, Steinbrenner AD, Staskawicz B: **Roq1 mediates recognition of the Xanthomonas and Pseudomonas effector proteins XopQ and HopQ1**. *Plant J* 2017, **92**:787-795.

12. Liu J, Elmore JM, Lin ZJ, Coaker G: **A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor**. *Cell Host Microbe* 2011, **9**:137-146.

13. Chung EH, da Cunha L, Wu AJ, Gao Z, Cherkis K, Afzal AJ, Mackey D, Dangl JL: **Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants**. *Cell Host Microbe* 2011, **9**:125-136.

14. Keestra AM, Winter MG, Auburger JJ, Frassle SP, Xavier MN, Winter SE, Kim A, Poon V, Ravesloot MM, Waldenmaier JF *et al.*: **Manipulation of small Rho GTPases is a pathogen-induced process detected by NOD1**. *Nature* 2013, **496**:233-237.

15. Sarris PF, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, Derbyshire P, Cevik V, Rallapalli G, Saucet SB *et al.*: **A plant immune receptor detects pathogen effectors that target WRKY transcription factors**. *Cell* 2015, **161**:1089-1100.

16. Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, Zhou B, Levaillant M, Adachi H, Yoshioka H *et al.*: **A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity**. *Cell* 2015, **161**:1074-1088.

17. Kroj T, Chanclud E, Michel-Romiti C, Grand X, Morel JB: **Integration of decoy domains derived from protein targets of**

pathogen effectors into plant immune receptors is
widespread. *New Phytol* 2016, **210**:618-626.

18. Cesari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y,
•• Shimamoto K, Dodds P, Terauchi R, Kroj T: **The NB-LRR proteins
RGA4 and RGA5 interact functionally and physically to confer
disease resistance**. *EMBO J* 2014, **33**:1941-1959.
This study identified two rice NLRs in head-to-head genomic orientation that
work together to recognize and respond to a fungal pathogen effector
protein. The RGA5 NLR acts as a sensor, exhibits an integrated domain and
directly binds the perceived AVR-Pia effector. The RGA4 NLR exhibits
canonical domain architecture and is required for innate immune signaling.

19. Wu CH, Abd-El-Haliem A, Bozkurt TO, Belhaj K, Terauchi R,
•• Vossen JH, Kamoun S: **NLR network mediates immunity to
diverse plant pathogens**. *Proc Natl Acad Sci U S A* 2017,
114:8113-8118.
This study identifies the NRC superclade that arose over 100 million years
ago. Members in the NRC superclade, found in asterids and caryophyl-
lales, act as helper NLRs that are required for the function of multiple
sensor NLRs. Both the NRC family and NRC-dependent NLRs cluster
phylogenetically and enable predictions of which NLRs require down-
stream helpers.

20. Li X, Kapos P, Zhang Y: **NLRs in plants**. *Curr Opin Immunol* 2015,
32:114-121.

21. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S,
Sobral BW, Young ND: **Plant disease resistance genes encode
members of an ancient and diverse protein family within the
nucleotide-binding superfamily**. *Plant J* 1999, **20**:317-332.

22. Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA,
Haring MA, Cornelissen BJ, Takken FL: **Mutations in the NB-ARC
domain of I-2 that impair ATP hydrolysis cause autoactivation**.
Plant Physiol 2006, **140**:1233-1245.

23. Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R,
Giraudat J, Leung J, Staskawicz BJ: **RPS2 of *Arabidopsis
thaliana*: a leucine-rich repeat class of plant disease
resistance genes**. *Science* 1994, **265**:1856-1860.

24. Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL,
Haring MA, Cornelissen BJ: **The tomato R gene products I-2 and
MI-1 are functional ATP binding proteins with ATPase activity**.
Plant Cell 2002, **14**:2929-2939.

25. Moffett P, Farnham G, Peart J, Baulcombe DC: **Interaction
between domains of a plant NBS-LRR protein in disease
resistance-related cell death**. *EMBO J* 2002, **21**:4511-4519.

26. Takken FL, Albrecht M, Tameling WI: **Resistance proteins:
molecular switches of plant defence**. *Curr Opin Plant Biol* 2006,
9:383-390.

27. Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, Qi Y, Wang HW,
•• Zhou JM, Chai J: **Reconstitution and structure of a plant NLR
resistosome conferring immunity**. *Science* 2019, **364**:6435
eaav5870.
This study reconstituted an active plant NLR innate immune complex. The
complex oligomerizes upon activation and cryo-electron microscopy revealed
a pentameric resistosome complex. The N-terminal alpha-helix of the NLR
ZAR1 exhibits a funnel-shaped structure upon resistosome formation.

28. Wang J, Wang J, Hu M, Wu S, Qi J, Wang G, Han Z, Qi Y, Gao N,
•• Wang HW *et al.*: **Ligand-triggered allosteric ADP release
primes a plant NLR complex**. *Science* 2019, **364**:6435 eaav5868.
This study reports cryo-electron microscopy structures of a plant NLR
innate immune complex in an inactive and intermediate state, revealing
differences in domain partitioning between plant and animal NLRs.
Pathogen effector modification of the PBL2 kinase recruits this kinase
to the NLR complex, which initiates receptor activation.

29. Wang G, Roux B, Feng F, Guy E, Li L, Li N, Zhang X, Lautier M,
Jardinaud MF, Chabannes M *et al.*: **The decoy substrate of a
pathogen effector and a pseudokinase specify pathogen-
induced modified-self recognition and immunity in plants**. *Cell
Host Microbe* 2015, **18**:285-295.

30. Los FC, Randis TM, Aroian RV, Ratner AJ: **Role of pore-forming
toxins in bacterial infectious diseases**. *Microbiol Mol Biol Rev*
2013, **77**:173-207.

31. Chai J, Shi Y: **Apoptosome and inflammasome: conserved
machineries for caspase activation**. *Natl Sci Rev* 2014, **1**:
101-118.

32. Feys BJ, Wiermer M, Bhat RA, Moisan LJ, Medina-Escobar N,
Neu C, Cabral A, Parker JE: **Arabidopsis SENESCENCE-
ASSOCIATED GENE101 stabilizes and signals within an
ENHANCED DISEASE SUSCEPTIBILITY1 complex in plant
innate immunity**. *Plant Cell* 2005, **17**:2601-2613.

33. Century KS, Holub EB, Staskawicz BJ: **NDR1, a locus of
Arabidopsis thaliana that is required for disease resistance to
both a bacterial and a fungal pathogen**. *Proc Natl Acad Sci U S A*
1995, **92**:6597-6601.

34. Knepper C, Savory EA, Day B: **Arabidopsis NDR1 is an integrin-
like protein with a role in fluid loss and plasma membrane-cell
wall adhesion**. *Plant Physiol* 2011, **156**:286-300.

35. Peart JR, Mestre P, Lu R, Malcuit I, Baulcombe DC: **NRG1, a CC-
NB-LRR protein, together with N, a TIR-NB-LRR protein,
mediates resistance against tobacco mosaic virus**. *Curr Biol*
2005, **15**:968-973.

36. Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL:
**Expanded functions for a family of plant intracellular immune
receptors beyond specific recognition of pathogen effectors**.
Proc Natl Acad Sci U S A 2011, **108**:16463-16468.

37. Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, Cho MJ,
Schultink A, Staskawicz BJ: **NRG1 functions downstream of
EDS1 to regulate TIR-NLR-mediated plant immunity in
*Nicotiana benthamiana***. *Proc Natl Acad Sci U S A* 2018, **115**:
E10979-E10987.

38. Wu CH, Kamoun W: **Tomato Prf requires NLR helpers NRC2
and NRC3 to confer resistance against the bacterial speck
pathogen *Pseudomonas syringae* pv. tomato**. *bioRxiv* 2019.
595744.

39. Kadota Y, Shirasu K, Zipfel C: **Regulation of the NADPH
Oxidase RBOHD during plant immunity**. *Plant Cell Physiol*
2015, **56**:1472-1480.

40. Kadota Y, Sklenar J, Derbyshire P, Stransfeld L, Asai S,
Ntoukakis V, Jones JD, Shirasu K, Menke F, Jones A *et al.*: **Direct
regulation of the NADPH oxidase RBOHD by the PRR-
associated kinase BIK1 during plant immunity**. *Mol Cell* 2014,
54:43-55.

41. Dubiella U, Seybold H, Durian G, Komander E, Lassig R,
Witte CP, Schulze WX, Romeis T: **Calcium-dependent protein
kinase/NADPH oxidase activation circuit is required for rapid
defense signal propagation**. *Proc Natl Acad Sci U S A* 2013,
110:8744-8749.

42. Zhang M, Chiang YH, Toruno TY, Lee D, Ma M, Liang X, Lal NK,
Lemos M, Lu YJ, Ma S *et al.*: **The MAP4 kinase SIK1 ensures
robust extracellular ROS burst and antibacterial immunity in
plants**. *Cell Host Microbe* 2018, **24**:379-391.e375.

43. Li L, Li M, Yu L, Zhou Z, Liang X, Liu Z, Cai G, Gao L, Zhang X,
Wang Y *et al.*: **The FLS2-associated kinase BIK1 directly
phosphorylates the NADPH oxidase RbohD to control plant
immunity**. *Cell Host Microbe* 2014, **15**:329-338.

44. Kadota Y, Liebrand TWH, Goto Y, Sklenar J, Derbyshire P,
•• Menke FLH, Torres MA, Molina A, Zipfel C, Coaker G *et al.*:
**Quantitative phosphoproteomic analysis reveals common
regulatory mechanisms between effector- and PAMP-
triggered immunity in plants**. *New Phytol* 2019, **221**:2160-2175.
This study investigated protein phosphorylation dynamics upon NLR
activation, revealing novel phosphorylated targets and overlap in phos-
phorylated proteins during PRR activation. While phosphorylation
occurred on similar residues of the NADPH oxidase, different kinases
are involved based on receptor type.

45. Ng DW, Abeysinghe JK, Kamali M: **Regulating the regulators: the
control of transcription factors in plant defense signaling**. *Int J
Mol Sci* 2018, **19**:E3737.

46. Jacob F, Kracher B, Mine A, Seyfferth C, Blanvillain-Baufume S,
•• Parker JE, Tsuda K, Schulze-Lefert P, Maekawa T: **A dominant-
interfering camta3 mutation compromises primary
transcriptional outputs mediated by both cell surface and
intracellular immune receptors in *Arabidopsis thaliana***. *New
Phytol* 2018, **217**:1667-1680.
This study compared the transcriptional response upon activation of plant
innate immunity and abiotic stress responses from over 40 datasets.
Abiotic stress and innate immune activation trigger a similar and rapid

105

general stress response, whose transcriptional outputs then diverge at later timepoints depending on the initial stimulus.

47. Bjornson M, Benn G, Song X, Comai L, Franz AK, Dandekar AM, Drakakaki G, Dehesh K: **Distinct roles for mitogen-activated protein kinase signaling and CALMODULIN-BINDING TRANSCRIPTIONAL ACTIVATOR3 in regulating the peak time and amplitude of the plant general stress response**. *Plant Physiol* 2014, **166**:988-996.

48. Benn G, Wang CQ, Hicks DR, Stein J, Guthrie C, Dehesh K: **A key general stress response motif is regulated non-uniformly by CAMTA transcription factors**. *Plant J* 2014, **80**:82-92.

49. Gong BQ, Guo J, Zhang N, Yao X, Wang HB, Li JF: **Cross-microbial protection via priming a conserved immune co-receptor through juxtamembrane phosphorylation in plants**. *Cell Host Microbe* 2019, **26**:810-822.

50. Ross AF: **Localized acquired resistance to plant virus infection in hypersensitive hosts**. *Virology* 1961, **14**:329-339.

51. Bernsdorff F, Doring AC, Gruner K, Schuck S, Brautigam A, Zeier J: **Pipecolic acid orchestrates plant systemic acquired resistance and defense priming via salicylic acid-dependent and -independent pathways**. *Plant Cell* 2016, **28**:102-129.

52. Riedlmeier M, Ghirardo A, Wenig M, Knappe C, Koch K, Georgii E, Dey S, Parker JE, Schnitzler JP, Vlot AC: **Monoterpenes support systemic acquired resistance within and between plants**. *Plant Cell* 2017, **29**:1440-1459.

53. Wenig M, Ghirardo A, Sales JH, Pabst ES, Breitenbach HH, Antritter F, Weber B, Lange B, Lenk M, Cameron RK *et al.*: **Systemic acquired resistance networks amplify airborne defense cues**. *Nat Commun* 2019, **10**:3813.

54. Luna E, Bruce TJ, Roberts MR, Flors V, Ton J: **Next-generation systemic acquired resistance**. *Plant Physiol* 2012, **158**:844-853.

55. Slaughter A, Daniel X, Flors V, Luna E, Hohn B, Mauch-Mani B: **Descendants of primed Arabidopsis plants exhibit resistance to biotic stress**. *Plant Physiol* 2012, **158**:835-843.

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

APPENDEX

Variation in microbial feature perception in the Rutaceae family with immune receptor conservation in citrus

Jessica Trinh*[1], Tianrun Li*[1], Jessica Y. Franco*[1], Tania Y. Toruño*[1], Danielle M. Stevens*[1], Shree P. Thapa[1], Justin Wong[1], Rebeca Pineda[2], Emmanuel Ávila de Dios[2], Tracy L. Kahn[2], Danelle K. Seymour[2], Chandrika Ramadugu[2], Gitta L. Coaker[1†]

[†] Corresponding author: Gitta Coaker, glcoaker@ucdavis.edu

*These authors contributed equally.

Abstract

Although much is known about the responses of model plants to microbial features, we still lack an understanding of the extent of variation in immune perception across members of a plant family. In this work, we analyzed immune responses in *Citrus* and wild relatives, surveying 86 Rutaceae genotypes with differing leaf morphologies and disease resistances. We found that responses to microbial features vary both within and between members. Species in two subtribes, the Balsamocitrinae and Clauseninae, can recognize flagellin (flg22), cold shock protein (csp22) and chitin, including one feature from *Candidatus* Liberibacter species (csp22$_{C\text{Las}}$), the bacterium associated with Huanglongbing. We investigated differences at the receptor level for the flagellin receptor FLAGELLIN SENSING 2 (FLS2) and the chitin receptor LYSIN MOTIF RECEPTOR KINASE 5 (LYK5) in citrus genotypes. We characterized two genetically linked *FLS2* homologs from 'Frost Lisbon' lemon (*Citrus ×limon*, responsive) and 'Washington navel' orange (*Citrus ×aurantium*, non-responsive). Surprisingly, *FLS2* homologs from responsive and non-responsive

genotypes were expressed in *Citrus* and functional when transferred to a heterologous system. 'Washington navel' orange weakly responded to chitin, whereas 'Tango' mandarin (*Citrus ×aurantium*) exhibited a robust response. LYK5 alleles were identical or nearly identical between the two genotypes and complemented the Arabidopsis (*Arabidopsis thaliana*) *lyk4/lyk5-2* mutant with respect to chitin perception. Collectively, our data indicate that differences in chitin and flg22 perception in these citrus genotypes are not the results of sequence polymorphisms at the receptor level. These findings shed light on the diversity of perception of microbial features and highlight genotypes capable of recognizing polymorphic pathogen features.

Author Contributions

I wrote code to plot the data in Figure 1A, 2A, 8B, and collected, analyzed, and plotted the data for Figure 5. I also wrote the of the section: Cultivated citrus genotypes contain functional chitin receptor homologs and Phylogenetic analyses and receptor comparisons. All authors were involved in editing the manuscript.

Plant Physiology®

Variation in microbial feature perception in the Rutaceae family with immune receptor conservation in citrus

Jessica Trinh [ID],[1,†] Tianrun Li [ID],[1,†] Jessica Y. Franco [ID],[1,†] Tania Y. Toruño [ID],[1,†] Danielle M. Stevens [ID],[1,†] Shree P. Thapa [ID],[1] Justin Wong [ID],[1] Rebeca Pineda [ID],[2] Emmanuel Ávila de Dios [ID],[2] Tracy L. Kahn,[2] Danelle K. Seymour [ID],[2] Chandrika Ramadugu [ID][2] and Gitta L. Coaker [ID][1,*]

1   Department of Plant Pathology, University of California, Davis, CA 95616, USA
2   Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

*Author for correspondence: glcoaker@ucdavis.edu
†These authors contributed equally.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Gitta Coaker.

Abstract

Although much is known about the responses of model plants to microbial features, we still lack an understanding of the extent of variation in immune perception across members of a plant family. In this work, we analyzed immune responses in *Citrus* and wild relatives, surveying 86 Rutaceae genotypes with differing leaf morphologies and disease resistances. We found that responses to microbial features vary both within and between members. Species in 2 subtribes, the Balsamocitrinae and Clauseninae, can recognize flagellin (flg22), cold shock protein (csp22), and chitin, including 1 feature from *Candidatus* Liberibacter species (csp22$_{CLas}$), the bacterium associated with Huanglongbing. We investigated differences at the receptor level for the flagellin receptor FLAGELLIN SENSING 2 (FLS2) and the chitin receptor LYSIN MOTIF RECEPTOR KINASE 5 (LYK5) in citrus genotypes. We characterized 2 genetically linked *FLS2* homologs from "Frost Lisbon" lemon (*Citrus* ×*limon*, responsive) and "Washington navel" orange (*Citrus* ×*aurantium*, nonresponsive). Surprisingly, *FLS2* homologs from responsive and nonresponsive genotypes were expressed in *Citrus* and functional when transferred to a heterologous system. "Washington navel" orange weakly responded to chitin, whereas "Tango" mandarin (*C.* ×*aurantium*) exhibited a robust response. LYK5 alleles were identical or nearly identical between the 2 genotypes and complemented the *Arabidopsis* (*Arabidopsis thaliana*) *lyk4/lyk5-2* mutant with respect to chitin perception. Collectively, our data indicate that differences in chitin and flg22 perception in these citrus genotypes are not the results of sequence polymorphisms at the receptor level. These findings shed light on the diversity of perception of microbial features and highlight genotypes capable of recognizing polymorphic pathogen features.

Introduction

The perception of microbial features has typically been assessed by using a single or few plant genotypes to make conclusions about perception. Recognition of conserved features of pathogens, known as microbe-associated molecular patterns (MAMPs), activates the plant immune system. MAMPs can be proteinaceous or structural pathogen features and are perceived by plant immune receptors. While many studies are focused on the immune responses of 1 representative genotype, responses to MAMPs exhibit variation within and between related species. For example, different genotypes of *Arabidopsis thaliana* contain FLAGELLIN SENSING 2 (FLS2) homologs that vary in binding specificity to an epitope of bacterial flagellin, and low binding specificity was associated with high bacterial proliferation (Vetter et al. 2012). Epitopes from 3 bacterial MAMPs were differentially

recognized across heirloom tomato genotypes, indicating the diversity in immune responses within a group of closely related plants (Veluchamy et al. 2014). The extent of natural variation in immune responses to microbial features remains largely unexplored.

The Rutaceae plant family contains ∼2,100 species with worldwide distribution, including the agriculturally important genus *Citrus* (Kubitzki et al. 2011). *Citrus* is the most extensively produced fruit crop in the world with 124.246 million tons of fruit produced in 2016 (Zhong and Nicolosi 2020). In the United States, the 2019 to 2020 growing season yielded a production value of ∼3.4 billion dollars for citrus products (USDA 2020). Florida alone is the second largest producer of orange juice in the world behind Brazil, with its citrus economy contributing billions of dollars to the state gross domestic product (Hodges and Spreen 2012). While oranges constitute more than half of worldwide citrus production, other relevant citrus products include tangerines, limes, lemons, and grapefruits (Liu et al. 2012). These different varieties of cultivated citrus are members of the genus *Citrus* in the family of Rutaceae, which contains several noncultivated *Citrus* relatives (Wu et al. 2018). Several systems of classification exist; in this study, we followed the classification of Swingle and Reece (1967). Rutaceae contains 6 subfamilies (Appelhans et al. 2021), and the subfamily to which citrus belongs, Aurantioideae, contains 2 tribes: Citreae and Clauseneae (Morton 2009). The Citreae contains 3 subtribes, which are Citrinae, Triphasiinae, and Balasmocitrinae. Clauseneae contains 3 subtribes: Micromelinae, Clauseninae, and Merrillinae (Nagano et al. 2018). The Citrinae subtribe contains all cultivated citrus genotypes and is the most economically important group in the Rutaceae family (Swingle and Reece 1967).

Cultivated citrus is a perennial crop that is vegetatively propagated through grafting (Castle 2010; Caruso et al. 2020). While asexual propagation methods maintain the desired combinations of traits in commercial cultivars, they prevent the exchange of genetic material (Uzun and Yesiloglu 2012; Wang et al. 2017). Because of this, crops that are primarily propagated asexually are susceptible to devastating impacts from newly introduced citrus diseases. Cultivated citrus varieties are susceptible to a variety of microbial pathogens including bacteria, filamentous fungi/oomycetes, and viruses. Breeding efforts often focus on developing rootstocks with resistance to these pathogens to fend off disease in the clonally propagated scion. Examples of citrus diseases with a substantial impact on production include citrus Huanglongbing (HLB) (Bové 2006; Wang 2019), citrus canker (Das 2003; Ference et al. 2018), citrus variegated chlorosis (Coleta-Filho et al. 2020), and fruit and root rots (Jaouad et al. 2020).

To protect themselves from pathogens, plants have evolved multiple defense mechanisms including MAMP perception. MAMPs can be proteinaceous, such as the flagellin or cold shock protein of bacteria (Wang et al. 2016) or not, such as bacterial lipopolysaccharide or fungal chitin (Newman et al. 2013). To detect MAMPs, plants possess pattern recognition receptors (PRRs) on the surface of their cells. PRRs include receptor-like kinases (RLKs) and receptor-like proteins (RLPs). RLKs consist of an extracellular domain, a transmembrane domain, and an intracellular kinase domain, whereas RLPs lack the intracellular kinase domain. Examples of PRR extracellular domains include leucine-rich repeats (LRRs), lysin motifs (LysM), and lectin domains, among others (Ngou et al. 2022). Binding of the MAMP to the PRR often results in heterodimer formation with a coreceptor to activate downstream signaling responses and host defenses ultimately leading to MAMP–triggered immunity (MTI). Hallmarks of MTI activation include apoplastic reactive oxygen species (ROS) production, intracellular mitogen-activated protein kinase (MAPK) activation, calcium influx, and global transcriptional reprogramming (Jeworutzki et al. 2010; Bigeard et al. 2015; Couto and Zipfel 2016; Saijo et al. 2018).

Extensive research has been performed in the last few decades to reveal PRRs that perceive various MAMPs in model and crop plants (Ngou et al. 2022). Some well-characterized receptors include the following: the *Arabidopsis* FLS2 receptor for a 22-amino acid epitope of bacterial flagellin (flg22; Gómez-Gómez and Boller 2000), the *Arabidopsis* LysM domain receptor (LYK4/5) for chitin (Miya et al. 2007; Cao et al. 2014), and the tomato (*Solanum lycopersicum*) CORE receptor for a 22-amino acid epitope of bacterial cold shock protein (csp22; Wang et al. 2016). However, more work needs to be done to discover immune receptors in tree crops and other nonmodel plants. Genome mining for citrus LRR-RLKs suggests that there are receptors capable of mediating host–pathogen interactions. Although RLKs have been predicted in the *Citrus* ×*aurantium* and *Citrus clementina* genomes, no immune receptors have been functionally validated in citrus (Magalhães et al 2016; Dalio et al. 2017). Previous studies have identified citrus FLS2 homologs, 1 of which is induced in response to bacterial flagellin (Shi et al. 2016). In addition, 1 Liberibacter-specific MAMP for the bacterial protein pksG was recognized in 3 out of 10 citrus genotypes (Chen et al. 2020).

Here, we have examined the responses of several genotypes encompassing both cultivated citrus and wild relatives to different microbial features in order to better understand the landscape of immune perception within the Rutaceae family. FLS2 orthologs are present in both monocots and dicots. We identified 2 *FLS2* homologs that are nearly identical from responding and nonresponding citrus genotypes. Surprisingly, *FLS2* homologs from responding and nonresponding genotypes were functional when transferred to a heterologous system, indicating that impaired flagellin perception is not due to differences at the receptor level. Most cultivated citrus and wild relatives can perceive chitin, and we were able to isolate a citrus homolog of the *Arabidopsis* chitin receptor *LYK5* and demonstrate its functionality in a nonhost species. We also identified citrus relatives that can perceive a conserved feature of *Candidatus* Liberibacter asiaticus (CLas), the bacterium associated with citrus HLB. These results highlight the importance of studying immunity in wild relatives, especially to identify potential genotypes with immune

mechanisms that can be transferred to disease-susceptible cultivars.

Results

Members of the Rutaceae family exhibit diversity in the perception of and magnitude of response to microbial features

To investigate the immune response capabilities of members of the Rutaceae family, we screened 86 genotypes for the perception of 3 common microbial features: chitin, flg22, and csp22. These genotypes were samples from the Givaudan Citrus Variety Collection (GCVC) at the University of California, Riverside. The GCVC is 1 of the most comprehensive collections of citrus diversity, including over 1,000 accessions that span the genus *Citrus* and related genera. This study includes representatives spanning known subtribes in Rutaceae, including both cultivated citrus and wild relatives. In total, we screened individuals from 2 subfamilies (Aurantioideae and Zanthoxyloideae) and representative taxa from all 6 subtribes of Aurantioideae, comprising over 30 different genera. The majority of selected genotypes fall within the Citrinae subtribe (56 genotypes), which includes the cultivated citrus types. Genotypes are referred to by their common name, if available, with the corresponding scientific name and accession number in Supplemental Table S1. To measure the ROS output of multiple genotypes, we have optimized a plate-based assay for high-throughput screening of leaves from both seedlings and mature trees. The genotypes we screened exhibit a variety of different leaf, branch, and fruit morphologies (Supplemental Table S1 and Fig. 1B). A luminol analog, L-012, chemically reacts with horseradish peroxidase and ROS to produce light, which is measured by the plate reader as relative light units (RLUs). RLUs from a ROS burst can be plotted as a curve over time, area under the curve, or in this case, the peak of ROS production (max RLU). The results from the ROS–based assay are presented as an average of max RLUs across multiple independent experiments in Fig. 1A.

The landscape of immune perception varied across genotypes, with the strength of the response to each elicitor segregating across members from most surveyed tribes (Figs. 1A, S1, and S2). There were differences in the proportion of genotypes responding to each elicitor. For example, nearly all genotypes screened are capable inducing ROS in response to chitin (75 out of 86 genotypes), but less than half of the screened genotypes are capable of inducing ROS in response to flg22 (40 out of 86 genotypes). More than half of the screened genotypes (45 out of 86 genotypes) are capable of inducing ROS in response to csp22. The luminol assay used for ROS production can also be impacted by secondary compounds, including phenolics, antioxidants, or reducing agents (Plieth 2018). The majority of genotypes were able to perceive chitin (76 out of 86), indicating that secondary compounds in leaves may not grossly affect ROS production in

these genotypes (Figs. 1 and S1). Because of widespread chitin response across cultivated citrus and wild relatives, it is likely that these genotypes share a conserved chitin receptor or multiple receptors capable of perceiving chitin of different lengths. There is substantial segregating variation in MAMP response across tribes. For example, within the Citrinae tribe, 'Tango' mandarin (*C. ×aurantium*) can respond to all 3 MAMPs, but kumquat (*Fortunella hindsii*) cannot respond to any of the 3. Twenty-four genotypes are capable of responding to all 3 MAMPs in addition to 'Tango' mandarin. Closely related genotypes, such as 'Tango' mandarin and 'Lee' mandarin, also have varying responses to MAMPs: 'Tango' mandarin responds to all 3 MAMPs, but 'Lee' mandarin can only respond to chitin (Figs. 1A and S1). This variation may be the result of differences at the receptor level or in downstream signaling components.

Rutaceae genotypes vary not only in their ability to respond to MAMPs but also in the magnitude of ROS production (Figs. 1, 2, S1, and S2). The magnitude of ROS production as a result of chitin, flg22, or csp22 treatment occurs across tribes as well as within members of a tribe. Although the majority of the screened Rutaceae genotypes are capable of perceiving chitin, some genotypes produce an average max RLU of <1,000, while others produce an average max RLU well over 10,000 in response to chitin (Supplemental Fig. S1). To categorize the strength in responses, 25th and 75th quartiles were computed for each MAMP and cutoffs were used. Figure 2A shows examples of "weak" (25th percentile or below), "medium" (between the 25th and 75th percentiles), and "strong" (75th percentile or greater) responders. Across tribes, we see that members of the Triphasiinae tribe, such as the trifoliate limeberry (*Triphasia trifolia*), are strong ROS responders to chitin, whereas some members of the Balsamocitrinae subtribe, such as the Chevalier's *Aeglopsis* (*Aeglopsis chevalieri*), are either weak or medium ROS responders (Fig. 2B). Within the Citrinae tribe, 'Tango' mandarin and Uganda cherry orange (*Citropsis schweinfurthii*) are strong responders to flg22, but 'King' tangor (*C. ×aurantium*) is a weak responder to flg22 (Fig. 2C). Trifoliate limeberry is also a strong responder to csp22 (Fig. 2D). The data showcase strong and weak ROS responders to MAMPs that are spread out within and between taxonomic groups.

In addition to the production of ROS, other common immune responses include MAPK activation, defense gene expression, and callose deposition. One of the challenges of studying Rutaceae is that many of the genotypes in this study do not have their genomes sequenced, making primer design for defense gene expression experiments difficult. Additionally, the thick, waxy leaves of citrus plants make it challenging to visualize callose deposition via microscopy. MAPKs are highly conserved across eukaryotes (Meng and Zhang 2013), making them viable immune markers to study MAMP responses in a variety of genotypes. MAPKs are phosphorylated upon MAMP perception, which can be detected via western blot. MAPK phosphorylation can be weakly induced in response to water or buffer treatment but is
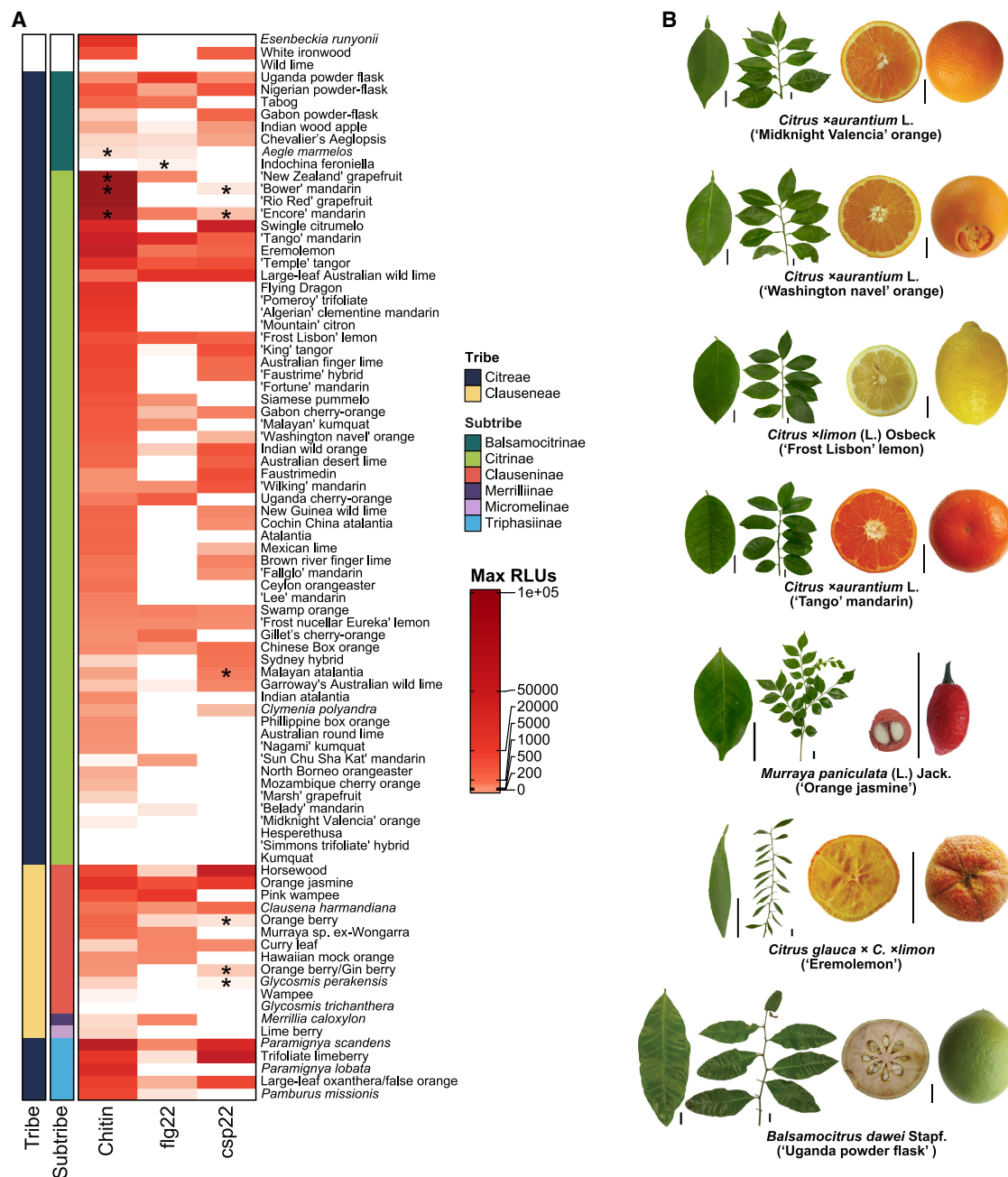
Figure 1. Genotypes within the Rutaceae family, including citrus, exhibit diverse responses to common MAMPs and possess differing leaf morphologies. **A)** Heat map compiling average max RLUs from ROS assays in genotypes within the Rutaceae family, organized by MAMP and phylogenetic relationship. Max RLUs are averages of at least 3 independent experiments and are represented as a heatmap, where *max RLU* = (*max RLU MAMP* − *max RLU water*). The threshold for no response is <90 RLUs. Asterisks indicate genotypes that exhibit a variable response, where 1 or 2 independent experiments shared a response. The MAMPs used are canonical features in the following concentrations: chitin (10 μM), flg22 (100 nM), and csp22 (100 nM). **B)** Leaf, branch, and fruit morphologies of selected genotypes grown under greenhouse and field conditions. The images were digitally extracted for comparison. Scale bars = 2 cm. Genotypes are referred to by common name unless otherwise unavailable.
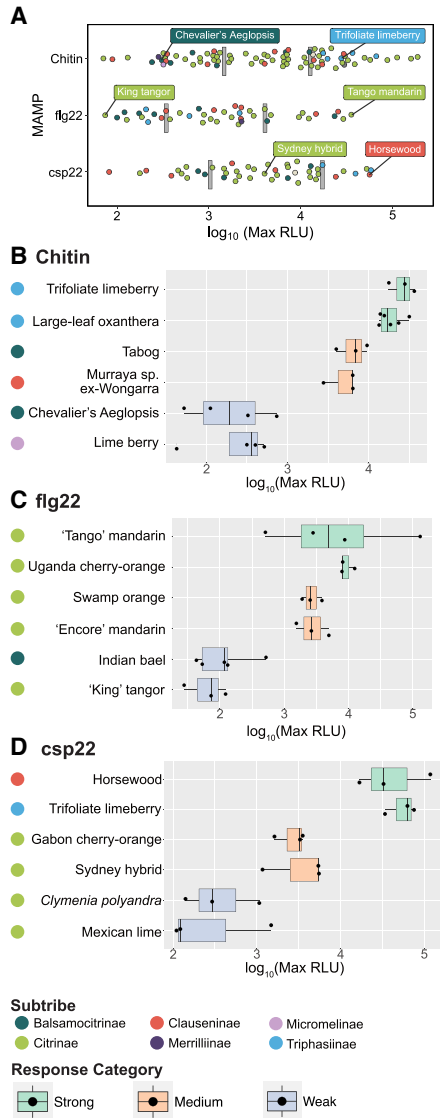
Figure 2. Rutaceae genotypes vary in the magnitude of their responses to perception of chitin, flagellin, and cold shock protein immunogenic epitopes. **A)** Distribution of all average max RLU values, with gray lines indicating 25th and 75th percentile markings. Box plots below are organized by the magnitude of their responses to chitin **B)**, flg22 **C)**, and csp22 **D)**. The MAMPs used are canonical features in the following concentrations: chitin (10 μM), flg22 (100 nM), and csp22 (100 nM). Max RLUs are averages of at least 3 independent experiments, where *max RLU = (max RLU MAMP − max RLU water)*. Data points on box plots represent the average max RLU for an individual experiment, with $n = 8$ leaf disks per experiment. Criteria for the response categories: "strong" responders are in the top 25th percentile, "medium" responders are between the 25th and 75th percentiles, and "weak" responders are in the bottom 25th percentile. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range.

strongly phosphorylated in response to immune activation (Asai et al. 2002; Zhang et al. 2013).

In order to investigate activation of other immune responses, we analyzed a subset of Rutaceae genotypes from 3 different subtribes (Balsamocitrinae, Clauseninae, and Citrinae) for MTI–induced MAPK activation in response to flg22 and chitin treatment. 'Midknight Valencia' orange (*C. ×aurantium*) is only able to respond to chitin based on ROS results and only exhibits MAPK phosphorylation upon chitin treatment (Fig. 3). 'Frost Lisbon' lemon (*Citrus ×limon*) responds to chitin and flg22 based on ROS results and induces sustained MAPK phosphorylation in response to chitin and flg22 treatment. For Orange jasmine and Uganda powder flask, 2 non-Citrinae genotypes, we can observe ROS production in response to chitin and flg22. While chitin-induced MAPK phosphorylation was robust in both genotypes, flg22 perception was only observed in 2 out of 4 MAPK trials for the Uganda powder flask (*Balsamocitrus dawei*). Orange jasmine (*Murraya paniculata*) exhibited weak, but reproducible, MAPK phosphorylation after flg22 treatment. Taken together, these data indicate that Rutaceae genotypes can respond to MAMPs by inducing ROS and MAPK activation.

Functional analyses of FLS2 in flagellin-responsive and nonresponsive citrus genotypes

FLS2 orthologs have been identified and functionally validated from diverse plant families including Brassicaceae (Gómez-Gómez and Boller 2000), Solanaceae (Robatzek et al. 2007), Vitaceae (Trdá et al. 2014), and Poaceae (Takai et al. 2008). Of the 86 genotypes, we surveyed, 41 were able to perceive flg22 (Figs. 1A and S1). Sweet orange genotypes, including 'Midknight Valencia' orange and 'Washington navel' orange (*C. ×aurantium*), were not able to elicit a ROS response to flg22, unlike the cultivated lemon genotypes 'Frost Lisbon' lemon and 'Frost nucellar Eureka' lemon (*C. ×limon*) (Figs. 1A and 4A). 'Washington navel' orange and 'Frost Lisbon' lemon are 2 widely grown citrus genotypes. Therefore, we investigated their response to flg22 in more detail. MAPK assays after treatment with flg22 verified that 'Washington navel' orange could not respond, while 'Frost Lisbon' lemon was able to induce MAPK phosphorylation (Fig. 4B). Similarly, flg22 treatment induced expression of the defense marker gene *WRKY22* in 'Frost Lisbon' lemon but not 'Washington navel' orange (Fig. 4C). Previously, 2 *FLS2* homologs (*FLS2-1* and *FLS2-2*) from 'Duncan' grapefruit (*C. ×aurantium*) and 'Sun Chu Sha Kat' mandarin (*C. ×aurantium*) were identified and demonstrated to be genetically linked (Shi et al. 2016). Similarly, when we analyzed the genomes of 'Washington navel' orange and 'Frost Lisbon' lemon, we identified *FLS2-1* and *FLS2-2* in syntenic chromosomal regions on haplotype 2 (Figs. 4D and S3). Interestingly, *FLS2-2* is absent in 'Frost Lisbon' lemon haplotype 1 and truncated in 'Washington navel' orange haplotype 1 (Fig. 4D). Due to high conservation between *FLS2-1* alleles from each haplotype, we were unable to distinguish their transcripts by reverse transcription
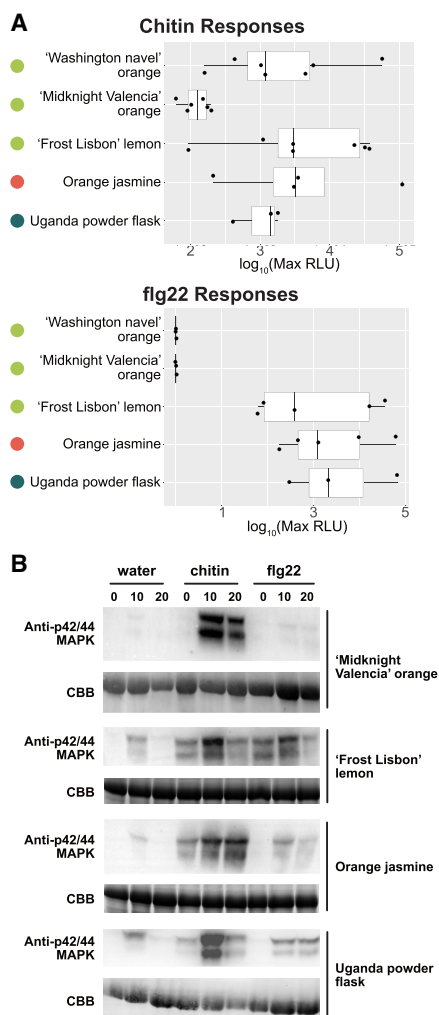
A

B

Figure 3. ROS and MAPK induction in response to MAMP treatment in cultivated citrus and wild relatives. **A)** Box plots showing the average max RLUs of selected Rutaceae genotypes in response to chitin and flg22. The MAMPs used are canonical features in the following concentrations: chitin (10 μM) and flg22 (100 nM). Max RLUs are averages of at least 3 independent experiments, where *max RLU = (max RLU MAMP − max RLU water)*. Data points on box plots represent the average max RLU for an individual experiment, with *n* = 8 leaf disks per experiment. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range. **B)** MAPK induction visualized at 0-, 10-, and 20-min post-induction with water or MAMP. MAMPs are applied to leaf punches at the following concentrations for MAPK assays: chitin (10 μM) and flg22 (100 nM), with water as a negative control. Western blots are performed with an anti-p42/44 MAPK antibody to visualize the MAPK bands and Coomassie Brilliant Blue (CBB) to verify equal loading of protein samples. All experiments were performed at least 3×; flg22 perception was not observed in 2 out of 4 trials for the Uganda powder flask.

quantitative PCR (RT-qPCR). Both genotypes exhibited low, but detectable baseline expression of each homolog using RT-qPCR (Supplemental Fig. S4). *FLS2-1* and *FLS2-2* had higher baseline transcript expression in 'Washington navel' orange, indicating that differential responsiveness is not due to resting-state expression (Supplemental Fig. S4). In both citrus genotypes, *FLS2-2* expression was induced after treatment with flg22, with 'Frost Lisbon' lemon exhibiting stronger induction (Fig. 4E). The immune responses (ROS and MAPK) we measured occur within 10-min post-MAMP treatment, and both citrus genotypes express *FLS2-1* and *FLS2-2* in the absence of flg22 perception. Therefore, it is unlikely that differences in early immune outputs would be regulated by de novo transcription of *FLS2* PRRs.

To test if compromised flagellin perception in 'Washington navel' orange is due to sequence polymorphisms in *FLS2*, we investigated the ability of each homolog to perceive flg22 using transcomplementation experiments in *Nicotiana benthamiana*. We used virus-induced gene silencing (VIGS) to silence endogenous *FLS2* in *N. benthamiana*, followed by *Agrobacterium*-mediated transient expression of *Arabidopsis FLS2* as well as citrus *FLS2-1* and *FLS2-2*. Forty-eight hours after transient expression, we assayed silenced plants for their ability to induce a ROS burst in response to flg22 treatment. As expected, *N. benthamiana FLS2*-silenced lines were unable to elicit a flg22-induced ROS burst, but *GUS*-silenced lines were able to perceive flg22 (Fig. 4F). Expression of *Arabidopsis* FLS2, 'Frost Lisbon' lemon FLS2-1 and FLS2-2, led to a ROS burst in response to flg22 and detectable using anti-HA western blotting (Fig. 4, F and G). Expression of 'Washington navel' orange FLS2-1, and to a lesser extent, FLS2-2, was also able to elicit a ROS burst in response to flg22 (Fig. 4F). While 'Washington navel' orange FLS2-1 was robustly expressed by western blot, FLS2-2 exhibited lower level expression, which may explain its reduced ROS burst (Fig. 4G). These data are consistent with the near identical amino acid similarity between FLS2-1 (99.16% to 99.41%) and FLS2-2 (97.41%) (Supplemental Fig. S3). Collectively, these results suggest that differences in flg22-mediated responses between both genotypes may not be regulated at the receptor level.

Cultivated citrus genotypes contain functional chitin receptor homologs

In our experiments, chitin is widely perceived across members of the Rutaceae, including both cultivated citrus types and wild relatives (Figs. 1A and S1), indicating that chitin perception is likely derived from a conserved receptor. Therefore, we sought to further investigate the presence of *LYK5*, the major chitin receptor (Cao et al. 2014; Erwig et al. 2017; Xue et al. 2019), and *CERK1* across the plant kingdom. To identify and assess the conservation of *A. thaliana* LYK5$_{At}$ and CERK1$_{At}$ across a variety of eudicots and monocots, we used an approach based on homology, phylogeny,
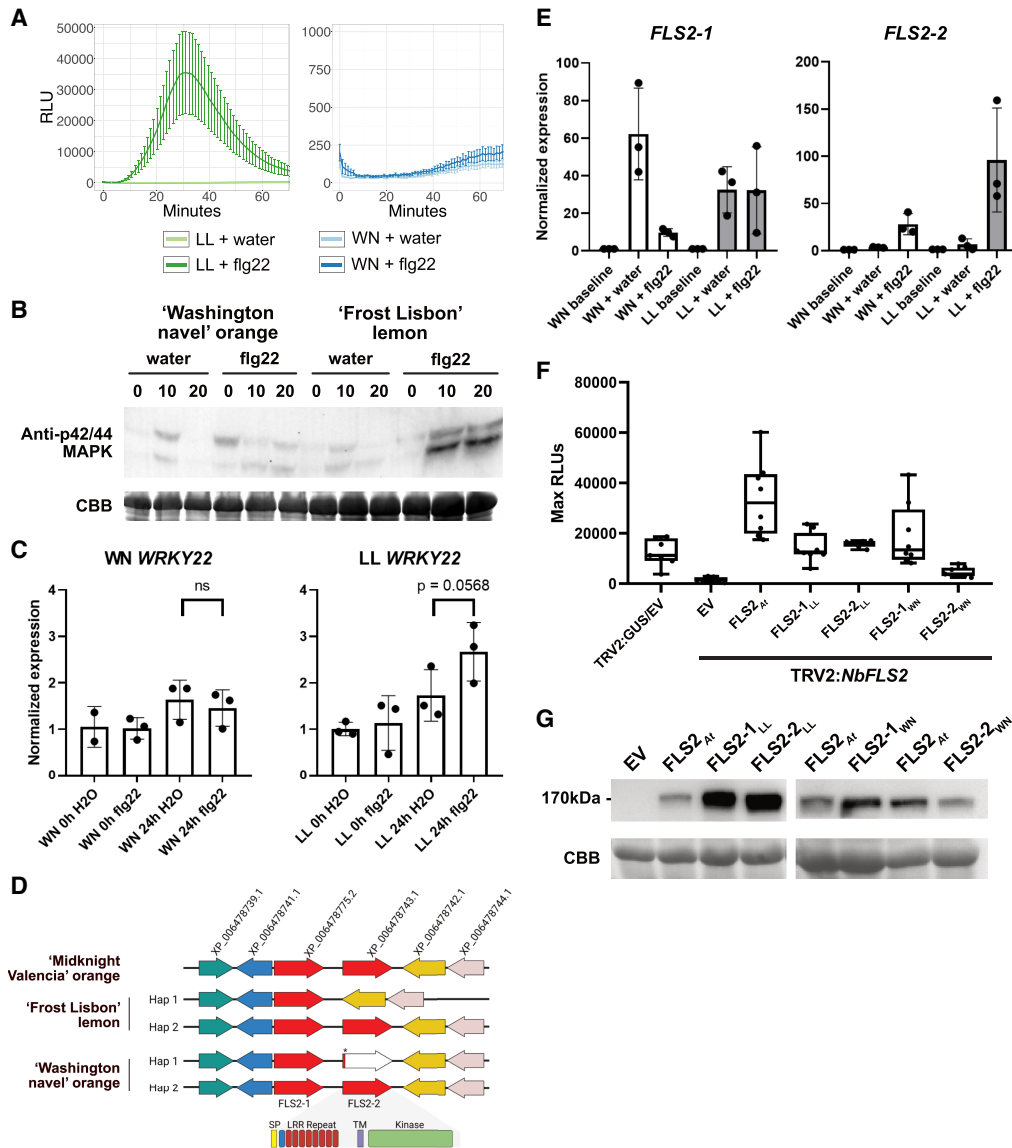
Figure 4. 'Washington navel' orange (WN) and 'Frost Lisbon' lemon (LL) contain FLS2 but differ in their response to bacterial flagellin. **A)** ROS curve for LL (left) and WN (right) when induced with either water or 100 nm flg22. *n* = 8 leaf disks; error bars denote standard error. Note the different scale on the *y*-axes. **B)** MAPK induction in response to either water or flg22 in WN versus LL using anti-p42/44 MAPK immunoblotting. Experiments were repeated 3 times. CBB = Coomassie Brilliant Blue. **C)** Normalized expression of *WRKY22* after induction with either water or 10 μm flg22. Significance was determined via 1-way ANOVA with a Šidák's multiple comparisons test to determine significance between water and flg22 treatments at 24 h. ns = nonsignificant. Bullet points represent 3 technical replicates from a single tree; error bars represent sd. Experiments were repeated 4× with similar results. **D)** Genome organization of *FLS2* homologs in citrus, with the 2 chromosomally linked homologs indicated. Haplotype data are shown for 'Frost Lisbon' lemon and 'Washington navel' orange. **E)** Expression of *FLS2-1* and *FLS2-2* transcripts measured via RT-qPCR at resting state and when induced with either flg22 or water, using citrus *GAPDH* as a reference gene. Error bars represent the sd (*n* = 3 biological replicates). Note the different axes scales between *FLS2-1* and *FLS2-2*. **F)** Transcomplementation experiments for FLS2 function in *N. benthamiana*. Two- to three-week post-silencing of *N. benthamiana* with tobacco rattle virus (TRV) targeting *GUS* (negative control) or endogenous *FLS2* (TRV2:NbFLS2), *FLS2* homologs from different plants were expressed using *Agrobacterium*-mediated transient expression. Forty-eight-hour post-*Agrobacterium* infiltration, leaf disks were subjected to ROS burst assays after treatment with 100 nm flg22, error bars = sd, *n* > 7. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range. **G)** Western blot demonstrating expression of all FLS2 proteins 48-h post-*Agrobacterium* infiltration. Top: anti-HA-HRP blot; bottom: Coomassie Brilliant Blue (CBB) staining. EV, empty vector; *At, Arabidopsis thaliana*.
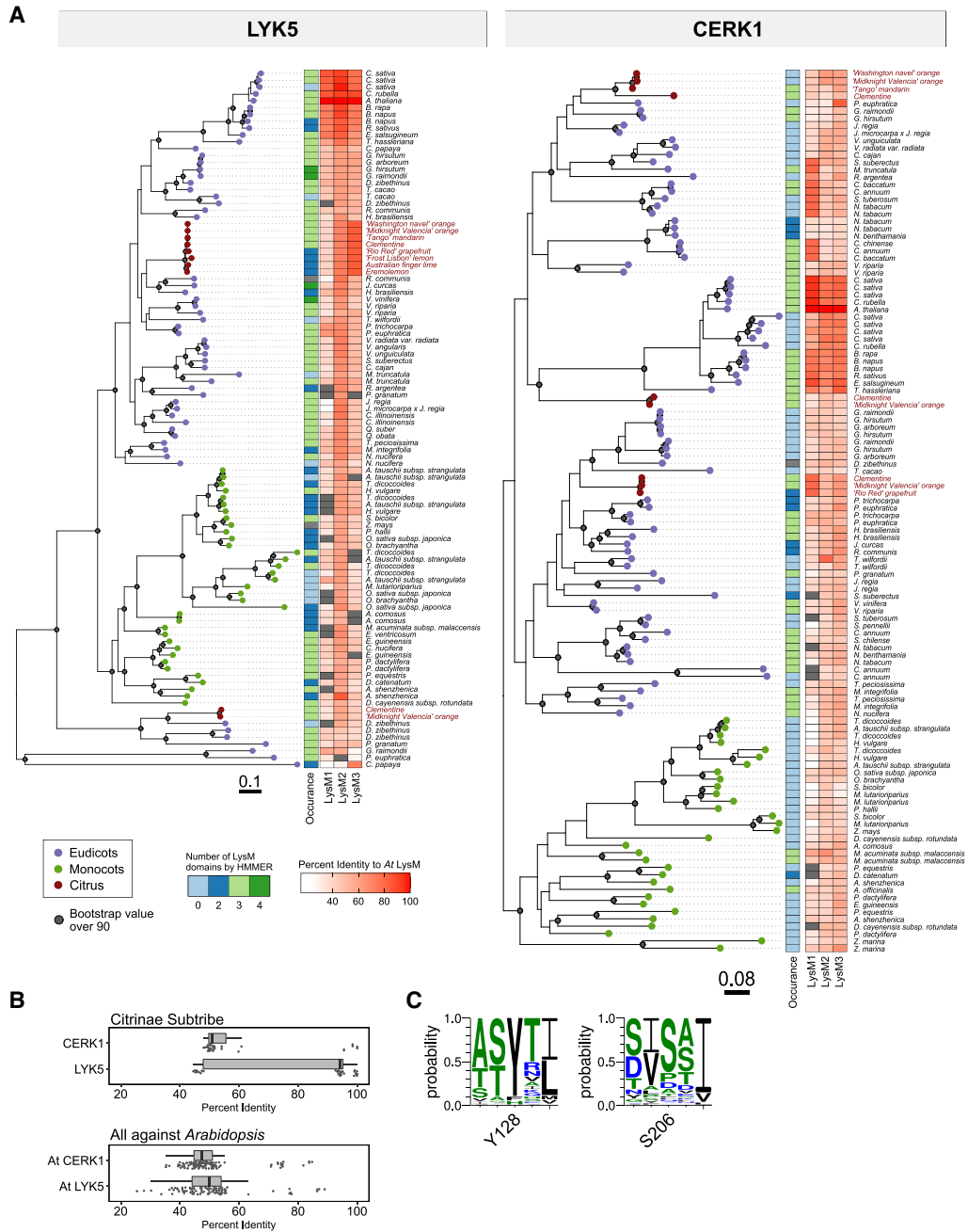
Figure 5. Phylogeny of LYK5 and CERK1 receptor homologs. **A)** Maximum likelihood phylogenetic tree of 102 LYK5 and LYK5-like homologs (top) and 120 CERK1 and CERK1-like homologs (bottom) from 66 plant species. In both trees, eudicots are labeled in purple, monocots are labeled in green, and sequences from citrus varieties are labeled in red. A total of 1,000 ultrafast bootstrap replicates were calculated, and values over 90 were plotted as a gray dot. To determine the number of LysM domains, hmmer (LysM domain, query ID: PF01476.19) and BLASTP were used. Similarity to the *A. thaliana* LysM domains by BLASTP from LYK5 and CERK1 was calculated and plotted. Scale bar indicates tree distance. **B)** All-by-all BLASTP of LysM receptor ectodomains. Top: BLASTP comparison of LYK5 and CERK1 homologs from the Citrinae tribe. Bottom: BLASTP comparison of all LYK5 and CERK1 plant homologs using *Arabidopsis* as a query. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range. **C)** Weblogos across 102 plant LYK5 homologs corresponding to critical residues for chitin binding in *A. thaliana* LYK5$_{At}$ Y128 and S206.

and hidden Markov models. Mid-rooted maximum likelihood trees show broad conservation of LYK5 and CERK1, with 41% of genotypes possessing multiple LYK5$_{At}$ homologs and 51% possessing multiple CERK1$_{At}$ homologs (Fig. 5A). There are 2 predominant LYK5 clades, a monophyletic monocot clade and a polyphyletic dicot clade including members from the Brassicaceae, Fabaceae, Malvaceae, and Rutaceae families. CERK1 displayed 2 major clades split by homologs from eudicots and monocots. Despite the diversification that can be found within the ectodomain of LYK5 and CERK1 homologs when compared against *Arabidopsis*, residues in LYK5$_{At}$ which are known to directly bind to chitin are conserved (Fig. 5, B and C) (Cao et al. 2014).

Additional citrus LYK5 members for Australian finger lime (*Citrus australasica*), 'Rio Red' grapefruit (*C.* ×*aurantium*), and 'Frost Lisbon' Lemon and Eremolemon (*Citrus glauca* × *C. limon*) were PCR amplified, sequenced, and plotted on the phylogeny.

LYK5 homologs within the Citrinae tribe exhibit low copy number and diversification, predominantly clustering in a single subclade (Fig. 5A). We attempted to PCR amplify additional *CERK1* homologs in Citrinae but were only able to amplify from 'Rio Red' grapefruit (*C.* ×*aurantium*). Within eudicots, citrus CERK1 homologs are polyphyletic, and both clementine and 'Midknight Valencia' orange carry multiple homologs (Fig. 5A). Ectodomains of citrus receptor homologs were compared, revealing a bimodal distribution for LYK5 and CERK1 and high amino acid similarity within citrus subclades (Fig. 5B).

Although most cultivated citrus genotypes can respond to chitin, there is still a wide range for the magnitude of the ROS response (Figs. 1A and S1). 'Washington navel' orange (*C.* × *aurantium*) is a sweet orange, and many modern type III mandarins are often derived from hybrids of sweet oranges and other mandarin types (Wu et al. 2018). 'Tango' mandarin has a stronger response to chitin, with a 5-fold stronger ROS burst (Fig. 6A). Chitin also activates MAPKs in both 'Washington navel' orange and 'Tango' mandarin, though the magnitude of the response varies in 'Tango' mandarin (Fig. 6B). Similarly, chitin treatment more robustly induced expression of the defense marker gene *WRKY22* in 'Tango' mandarin compared to 'Washington navel' orange (Fig. 6C). We identified *LYK5* alleles on each chromosome of 'Washington navel' orange and 'Tango' mandarin (type III; Supplemental Fig. S5). Both genotypes contain homologous *LYK5* genes in syntenic genetic regions that are highly similar to each other with conserved chitin binding residues (99.6% amino acid similarity; Figs. 6D and S5). Moreover, 'Washington navel' orange and 'Tango' mandarin possess 1 identical allele of *LYK5* (allele 1) and a second nearly identical allele with 2 amino acid polymorphisms (allele 2; Supplemental Fig. S5). One polymorphic residue is between transmembrane and kinase domain, while the second is in the kinase domain but not in a known critical residue (Supplemental Fig. S5). These *LYK5*

homologs are also expressed similarly in both genotypes using RT-qPCR (Fig. 6E).

To validate the functionality of *LYK5* allele 1, we complemented *Arabidopsis lyk4/lyk5-2* with the 'Tango' mandarin LYK5 (identical to 'Washington navel' orange allele 1, referred to as LYK5$_{TM}$) or *Arabidopsis* LYK5$_{At}$. Two independent *Arabidopsis* transgenic lines expressing LYK5$_{TM}$ as well as LYK5$_{At}$ regain the ability to produce ROS in response to chitin treatment (Fig. 7A). Immunoblot analyses against the HA epitope tag verified protein expression of all transgenes (Fig. 7B). We were able to further confirm the functionality of this receptor with MAPK assays. Both LYK5$_{TM}$ and LYK5$_{At}$ complementation lines exhibited MAPK phosphorylation upon chitin treatment (Fig. 7C). Taken together, we have demonstrated that cultivated citrus can respond to chitin and possess *LYK5* and *CERK1* homologs. The LYK5 allele 1 can also function as a chitin receptor in *Arabidopsis*.

Members from the Citrinae, Balsamocitrinae, and Clauseninae subtribes are capable of perceiving csp22 from an important citrus pathogen

In tomato, the CORE RLK perceives csp22, generating resistance to the bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 when expressed in *Arabidopsis* (Wang et al. 2016). The closest *N. benthamiana* homolog of tomato CORE is able to induce ROS production in response to csp22 treatment after transient expression (Wang et al. 2016). In order to gain insight into candidate citrus csp22 receptors, we analyzed citrus genomes for the presence of RLKs with similarity to the CORE receptor. However, the closest citrus RLK to either *Nicotiana* or *Solanum* CORE receptors had limited sequence similarity (Supplemental Fig. S8). Expression of the receptor recognizing csp22 is developmentally regulated and expressed in flowering *N. benthamiana* and tomato (Saur et al. 2016; Wang et al. 2016), which makes it possible to use *Agrobacterium*-mediated transient expression to investigate CORE in young *N. benthamiana*. We investigated the RLKs identified from 'Frost nucellar Eureka' lemon and 'Washington navel' orange, which are responsive to csp22. However, when these RLKs were heterologously expressed in 30-d-old *N. benthamiana*, they failed to confer csp22 responsiveness, in contrast to expression of *NbCORE*. All proteins were expressed by immunoblot analysis (Supplemental Fig. S8C). These data suggest that Rutaceae possesses an independently derived csp22 receptor.

Proteinaceous MAMPs are often conserved across pathogens. However, due to strong selection pressure, some pathogens have evolved immunogenic epitopes that cannot be perceived, while still retaining the presence of the entire protein (Cheng et al. 2021). The csp22 epitope from CLas (csp22$_{CLas}$) contains several polymorphisms when compared to the canonical csp22 sequence (Fig. 8A). Therefore, we investigated if there were members of the Rutaceae family that could perceive csp22$_{CLas}$ (Figs. 8, B to D, and S7). The vast majority of Rutaceae genotypes that can respond to canonical

csp22 cannot perceive csp22$_{CLas}$ using ROS production as an output. Notably, members from the Balsamocitrinae and Clauseninae subtribes can perceive both canonical csp22 and csp22$_{CLas}$, such as Uganda powder flask and *Clausena harmandiana*. There is 1 member of the Citrinae tribe, the 'Algerian clementine' mandarin (*C. ×aurantium*), that perceives csp22$_{CLas}$, but not canonical csp22. *C. harmandiana* is able to induce MAPK phosphorylation in response to csp22$_{CLas}$ compared to the nonresponding genotype 'Midknight Valencia' orange (Fig. 8D). Genotypes that respond to csp22$_{CLas}$ also exhibited some level of reduced symptomology to HLB disease in field trials with mature trees (Figs. 8 and S7; Ramadugu et al. 2016). However, not all genotypes with HLB tolerance can respond to csp22$_{CLas}$ (Supplemental Figs. S7 and S8). Data generated from our ROS screens in members of Rutaceae reveal members that can be used to identify receptors for transfer to CLas-susceptible citrus cultivars.

Discussion

Here, we have investigated variation in MAMP perception within the Rutaceae family to determine the landscape of perception in citrus and citrus relatives. Variations in MAMP perception have been noted within genotypes of the same species, such as in tomato (Roberts et al. 2019) and in Arabidopsis (Vetter et al 2012). Even in close relatives, the perception of a potent immune elicitor such as flg22 varies widely (Veluchamy et al. 2014). Much more diversity remains to be discovered by analyzing multiple genotypes. Perennial plants like citrus are largely unexplored due to long lifespans lengthening the time required to perform experiments, large field or greenhouse space required to grow tree crops, reduced access to diverse genotypes, and a lack of genomic resources.

There are multiple potential reasons why studies have observed variation in MAMP perception among related species. While some species may contain the same receptor homolog, the presence of the homolog does not always correspond with strong MAMP perception (Vetter et al. 2012; Trdá et al. 2014). In this study, the LYK5$_{At}$ homologs in 'Washington navel' orange and 'Tango' mandarin are identical for allele 1 and only differ by 2 amino acid changes in allele 2. These polymorphisms have not been previously described as important for LYK5 receptor function (Cao et al. 2014). However, 'Tango' mandarin is a much stronger ROS responder to chitin than 'Washington navel' orange. Similarly, 'Washington navel' orange does not respond to flg22, while 'Frost Lisbon' lemon does. Similar to our LYK5 results, both responsive and nonresponsive genotypes contain a functional *FLS2* homolog when expressed in *Arabidopsis* or *N. benthamiana*. In another study, Vetter et al. 2012 noted that the variation in FLS2 protein abundance for certain genotypes can reflect the variation in flg22 binding. While we were unable to determine protein abundance for LYK5 and FLS2 in citrus, both receptors were transcriptionally expressed at a similar basal level in responding and nonresponding genotypes. Minor variations in the rice OsCERK1 coreceptor have been linked to variation in mycorrhizal symbiosis; thus, it is possible minor allelic variation could also explain responsiveness to flagellin or chitin in citrus (Huang et al. 2020).

The segregation of immune response outputs has been observed previously, where flg22 from CLas induces defense gene induction but no ROS burst in 'Sun Chu Sha Kat' mandarin (Shi et al. 2018). Overexpression of the *N. benthamiana* FLS2 receptor in 'Hamlin' orange (*C. ×aurantium*) is able to confer flg22 responsiveness, indicating that boosting PRR expression may be a viable strategy to gain MAMP recognition (Hao et al. 2016). There is also a possibility that downstream signaling components may play a role in the presence and magnitude of immune responses. Signaling has not been investigated in detail in perennial crops, and further research may reveal if downstream signaling components have a role in altering MAMP responsiveness in citrus.

Flagellin perception is widespread amongst plants, mainly conferred by the receptor FLS2 (Saijo et al. 2018). Additional receptors were identified based on homology to *Arabidopsis* FLS2 in tomato, grapevine, citrus, and rice (Robatzek et al. 2007; Takai et al. 2008; Trdá et al. 2014; Shi et al. 2016). From our study, LYK5 and CERK1 homologs are also widespread and cluster separately between dicots and monocots. Rice utilizes a different LysM receptor (CEBiP) along with the CERK1 coreceptor for chitin perception (Kaku et al. 2006; Shimizu et al. 2010; Lee et al. 2014). Cotton (*Gossypium hirsutum*), a dicot, has a wall-associated kinase that interacts with LYK5 and CERK1 to promote chitin-induced dimerization (Wang et al. 2020). These results are consistent with an ancient acquisition of chitin perception in dicots, which may explain why a vast majority of the Rutaceae genotypes we evaluated are capable of producing an immune response to chitin. For MAMPs that can be recognized by a broad range of species, identifying receptors based on homology is a useful tactic.

Recent studies have identified immune receptor homologs that are capable of perceiving polymorphic flg22 epitopes that are not perceived by the canonical *Arabidopsis* FLS2 receptor. Flg22 from *Agrobacterium tumefaciens* contains several polymorphisms that prevent perception in *Arabidopsis* (Felix et al. 1999). Fürst et al. (2020) identified a flagellin-sensing receptor from wild grape with expanded ligand perception, FLS2^{XL}, capable of sensing both canonical flg22 and the *Agrobacterium* flg22 epitopes. The *Ralstonia solanacearum* flg22 is also highly polymorphic and is not recognized by tomato (Pfund et al. 2004). Wei et al. (2020) identified a FLS2/BAK1 complex in soybean that is capable of sensing the *Ralstonia* flg22. In this study, we have identified Rutaceae genotypes that are capable of recognizing canonical csp22 and csp22$_{CLas}$. In tomato, the CORE RLK is responsible for csp22 recognition (Wang et al. 2016). However, no obvious homolog of the *CORE* receptor has been identified in citrus genomes so far. While homology can be a fruitful approach
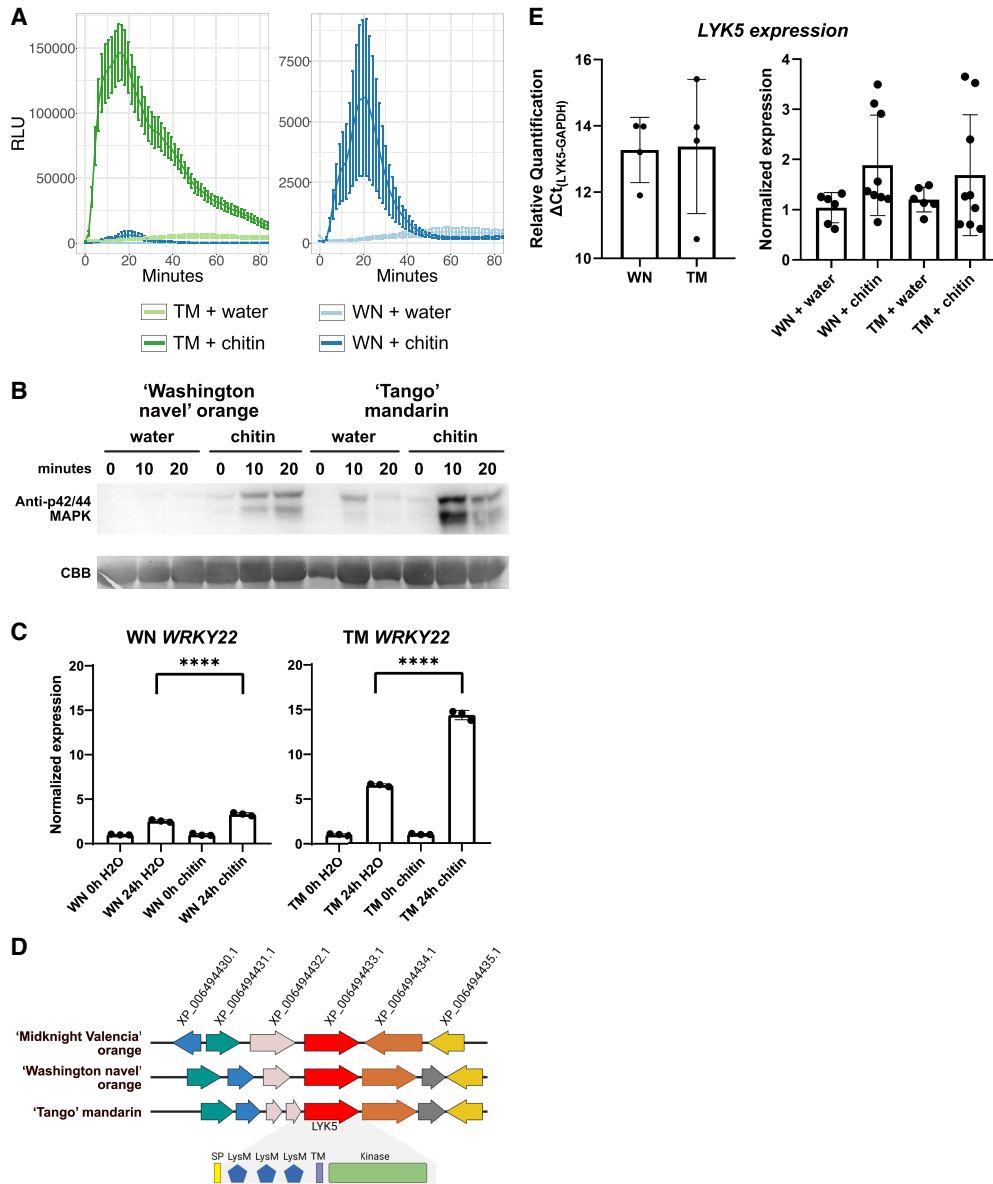
Figure 6. 'Washington navel' orange (WN) and 'Tango' mandarin (TM) contain LYK5 homologs but differ in magnitude of their chitin response. **A)** ROS curve for 'Washington navel' orange and 'Tango' mandarin (left) and 'Washington navel' orange only (right) when induced with either water or 10 μM chitin. $n = 8$ leaf disks; error bars denote standard error. Note the different scale on the *y*-axes. **B)** MAPK induction in response to either water for chitin in 'Washington navel' orange versus 'Tango' mandarin using anti-p42/44 MAPK immunoblotting. Experiments were repeated 4 times. CBB = Coomassie Brilliant Blue. **C)** Normalized expression of *WRKY22* after induction with either water or 10 μM chitin. Significance was determined via 1-way ANOVA with a Šidák's multiple comparisons test to determine significance between water and chitin treatments at the 24-h mark. Asterisks represent significance thresholds, with **** meaning $P \leq 0.0001$. Bullet points represent technical replicates from a single tree. Experiments were repeated 4×; significant 'Washington navel' orange induction with chitin was only observed in 2 out of 4 trials. **D)** Genome organization of *LYK5* in 'Midknight Valencia' orange, 'Washington navel' orange, and 'Tango' mandarin, with the LYK5 domain in 'Tango' mandarin expanded to show functional domains. Arrows indicate the difference in amino acid sequence between 'Tango' mandarin and 'Washington navel' orange. **E)** Transcript expression of citrus *LYK5* measured via RT-qPCR at resting state, using citrus *GAPDH* as a reference gene. Error bars represent the SD ($n = 4$ biological replicates). **E)** Transcript expression of citrus *LYK5* transcript via qPCR at resting state (left, ΔCt) and when induced with water or chitin (right), using citrus *GAPDH* as a reference gene. Error bars represent the SD ($n = 4$ biological replicates).
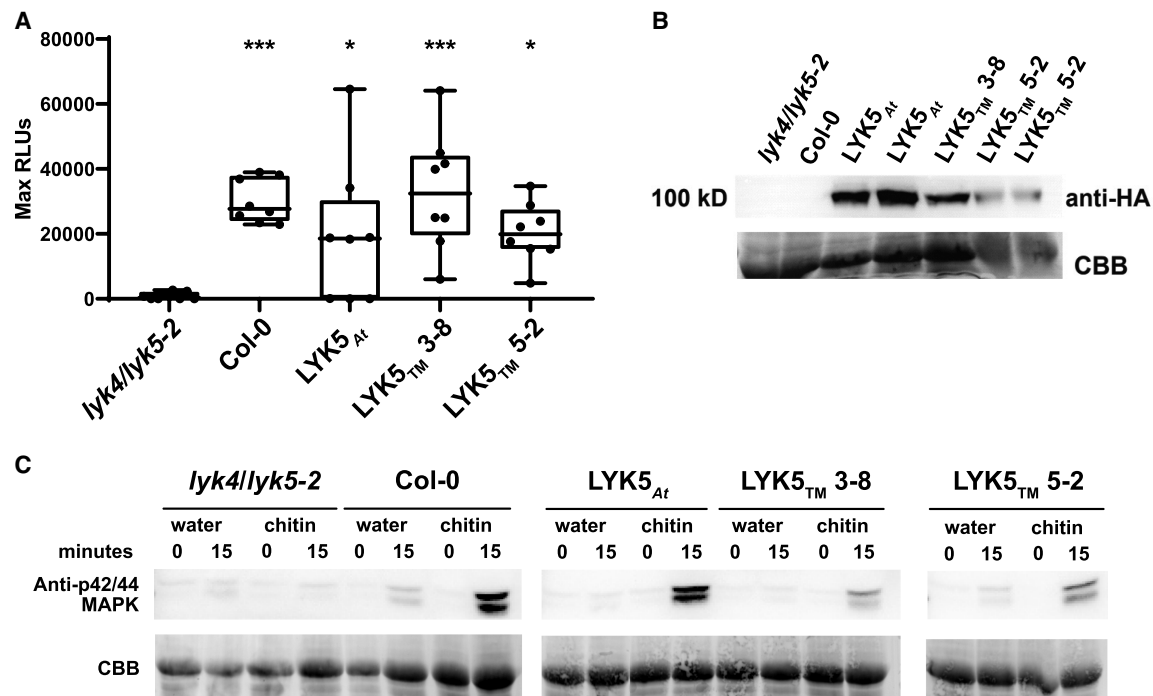
Figure 7. The 'Tango' mandarin LYK5$_{At}$ homolog can complement an *Arabidopsis* chitin perception mutant. **A)** ROS production of *Arabidopsis lyk4/lyk5-2* knockouts complemented with the indicated LYK5 constructs after treatment with 10 μM chitin. We complemented *Arabidopsis* with *LYK5* allele 1, which is referred to as *LYK5*$_{TM}$ and is identical between 'Tango' mandarin and 'Washington navel' orange. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range. Significance of results was determined via ordinary 1-way ANOVA, with a post hoc Dunnett's multiple comparison to the *lyk4/lyk5-2* knockout. Asterisks represent significance thresholds: ***P = 0.0001 to 0.001 and *P = 0.01 to 0.05. **B)** Anti-HA-HRP immunoblots visualize the expression of LYK5-HA homologs in *Arabidopsis*; CBB = Coomassie Brilliant Blue (CBB). **C)** MAPK induction in response to either water for chitin in LYK5-complemented *Arabidopsis lyk4/lyk5-2*, using anti-p42/44 MAPK immunoblotting. CBB = Coomassie Brilliant Blue. All experiments were performed 3 times with similar results.

to identify candidate receptors, convergent receptor evolution to recognize the same MAMP is also possible. There is likely an independently evolved receptor that members of the Rutaceae possess to recognize csp22 epitopes. Comparative genomics of csp22-responsive and csp22-nonresponsive citrus genotypes, or segregating populations if available, could be used to identify candidate receptor(s) for csp22 epitopes for future functional validation.

HLB induces some hallmarks of defense in susceptible plants, including callose and elevated ROS production, indicating that it can be a pathogen-triggered immune disease (Ma et al. 2022). The HLB susceptible orange genotypes we analyzed were unable to robustly respond to most MAMPs, unlike the more HLB tolerant 'Frost Lisbon' lemon. These results are consistent with weak, continuous, and ineffective defense activation in HLB susceptible citrus resulting in detrimental immune responses. It is possible that introduction of multiple PRRs capable of robustly inducing defense against CLas, including the csp22$_{CLas}$ receptor, may result in active pathogen clearing. *C.* ×*aurantium* overexpressing the SA receptor and master immune regulator NPR1 exhibited

increased tolerance to HLB and decreased pathogen titers (Dutt et al. 2015; Robertson et al. 2018; Peng et al. 2021). Appropriate regulation of defense and careful introduction of candidate receptors/genes should be considered with respect to HLB mitigation.

Our study highlights the diversity of immune response in a genetically diverse plant family. We identified relatives of citrus that are capable of responding to different MAMPs, opening up opportunities to study relatives with potential novel mechanisms of immune signaling. The transfer of MAMP receptors to susceptible plants can generate resistance to pathogens (Hao et al. 2016; Fürst et al. 2020; Wei et al. 2020). ROS–based immune phenotyping can be a high-throughput method to accelerate selection of promising individuals in a breeding program. Individuals that can respond to unique MAMPs are likely to have unique immune signaling components that can be transferred to susceptible citrus varieties. A greater understanding of immune perception repertoires in economically important plant genotypes will also facilitate the design stacks of receptors or signaling components for transfer and disease control. Similar strategies
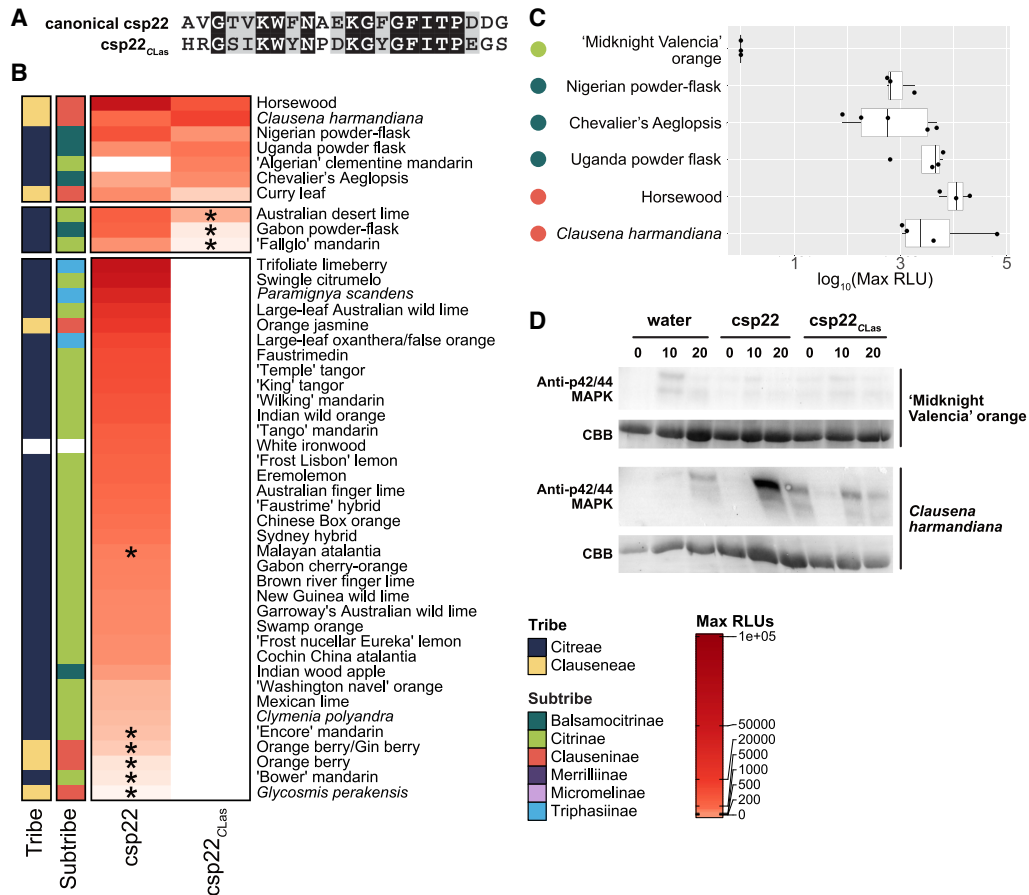
Figure 8. Three Rutaceae tribes can respond to a polymorphic csp22 from an important citrus pathogen. **A)** Alignment of the canonical csp22 sequence to the csp22 of CLas (csp22$_{CLas}$). **B)** Heat map compiling average max RLUs from ROS assays in genotypes within the Rutaceae family, organized by MAMP and phylogenetic relationship. Max RLUs are averages of at least 3 independent experiments, where $max\ RLU = (max\ RLU\ MAMP - max\ RLU\ water)$. The threshold for no response is <90 RLUs. Asterisks indicate genotypes that exhibit a variable response, where 1 or 2 independent experiments revealed a response. The MAMPs used for treatments are canonical csp22 and csp22 from CLas. **C)** Box plot of max RLUs for citrus relatives that can respond to 200 nM csp22$_{CLas}$. Max RLUs are averages of at least 3 independent experiments. Data points on box plots represent the average max RLU for an individual experiment, with $n = 8$ leaf disks per experiment. The max RLU are plotted on a log10 scale. The bar within the box plot depicts the median of the data, where the box boundaries represent the interquartile range (between the 25th and 75th percentiles) of the data. Box whiskers represent the minimum or maximum values of the data within 1.5× of the interquartile range. **D)** MAPK induction by csp22 or csp22$_{CLas}$ in either 'Midknight Valencia' orange or *C. harmandiana*, visualized by p42/44 MAPK antibody immunoblotting. CBB = Coomassie Brilliant Blue.

have resulted in durable resistance against fungal and oomycete pathogens in crop plants (Ghislain et al. 2019; Luo et al. 2021). This work has opened up interesting avenues to identify additional receptors in nonmodel species and highlight genotypes capable of recognizing polymorphic pathogen features.

Materials and methods

Plant materials and growth conditions
Eighty-six Rutaceae genotypes were tested for MAMP responsiveness under field and greenhouse conditions using mature trees (>5 yr old). Field-grown trees in the GCVC in

Riverside, CA (http://www.citrusvariety.ucr.edu), were analyzed between the spring and fall from 2018 to 2021. Supplemental Table S1 includes all genotypes analyzed, their accession IDs, and their location. When collecting Rutaceae samples, branches of selected Rutaceae genotypes were retrieved either from the greenhouse or the field, selecting branches with fully expanded leaves that still retained flexibility. Branches were stored by placing the cut side of the branch into a wet floral block (Oasis #10-00020-CASE) until processing.

A. thaliana seeds (Col-0 or *lyk4/lyk5-2* mutant) were stratified for 2 d in the dark at 4 °C before sowing onto soil or half-strength Murashige and Skoog (MS) media (Cao et al. 2014).

Plants were grown in a Conviron growth chamber at 23 °C and 70% relative humidity with a 10-h light/14-h dark photoperiod (100 μM m−2 s−1). Ten- to 14-day-old seedlings grown on MS were used for MAPK phosphorylation and MTI marker gene induction assays, and 4-wk-old soil grown plants were used for ROS and MAPK assays.

N. benthamiana was grown in a growth chamber at 26°C with a 16-h light/8-h dark photoperiod (180 μM m−2 s−1). Thirty-day-old (before flowering) plants were used for *Agrobacterium*-mediated transient protein expression, and 3- to 4-wk-old plants were used to examine *LYK5* transgene expression.

MAMPs
Immunogenic epitopes for flg22 and csp22 peptides were synthesized using Genscript (≥95% purity, Piscataway, NJ, USA). Hexaacetyl-chitohexaose (chitin) (Megazyme #O-CHI6) was diluted in water. The canonical flg22 epitope (QRLSTGSRINSAKDDAAGLQIA) is based on the sequence information from *Pseudomonas aeruginosa*. The canonical csp22 sequence (AVGTVKWFNAEKGFGFITPDGG) is from *P. syringae*. The csp22 CLas sequence (HRGSIKWYNPDKGYGFITPEGS) is identical to the sequence from the CLas strain psy62 (CLIBASIA_04060).

ROS burst assay
Leaf disks were collected using a #1 cork borer (4 mm) and floated overnight in 200-μL demineralized water in a Corning Costar 96-Well White Solid Plate (Fisher #07-200-589) with a plastic lid to prevent evaporation of the water. On the subsequent day, water was replaced with 100 μL of an assay solution containing MAMP. The assay solution contained 20 μM L-012 (a luminol derivative from Wako Chemicals USA #120-04891), 10 mg mL−1 horseradish peroxidase (Sigma), and MAMP. Concentrations used for MAMP treatments were as follows: 100 nM flg22, 10 μM chitin, 100 nM csp22, or 200 nM csp22$_{CLas}$. Luminescence was measured using a GloMax-Multi + Reader (Promega) or TriStar LB 941 plate reader (Berthold Technologies). At least 8 leaf discs from a single leaf were used in each replication, and the experiments were repeated at least 3 times. Samples that only responded once were considered variable.

For each plate, the average maximum RLUs for each tested MAMP–genotype combination were calculated from the maximum RLU of the 8 leaf disks after subtraction by the average water RLU. Across all ROS plates, average maximum RLUs were calculated based on all plates ran for each MAMP–genotype, and a heatmap was created from the accumulated ROS data via custom R scripts (Github repository: DanielleMStevens/Divergent_citrus_response_to_PAMPs) including the following R packages: ComplexHeatmap (v2.5.1; Gu et al. 2016) and circlize (v0.4.8; Gu et al. 2014).

MAPK induction assay
For MAPK induction assays in citrus, leaf disks from the same leaf were collected using a #6 cork borer (12 mm) from accessions grown in UC Davis or UC Riverside greenhouses and floated overnight in 1-mL deionized water in a 24-well tissue culture plate (VWR #10062-896) with a plastic lid to prevent evaporation of the water. On the subsequent day, water was replaced with 500 μL of either water or water containing MAMP before pressure-infiltrating for 2 min at 30 mm Hg in a vacuum desiccator (SP Bel-Art #F42025-0000). Leaf disks were individually collected at 0, 10, and 20 min after vacuum infiltration, flash frozen in liquid nitrogen, and ground up with pestles attached to an electric grinder (Conos AC-18S electric torque screwdriver) before adding 200-μL extraction buffer and grinding until homogenous. Protein extraction buffer contained 50 mM HEPES (pH 7.5), 50 mM NaCl, 10 mM EDTA, 0.2% v/v Triton X-100, Pierce Protease Inhibitor Mini Tablets, EDTA-free (Thermo #A32955), Pierce Phosphatase Inhibitor Mini Tablets (Thermo #A32957). Samples were centrifuged at 15,000 × g for 10 min to pellet cell debris.

Protein concentrations were quantified with the Pierce 660 nm Protein Assay Reagent (Thermo #22660) with Ionic Detergent Compatibility Reagent (Thermo #22663). MAPKs were visualized by anti-p44/42 MPK immunoblotting (1:2,000, Cell Signaling Technology #4370L) with goat antirabbit HRP secondary antibody (1:3,000, Bio-Rad #170-5046). Membranes were developed using the SuperSignal West Pico Chemiluminescent Substrate kit (Fisher #PI34578) and visualized on a ChemiDoc Touch Gel Imaging System (BioRad #1708370).

For MAPK induction assays in *Arabidopsis*, 5-d-old seedlings were transplanted into a 48-well tissue culture plate (Costar #3548) supplemented with half-strength MS liquid media. After 9 d, MS liquid media were replaced with 500 μL of either water or water containing 10 μM chitin. Three seedlings per treatment were collected at 0 and 15 min after chitin induction. Protein extraction and western blotting were conducted as described above. Experiments were performed 3 times.

RT-qPCR
To examine the expression of *LYK5* homologs, citrus leaves were harvested to make leaf punches with a #9 cork borer (22.5 mm). Each leaf disk was placed in a 12-well plate (VWR #10062-894) with 1-mL water and kept overnight at room temperature to allow the samples to recover from wounding. On the subsequent day, water was replaced with either 1-mL water or water containing 1-mL 10 μM chitin before vacuum infiltrating for 1.5 min at 30 mm Hg in a vacuum desiccator (SP Bel-Art #F42025-0000). Three leaf disks per treatment were collected at 0- and 24-h postinfiltration, and RNA extraction was performed as described above. To examine the expression of *FLS2* homologs, citrus leaves were syringe-infiltrated with water or 1 μM flg22 and plant

samples harvested 6 hpi. RNA extraction was performed as described above. Samples for resting state expression of *LYK5* and *FLS2* homologs were taken at 0 h as described above without infiltrating with water or MAMP.

To examine the expression of MTI marker genes, individual citrus leaves were infiltrated on trees with either water or 10 μM MAMP. At 0 and 24 h after infiltration, a #6 cork borer (12 mm) was used to make 6 leaf punches of infiltrated areas from the same leaf and was pooled as 1 sample. Leaf punches were manually ground with liquid nitrogen into a fine powder before transferring the powder to tubes to perform RNA extractions. RNA was extracted from plant samples with TRIzol (Fisher #15596018), following the manufacturer's instructions. DNase treatments for RNA preps were performed with RQ1 RNase-Free DNase (Promega #PR-M6101). cDNA synthesis was performed with the MMLV Reverse Transcriptase (Promega #PRM1705) kit.

A table of RT-qPCR primers used is listed in Supplemental Table S2. *Citrus* glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*; Pang et al. 2020) was used as the reference gene for RT-qPCR reactions. qPCR reactions were performed with SsoFast EvaGreen Supermix with Low ROX (BioRad #1725211) in a 96-well white PCR plate (BioRad #HSP9601) according to the manufacturer's instructions. Fold induction of gene expression was determined using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001), normalizing to water-treated and 0-h timepoints. Significance was determined via 1-way ANOVA with a Šidák's multiple comparisons test to determine significance between water and MAMP treatments at the 24-h mark. Marker gene expression experiments were performed 4 times.

Phylogenetic analyses and receptor comparisons

Plant genotypes used to build LYK5 and CERK1 phylogenies can be found in Supplemental Tables S3 and S4, respectively. Using the *A. thaliana* (NCBI taxid: 3702) $CERK1_{At}$ (NCBI RefSeq: NP_566689.2) and $LYK5_{At}$ (NCBI RefSeq: NP_180916.1) as queries, a BLASTP search was performed to mine for homologs across plant species (BLAST Suite v2.11.0+, query coverage cutoff 80%, *E*-value cutoff $1e^{-50}$, defaults otherwise). Partial protein hits were removed. Hits were then compared to all LysM RLKs in *A. thaliana* (CERK1, LYK2, LYK3, LYK4, and LYK5), and those closer to another LYK were removed. $LYK5_{At}$ hits from members of the Solanaceae family were removed because they were closer to $LYK4_{At}$ than $LYK5_{At}$. A multiple sequence alignment was built using MAFFT (v7.310, −reorder, −maxiterate 1000, −localpair, defaults otherwise). TrimAl was used to trim each multiple sequence alignment for large gaps (v1.4.rev15, automated1, defaults otherwise). A maximum likelihood tree was built from the alignment using iqtree (v2.1.2, —bb 1000, -T AUTO, -st AA, -v -m MFP -safe, defaults otherwise), mid-rooted, and visualized using R packages phangorn (v2.7.1) and ggtree (v3.1.2.991).

We sought to characterize the number of chitin-binding LysM domains in LYK5 and CERK1 homologs. The standard domain prediction software Interproscan was unable to accurately predict even well-characterized $LYK5_{At}$ and $CERK1_{At}$ receptors. To improve assessment of LysM domain frequency, we used homology and a hidden Markov model approach by BLASTP and HMMER, respectively. Hmmersearch using the LysM domain (query ID: PF01476.19) as a query identified the number of LysM domains (hmmer v3.1b2, -E 1e-5, defaults otherwise) (Eddy 2011). Manually extracted LysM domains from *A. thaliana* Col-0 were used to build a local BLAST database to calculate similarity of each LysM domain (coverage cutoff 80%, *E*-value cutoff $1e^{-50}$, defaults otherwise). All LysM domain analyses were plotted onto the receptor trees in R.

Ectodomains of RLKs were extracted and an all-by-all comparison was computed using BLASTP. Similarity to either $CERK1_{At}$ and $LYK5_{At}$ or all citrus homologs to each other was plotted using R packages ggplot2 (v3.3.5) and ggbeeswarm (v0.6.0). Weblogos were generated from the multiple sequence alignment corresponding to $LYK5_{At}$ residues Y128 and S206 using WebLogo3 (Crooks et al. 2004).

The phylogenetic tree of citrus RLKs using CORE as a query was built with CORE homologs from other solanaceous plants (*S. lycopersicum*: Solyc03g096190; *Solanum pennellii*: XP_015068909.1; *N. benthamiana*: Niben101 Scf02323g01010.1; *Nicotiana sylvestris*: XP_009803840.1; and *Nicotiana tabacum*: XP_016470062.1). Protein sequences were aligned via MAFFT (tree building was performed by iqtree) and were visualized in R similarly as described above. Sequences of cloned CRLKs from 'Frost nucellar Eureka' lemon (*C. ×limon*) and 'Washington navel' orange (*C. ×aurantium*) are deposited in GenBank (ON863917 and ON863918, respectively).

For more information and raw files, see Github repository: DanielleMStevens/Divergent_citrus_response_to_PAMPs.

Cloning citrus LYK5 homologous sequences and transcomplementation in *Arabidopsis*

Putative LYK5 sequences from Australian finger lime (*C. australasica*), 'Rio Red' grapefruit (*C. ×aurantium*), 'Frost Lisbon' lemon (*C. ×limon*), 'Tango' mandarin (*C. ×aurantium*), and Eremolemon (*C. glauca* × *C. ×limon*) were amplified using iProof DNA polymerase (Bio-Rad #BR0114). The PCR products were initially cloned into a pENTR/D-TOPO backbone (Invitrogen #K2400-20), and then *LYK5* from 'Tango' mandarin (TM) was selected and inserted into a modified pGWB14 binary destination vector with the *Arabidopsis* ubiquitin 10 promoter using Gateway LR Clonase II enzyme mix (Invitrogen #11791-100). The *Arabidopsis lyk4/lyk5-2* mutant (Cao et al. 2014) was transformed using a floral dip method with pUBQ10::TM_LYK5-HA and pUBQ10::LYK5-HA from Arabidopsis (Zhang et al. 2006). Experiments were performed 3 times with T4 homozygous lines.

VIGS

VIGS was performed as described in (Chakravarthy et al. 2010). Two-week-old *N. benthamiana* seedlings were

infiltrated with pTRV1(RNA1), along with silencing constructs: pTRV2:*GUS*, pTRV2:*PDS*, and pTRV2:*FLS2* (Chakravarthy et al. 2010). After 2 to 3 weeks, silenced plants were infiltrated with *Agrobacterium* harboring FLS2 constructs for transient expression and ROS assays. *Arabidopsis* Col-0 *FLS2*, 'Washington navel' orange *FLS2*, and 'Frost Lisbon' lemon *FLS2* homologs were amplified and inserted into a modified pGWB14 binary destination vector with the *Arabidopsis* ubiquitin 10 promoter as described above. Plasmids were transformed via electroporation into *Agrobacterium* C58C1, and VIGS-silenced *N. benthamiana* plants were infiltrated with *Agrobacterium* suspensions of OD$_{600}$ 0.6. Forty-eight hours after infiltration, leaf disks were collected using a #1 cork borer (4 mm) for ROS assays after challenging with 100 nM flg22 as described above. To visualize the expression of FLS2 homologs, additional leaf disks were collected using a #7 cork borer at 48 hpi for protein extraction. Leaf disks were homogenized in 100-μL Laemmli buffer and boiled for 5 min. Western blotting was conducted as described above and visualized with anti HA-HRP antibody (Roche 39 #12013819001; 1:2,000).

Transient expression of CORE in *Nicotiana*

RLKs with the highest similarity to tomato COREs from 'Frost nucellar Eureka' lemon (*C.* × *limon*) and 'Washington navel' orange were amplified using iProof DNA polymerase (Bio-Rad #BR0114). PCR products were cloned into a pEARLY103 backbone (Earley et al. 2006) with expression mediated by a 35*S* promoter, and plasmids were transformed via electroporation into *Agrobacterium* GV3101. *N. benthamiana* CORE was used as a positive control (Wang et al. 2016). To test the function of CORE homologs, young (nonflowering) *N. benthamiana* plants were infiltrated with *Agrobacterium* suspensions of OD$_{600}$ 0.25. Twenty-four hours after infiltration, leaf disks were collected using a #1 cork borer (4 mm) for ROS assays as described above. To visualize the expression of CORE homologs, additional leaf disks were collected using a #7 cork borer at 48 hpi for protein extraction. Leaf disks were homogenized in 100-μL Laemmli buffer and boiled for 5 min. Western blotting was conducted as described above and visualized with anti GFP-HRP antibody (Miltenyi Biotec #130-091-833).

Comparison of LYK5 and FLS2 homologs in citrus

Genome sequences surrounding *LYK5* and *FLS2* homologs in cultivated citrus were compared based on BLASTP hits with an *E*-value of 0.001 (Altschul et al. 1990) against de novo assembled genomes. Contigs were assembled using wtgbt2 (Ruan and Li 2020), and long-read sequencing was performed on the Pacific Biosciences Sequel II platform in the CLR sequencing mode. LYK5 and FLS2 protein sequences were aligned using MUSCLE v5.1 (Edgar 2021).

Accession numbers

Sequence data from this article can be found in the GenBank data libraries under these accession numbers: *LYK5* from "Washington navel" orange and "Tango" mandarin: ON685188, ON685189, ON685190, ON685191; *FLS2* from "Washington navel" orange and "Frost Lisbon" lemon: OP718785, OP718786, OP718788, OP718789; RLKs from "Frost nucellar Eureka" lemon and "Washington navel" orange in Supplemental Fig. S8B: ON863917 and ON863918.

Author contributions

J.Y.F., T.Y.T, C.R., and G.L.C. designed the research, and J.T., T.L., J.Y.F., T.Y.T., C.R., R.P., and J.W. performed the research. J.T., T.L., D.M.S., J.Y.F., T.Y.T., S.P.T., C.R., and G.C. analyzed the data, and J.T., T.L., D.M.S., and G.C. wrote the paper. D.K.S. and E.A.d.-D. provided citrus sequencing data. T.L.K. provided guidance and access to greenhouse-grown Rutaceae genotypes.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Table S1. Information about the Rutaceae genotypes used in this study.

Supplemental Table S2. Primers used in this study.

Supplemental Table S3. Information about the LYK5 sequences used to build Fig. 5A.

Supplemental Table S4. Information on the CERK1 sequences used to build Fig. 5A.

Supplemental Figure S1. Box plots containing the distribution of ROS data from Fig. 1 for all nonvariable Rutaceae genotypes.

Supplemental Figure S2. Representative ROS curves for the genotypes shown in Fig. 2.

Supplemental Figure S3. Amino acid alignments of FLS2 homologs.

Supplemental Figure S4. Baseline expression of *FLS2-1* and *FLS2-2*.

Supplemental Figure S5. Amino acid alignments of LYK5 homologs.

Supplemental Figure S6. Full heat map featuring all genotypes tested with either canonical csp22 or *C*Las csp22, related to Fig. 7.

Supplemental Figure S7. Visualization of MAMP perception with HLB disease susceptibility.

Supplemental Figure S8. The closest citrus RLKs to CORE from Solanaceous plants cannot perceive csp22.

Data availability

The data that support the findings of this study are available in GenBank, with raw data available from the corresponding author upon request. Software used to build the heatmap and parts of Fig. 5 are deposited in GitHub under the repository DanielleMStevens/Divergent_citrus_response_to_PAMPs.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:**215**(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Appelhans MS, Bayly MJ, Heslewood MM, Groppo M, Verboom GA, Forster PI, Kallunki JA, Duretto MF. A new subfamily classification of the Citrus family (Rutaceae) based on six nuclear and plastid markers. TAXON 2021:**70**(5):1035–1061. https://doi.org/10.1002/tax.12543

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W-L, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J. MAP kinase signalling cascade in *Arabidopsis* innate immunity. Nature 2002:**415**(6875): 977–983. https://doi.org/10.1038/415977a

Bigeard J, Colcombet J, Hirt H. Signaling mechanisms in pattern-triggered immunity (PTI). Molecular plant. 2015:**8**(4):521–539. https://doi.org/10.1016/j.molp.2014.12.022

Bové JM. Huanglongbing: a destructive, newly-emerging, century-old disease of *Citrus*. J Plant Pathol. 2006:**88**(1):7–37. http://dx.doi.org/10.4454/jpp.v88i1.828

Cao Y, Liang Y, Tanaka K, Nguyen CT, Jedrzejczak RP, Joachimiak A, Stacey G. The kinase LYK5 is a major chitin receptor in *Arabidopsis* and forms a chitin-induced complex with related kinase CERK1. eLife 2014:**3**:e03766. https://doi.org/10.7554/eLife.03766

Caruso M, Smith MW, Froelicher Y, Russo G, Gmitter FG. Chapter 7 —traditional breeding. In: Talon M, Caruso M, Gmitter FG, editors. The genus Citrus. Cambridge (UK): Woodhead Publishing. p. 129–148. https://doi.org/10.1016/B978-0-12-812163-4.00007-3

Castle WS. A career perspective on *Citrus* rootstocks, their development, and commercialization. HortScience 2010:**45**(1):11–15. https://doi.org/10.21273/HORTSCI.45.1.11

Chakravarthy S, Velásquez AC, Ekengren SK, Collmer A, Martin GB. Identification of *Nicotiana benthamiana* genes involved in pathogen-associated molecular pattern–triggered immunity. Mol Plant Microbe Interact. 2010:**23**(6):715–726. http://dx.doi.org/10.1094/MPMI-23-6-0715

Chen Y, Bendix C, Lewis JD. Comparative genomics screen identifies microbe-associated molecular patterns from *Candidatus* Liberibacter sp. that elicit immune responses in plants. Mol Plant

Microbe Interact. 2020:**33**(3):539–552. https://doi.org/10.1094/MPMI-11-19-0309-R

Cheng JHT, Bredow M, Monaghan J, diCenzo GC. Proteobacteria contain diverse flg22 epitopes that elicit varying immune responses in *Arabidopsis thaliana*. MPMI 2021:**34**(5):504–510. https://doi.org/10.1094/MPMI-11-20-0314-SC

Coletta-Filho HD, Castillo AI, Laranjeira FF, de Andrade EC, Silva NT, de Souza AA, Bossi ME, Almeida RPP, Lopes JRS. Citrus variegated chlorosis: an overview of 30 years of research and disease management. Trop. Plant pathol. 2020:**45**:175–191. https://doi.org/10.1007/s40858-020-00358-5

Couto D, Zipfel C. Regulation of pattern recognition receptor signalling in plants. Nat Rev Immunol. 2016:**16**(9):537–552. https://doi.org/10.1038/nri.2016.77

Crooks GE, Hon G, Chandonia J-M, Brenner SE. Weblogo: a sequence logo generator. Genome Res. 2004:**14**(6):1188–1190. https://doi.org/10.1101/gr.849004

Dalio RJD, Magalhães DM, Rodrigues CM, Arena GD, Oliveira TS, Souza-Neto RR, Picchi SC, Martins PMM, Santos PJC, Maximo HJ, et al. PAMPs, PRRs, effectors and R-genes associated with citrus–pathogen interactions. Ann Bot. 2017:**119**(5):749–774. https://doi.org/10.1093/aob/mcw238

Das AK. Citrus canker—A review. J Appl Hortic. 2003:**5**((01|1)):52–60. https://doi.org/10.37855/jah.2003.v05i01.15

Dutt M, Barthe G, Irey M, Grosser J. Transgenic *Citrus* expressing an *Arabidopsis* NPR1 gene exhibit enhanced resistance against Huanglongbing (HLB; Citrus greening). PLoS One 2015:**10**(9): e0137134. https://doi.org/10.1371/journal.pone.0137134

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. 2006:**45**(4):616–629. https://doi.org/10.1111/j.1365-313X.2005.02617.x

Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011:**7-**(10):e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edgar RC. MUSCLE V5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. bioRxiv. 2021. https://doi.org/10.1101/2021.06.20.449169, preprint: not peer reviewed.

Erwig J, Ghareeb H, Kopischke M, Hacke R, Matei A, Petutschnig E, Lipka V. Chitin-induced and CHITIN ELICITOR RECEPTOR KINASE1 (CERK1) phosphorylation-dependent endocytosis of *Arabidopsis thaliana* LYSIN MOTIF-CONTAINING RECEPTOR-LIKE KINASE5 (LYK5). New Phytol. 2017:**215**(1):382–396. https://doi.org/10.1111/nph.14592

Felix G, Duran JD, Volko S, Boller T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. 1999:**18**(3):265–276. https://doi.org/10.1046/j.1365-313X.1999.00265.x

Ference CM, Gochez AM, Behlau F, Wang N, Graham JH, Jones JB. Recent advances in the understanding of *Xanthomonas citri* ssp. *citri* pathogenesis and citrus canker disease management. Mol Plant Pathol. 2018:**19**(6):1302–1318. https://doi.org/10.1111/mpp.12638

Fürst U, Zeng Y, Albert M, Witte AK, Fliegmann J, Felix G. Perception of *Agrobacterium tumefaciens* flagellin by FLS2XL confers resistance to crown gall disease. Nat Plants. 2020:**6**(1):22–27. https://doi.org/10.1038/s41477-019-0578-6

Ghislain M, Byarugaba AA, Magembe E, Njoroge A, Rivera C, Román ML, Tovar JC, Gamboa S, Forbes GA, Kreuze JF, et al. Stacking three late blight resistance genes from wild species directly into African highland potato varieties confers complete field resistance to local blight races. Plant Biotechnol J. 2019:**17**(6):1119–1129. https://doi.org/10.1111/pbi.13042

Gómez-Gómez L, Boller T. FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. Mol Cell. 2000:**5**(6):1003–1011. https://doi.org/10.1016/S1097-2765(00)80265-8

Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016:**32-**(18):2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics 2014:**30**(19): 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Hao G, Pitino M, Duan Y, Stover E. Reduced susceptibility to *Xanthomonas citri* in transgenic *Citrus* expressing the FLS2 receptor from *Nicotiana benthamiana*. Mol Plant Microbe Interact. 2016:**29**(2):132–142. https://doi.org/10.1094/MPMI-09-15-0211-R

Hodges AW, Spreen TH. Economic impacts of *Citrus* greening (HLB) in Florida, 2006/07–2010/11: FE903/FE903, 1/2012. EDIS 2012:**2012**(1): 1–6. https://doi.org/10.32473/edis-fe903-2012

Huang R, Li Z, Mao C, Zhang H, Sun Z, Li H, Huang C, Feng Y, Shen X, Bucher M, et al. Natural variation at OsCERK1 regulates arbuscular mycorrhizal symbiosis in rice. New Phytol. 2020:**225**(4):1762–1776. https://doi.org/10.1111/nph.16158

Jaouad M, Moinina A, Ezrari S, Lahlali R. Key pests and diseases of citrus trees with emphasis on root rot diseases: an overview. Moroccan J Agric Sci. 2020:**1**(3):149–160. http://dx.doi.org/MJAS_Jaouad_1-3

Jeworutzki E, Roelfsema MRG, Anschütz U, Krol E, Elzenga JTM, Felix G, Boller T, Hedrich R, Becker D. Early signaling through the *Arabidopsis* pattern recognition receptors FLS2 and EFR involves Ca^{2+}- associated opening of plasma membrane anion channels. Plant J. 2010:**62**(3):367–378. https://doi.org/10.1111/j.1365-313X.2010.04155.x

Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiyama C, Dohmae N, Takio K, Minami E, Shibuya N. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. PNAS 2006:**103**(29):11086–11091. https://doi.org/10.1073/pnas.0508882103

Kubitzki K, Kallunki JA, Duretto M, Wilson PG. Rutaceae In: **Kubitzki K**, editor. Flowering plants. Eudicots: Sapindales, Cucurbitales, Myrtaceae, the families and genera of vascular plants. Berlin, Heidelberg: Springer; 2011. pp. 276–356. https://doi.org/10.1007/978-3-642-14397-7_16

Lee W-S, Rudd JJ, Hammond-Kosack KE, Kanyuka K. *Mycosphaerella graminicola* LysM effector-mediated stealth pathogenesis subverts recognition through both CERK1 and CEBiP homologues in wheat. MPMI 2014:**27**(3):236–243. https://doi.org/10.1094/MPMI-07-13-0201-R

Liu Y, Heying E, Tanumihardjo SA. History, global distribution, and nutritional importance of *Citrus* fruits. Compr Rev Food Sci Food Safety. 2012:**11**(6):530–545. https://doi.org/10.1111/j.1541-4337.2012.00201.x

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 2001:**25**(4):402–408. https://doi.org/10.1006/meth.2001.1262

Luo M, Xie L, Chakraborty S, Wang A, Matny O, Jugovich M, Kolmer JA, Richardson T, Bhatt D, Hoque M, et al. A five-transgene cassette confers broad-spectrum resistance to a fungal rust pathogen in wheat. Nat Biotechnol. 2021:**39**(5):561–566. https://doi.org/10.1038/s41587-020-00770-x

Ma W, Pang Z, Huang X, Xu J, Pandey SS, Li J, Achor DS, Vasconcelos FNC, Hendrich C, Huang Y, et al. Citrus Huanglongbing is a pathogen-triggered immune disease that can be mitigated with antioxidants and gibberellin. Nat Commun. 2022:**13**(1):529. https://doi.org/10.1038/s41467-022-28189-9

Magalhães DM, Scholte LLS, Silva NV, Oliveira GC, Zipfel C, Takita MA, De Souza AA. LRR-RLK family from two *Citrus* species: genome-wide identification and evolutionary aspects. BMC Genomics 2016:**17**(1):623. https://doi.org/10.1186/s12864-016-2930-9

Meng X, Zhang S. MAPK cascades in plant disease resistance signaling. Annu Rev Phytopathol. 2013:**51**(1):245–266. https://doi.org/10.1146/annurev-phyto-082712-102314

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. PNAS 2007:**104**(49):19613–19618. https://doi.org/10.1073/pnas.0705147104

Morton CM. Phylogenetic relationships of the Aurantioideae (Rutaceae) based on the nuclear ribosomal DNA ITS region and three noncoding chloroplast DNA regions, atpB-rbcL spacer, rps16, and trnL-trnF. Org Divers Evol. 2009:**9**(1):52–68. https://doi.org/10.1016/j.ode.2008.11.001

Nagano Y, Mimura T, Kotoda N, Matsumoto R, Nagano AJ, Honjo MN, Kudoh H, Yamamoto M. Phylogenetic relationships of Aurantioideae (Rutaceae) based on RAD-seq. Tree Genet Genomes. 2018:**14**(1):6. https://doi.org/10.1007/s11295-017-1223-z

Newman M-A, Sundelin T, Nielsen J, Erbs G. MAMP (microbe-associated molecular pattern) triggered immunity in plants. Front Plant Sci. 2013:**4**(139):1–14. https://doi.org/10.3389/fpls.2013.00139

Ngou BPM, Ding P, Jones JDG. Thirty years of resistance: zig-zag through the plant immune system. Plant Cell 2022:**34**(5): 1447–1478. https://doi.org/10.1093/plcell/koac041

Nguyen HP, Chakravarthy S, Velásquez AC, McLane HL, Zeng L, Nakayashiki H, Park D-H, Collmer A, Martin GB. Methods to study PAMP-triggered immunity using tomato and *Nicotiana benthamiana*. Mol Plant Microbe Interact. **23**(8):991–999. http://dx.doi.org/10.1094/MPMI-23-8-0991

Pang Z, Zhang L, Coaker G, Ma W, He S-Y, Wang N. Citrus CsACD2 is a target of *Candidatus* Liberibacter Asiaticus in Huanglongbing disease. Plant Physiol. 2020:**184**(2):792–805. https://doi.org/10.1104/pp.20.00348

Peng A, Zou X, He Y, Chen S, Liu X, Zhang J, Zhang Q, Xie Z, Long J, Zhao X. Overexpressing a NPR1-like gene from *Citrus paradisi* enhanced Huanglongbing resistance in *C. sinensis*. Plant Cell Rep. 2021:**40**(3):529–541. https://doi.org/10.1007/s00299-020-02648-3

Pfund C, Tans-Kersten J, Dunning FM, Alonso JM, Ecker JR, Allen C, Bent AF. Flagellin is not a major defense elicitor in *Ralstonia solanacearum* cells or extracts applied to *Arabidopsis thaliana*. MPMI. 2004:**17**(6):696–706. https://doi.org/10.1094/MPMI.2004.17.6.696

Plieth C. Redox modulators determine luminol luminescence generated by porphyrin-coordinated iron and may repress "suicide inactivation.". ACS Omega. 2018:**3**(9):12295–12303. https://doi.org/10.1021/acsomega.8b01261

Ramadugu C, Keremane ML, Halbert SE, Duan YP, Roose ML, Stover E, Lee RF. Long-Term field evaluation reveals Huanglongbing resistance in *Citrus* relatives. Plant Dis. 2016:**100**(9):1858–1869. https://doi.org/10.1094/PDIS-03-16-0271-RE

Robatzek S, Bittel P, Chinchilla D, Köchner P, Felix G, Shiu S-H, Boller T. Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of *Arabidopsis* FLS2 exhibiting characteristically different perception specificities. Plant Mol Biol. 2007:**64**(5):539–547. https://doi.org/10.1007/s11103-007-9173-8

Roberts R, Mainiero S, Powell AF, Liu AE, Shi K, Hind SR, Strickler SR, Collmer A, Martin GB. Natural variation for unusual host responses and flagellin-mediated immunity against *Pseudomonas syringae* in genetically diverse tomato accessions. New Phytol. 2019:**223**(1):447–461. https://doi.org/10.1111/nph.15788

Robertson CJ, Zhang X, Gowda S, Orbović V, Dawson WO, Mou Z. Overexpression of the *Arabidopsis* NPR1 protein in citrus confers tolerance to Huanglongbing. J Citrus Pathol. 2018:**5**(1):1–8. https://doi.org/10.5070/C451038911

Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020:**17**(2):155–158. https://doi.org/10.1038/s41592-019-0669-3

Saijo Y, Loo EP, Yasuda S. Pattern recognition receptors and signaling in plant–microbe interactions. Plant J. 2018:**93**(4):592–613. https://doi.org/10.1111/tpj.13808

Saur IML, Kadota Y, Sklenar J, Holton NJ, Smakowska E, Belkhadir Y, Zipfel C, Rathjen JP. NbCSPR underlies age-dependent immune responses to bacterial cold shock protein in *Nicotiana benthamiana*. Proc Natl Acad Sci USA. 2016:**113**(12):3389–3394. https://doi.org/10.1073/pnas.1511847113

Shi Q, Febres VJ, Jones JB, Moore GA. A survey of FLS2 genes from multiple citrus species identifies candidates for enhancing disease

resistance to *Xanthomonas citri* ssp. *citri*. Hortic Res. 2016:**3**(1):16022. https://doi.org/10.1038/hortres.2016.22

Shi Q, Febres VJ, Zhang S, Yu F, McCollum G, Hall DG, Moore GA, Stover E. Identification of gene candidates associated with Huanglongbing tolerance, using '*Candidatus* Liberibacter asiaticus' flagellin 22 as a proxy to challenge *Citrus*. MPMI 2018:**31**(2): 200–211. https://doi.org/10.1094/MPMI-04-17-0084-R

Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, Minami E, Okada K, Yamane H, Kaku H, et al. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J. 2010:**64**(2):204–214. https://doi.org/10.1111/j.1365-313X.2010.04324.x

Swingle WT, Reece PC. The botany of *Citrus* and its wild relatives. In: **Reuther W, Webber HJ, Batchelor LD**, editors. The Citrus Industry. Vol I. USA: Published by University of California Riverside; USA. 1967. pp. 190–430. http://citruspages.free.fr/CI/Vol1/Chapter3.html

Takai R, Isogai A, Takayama S, Che F-S. Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. MPMI 2008:**21**(12): 1635–1642. https://doi.org/10.1094/MPMI-21-12-1635

Trdá L, Fernandez O, Boutrot F, Héloir M-C, Kelloniemi J, Daire X, Adrian M, Clément C, Zipfel C, Dorey S, et al. The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium *Burkholderia phytofirmans* and plant pathogenic bacteria. New Phytol. 2014:**201**(4):1371–1384. https://doi.org/10.1111/nph.12592

USDA, National Agricultural Statistics Service. 2020. Citrus Fruits: 2020 Summary.

Uzun A, Yesiloglu T. Genetic diversity in Citrus. London (UK): InTech; 2012

Veluchamy S, Hind SR, Dunham DM, Martin GB, Panthee DR. Natural variation for responsiveness to flg22, flgII-28, and csp22 and *Pseudomonas syringae* pv. tomato in heirloom tomatoes. PLoS One 2014:**9**(9):e106119. https://doi.org/10.1371/journal.pone.0106119

Vetter MM, Kronholm I, He F, Häweker H, Reymond M, Bergelson J, Robatzek S, de Meaux J. Flagellin perception varies quantitatively in *Arabidopsis thaliana* and its relatives. Mol Biol Evol. 2012:**29**(6): 1655–1667. https://doi.org/10.1093/molbev/mss011

Wang N. The citrus Huanglongbing crisis and potential solutions. Mol Plant. 2019:**12**(5):607–609. https://doi.org/10.1016/j.molp.2019.03.008

Wang L, Albert M, Einig E, Fürst U, Krust D, Felix G. The pattern-recognition receptor CORE of Solanaceae detects bacterial cold-shock protein. Nat Plants. 2016:**2**(12):1–9. https://doi.org/10.1038/nplants.2016.185

Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, Wu G, Tian S, Chen C, Liu Y, et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 2017:**49**(5):765–772. https://doi.org/10.1038/ng.3839

Wang P, Zhou L, Jamieson P, Zhang L, Zhao Z, Babilonia K, Shao W, Wu L, Mustafa R, Amin I, et al. The cotton wall-associated kinase GhWAK7A mediates responses to fungal wilt pathogens by complexing with the chitin sensory receptors. Plant Cell 2020:**32**(12): 3978–4001. https://doi.org/10.1105/tpc.19.00950

Wei Y, Balaceanu A, Rufian JS, Segonzac C, Zhao A, Morcillo RJL, Macho AP. An immune receptor complex evolved in soybean to perceive a polymorphic bacterial flagellin. Nat Commun. 2020:**11**(1): 3763. https://doi.org/10.1038/s41467-020-17573-y

Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, Domingo C, Tadeo FR, Carbonell-Caballero J, Alonso R, et al. Genomics of the origin and evolution of Citrus. Nature 2018:**554-**(7692):311–316. https://doi.org/10.1038/nature25447

Xue D-X, Li C-L, Xie Z-P, Staehelin C. LYK4 is a component of a tripartite chitin receptor complex in *Arabidopsis thaliana*. J Exp Bot. 2019:**70**(19):5507–5516. https://doi.org/10.1093/jxb/erz313

Zhang W, Fraiture M, Kolb D, Löffelhardt B, Desaki Y, Boutrot FFG, Tör M, Zipfel C, Gust AA, Brunner F. *Arabidopsis* RECEPTOR-LIKE PROTEIN30 and receptor-like kinase SUPPRESSOR OF BIR1–1/EVERSHED mediate innate immunity to necrotrophic fungi. Plant Cell 2013:**25**(10):4227–4241. https://doi.org/10.1105/tpc.113.117010

Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat Protoc. 2006:**1**(2):641–646. https://doi.org/10.1038/nprot.2006.97

Zhong G, Nicolosi, E. Citrus origin, diffusion, and economic importance. In: **Gentile A, La Malfa S, Deng Z**, editors. The Citrus genome, compendium of plant genomes. Cham: Springer International Publishing. 2020. pp. 5–21. https://doi.org/10.1007/978-3-030-15308-3_2