Characterization of a Natural *Arabidopsis thaliana* – *Pseudomonas viridiflava* Pathosystem

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Alejandra Duque Jaramillo

aus Medellín, Kolumbien

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	04.10.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Detlef Weigel
2. Berichterstatter:	Prof. Dr. Georg Felix

Table of contents

Summary

Disease in plants can be caused by a variety of microorganisms, including bacteria. To fight infection, plants are equipped with an immune system that recognizes pathogens and activates a defense response mediated by the hormones salicylic acid and/or jasmonic acid and ethylene. One of the most important bacterial plant pathogens are strains from the *Pseudomonas* genus, able to infect crops and wild plants. The *Pseudomonas syringae* complex comprises most of the phytopathogens of this genus, including the model strain *P. syringae* pv. *tomato* DC3000 (DC3000), widely used in pathogenicity studies.

P. viridiflava, a globally-distributed natural pathogen of the model plant *Arabidopsis thaliana*, also belongs to the *P. syringae* complex but is genetically and phenotypically distinct from well-characterized DC3000. Despite *P. viridiflava* being the most abundant *Pseudomonas* species in *A. thaliana* populations, little is known about the mechanisms of bacterial virulence and plant resistance in this pathosystem. In this thesis, I characterized the natural *A. thaliana* - *P. viridiflava* pathosystem by combining genetics, transcriptomics and metabolomics to identify resistance mechanisms in the host. I also used a computational framework to identify virulence-related specialized metabolites in the pathogen.

In the first chapter, I investigated how *P. viridiflava* interacts with *A. thaliana*, and contrasted this with the model pathogen DC3000. I uncovered that the jasmonic acid/ethylene pathway is involved in defense against *P. viridiflava*, likely through an increase in jasmonic acid levels. Infection elicited a similar response in resistant and susceptible hosts, but the timing was different: changes occurred faster in the resistant host. In the second chapter, I explored how potential specialized metabolites encoded by *P. viridiflava* might be associated with differences in their virulence. I described the large biosynthetic potential of a collection of *Pseudomonas* genomes from the *A. thaliana* phyllosphere, and found that this biosynthetic potential is dominated by non-ribosomal peptide synthetases. I then identified gene cluster families with a putative role in *P. viridiflava* virulence, one of them related to the siderophore pyoverdine.

Overall, this thesis presents an integrative approach to the study of plant-microbe interactions, and provides the baseline for further studies on the interactions between *A. thaliana* and *P. viridiflava*. This pathosystem better represents the interaction dynamics in natural populations and has the potential to address ecologically-relevant questions about adaptation and co-evolution of host and pathogen.

Zusammenfassung

Pflanzenkrankheiten können durch eine Vielzahl von Mikro-Organismen, einschließlich Bakterien verursacht werden. Diese Mikroorganismen können durch das pflanzliche Immunsystem erkannt werden und die Infektion wird über die Hormon-induzierte Antwort mittels Salicylsäure und/oder Jasmonsäure und Ethylen bekämpft. Zu den wichtigsten bakteriellen Pflanzenpathogenen gehören Stämme der Gattung *Pseudomonas*, welche beides Nutz- und Wildpflanzen infizieren können. Der *Pseudomonas syringae*-Komplex umfasst die meisten Phytopathogene, darunter den Modellstamm *P. syringae* pv. *tomato* DC3000 (DC3000), der häufig in pathologischen Studien verwendet wird.

P. viridiflava, ein weltweit verbreiteter natürlicher Erreger der Modellpflanze *Arabidopsis thaliana*, gehört ebenfalls zum *P. syringae*-Komplex, dieser unterscheidet sich genetisch und phänotypisch vom gut charakterisiertem DC3000. Obwohl *P. viridiflava* die häufigste *Pseudomonas*-Art in *A. thaliana*-Populationen ist, sind die Mechanismen der bakteriellen Infektion und der Pflanzenresistenz in diesem Pathosystem bisher nicht ausreichend erforscht. In dieser Arbeit habe ich das natürliche *A. thaliana* - *P. viridiflava* Pathosystem durch die Kombination von Genetik, Transkriptomik und Metabolomik charakterisiert, um Resistenzmechanismen auf dem Wirt zu identifizieren. Zudem verwendete ich ein computergestütztes Rahmenwerk, um virulenzbezogene spezialisierte Metaboliten im Pathogen zu identifizieren.

Das erste Kapitel befasst sich mit der Interaktion zwischen *P. viridiflava* mit *A. thaliana*, diese Interaktion wird mit dem Modellpathogen DC3000 verglichen. Es zeigte sich aufgrund des Anstieges vom Jasmonsäurespiegels, dass der Jasmonsäure-Ethylen-Stoffwechselweg an der Abwehr von *P. viridiflava* beteiligt ist. Die Infektion löste in resistenten und anfälligen Wirten eine ähnliche Reaktion aus, aber der Zeitpunkt war unterschiedlich: In den resistenten Wirten traten die Veränderungen schneller auf. Der Inhalt des zweiten Kapitels umfasst die Thematik der Untersuchung von Zusammenhängen zwischen potentiellen spezialisierten Metaboliten, die in *P. viridiflava* kodieren, und den Grad der Infektion. Beschrieben wird das biosynthetische Potenzial einer Sammlung von *Pseudomonas*-Genomen aus der *A. thaliana* Phyllosphäre. Es zeigte sich, dass dieses Potenzial von nicht-ribosomalen Peptidsynthetasen dominiert wird. Identifiziert wurden Gencluster-Familien, die vermutlich eine Rolle bei der Virulenz spielen, eine davon im Zusammenhang mit dem Siderophor Pyoverdin.

Ziel dieser Arbeit ist die Untersuchung des integrativen Ansatzes zu den Wechselwirkungen zwischen Pflanzen und Mikroben, der die Grundlage für weitere Studien über die Wechselwirkungen zwischen *A. thaliana* und *P. viridiflava* bildet. Dieses Pathosystem stellt eine natürliche Situation besser dar und hat das Potenzial, ökologisch relevante Fragen zur Anpassung und Koevolution von Wirt und Pathogen zu beantworten.

Introduction

1. The model organism *Arabidopsis thaliana*

Arabidopsis thaliana is a small, annual or winter annual weed native to Eurasia and North Africa. It belongs to the Brassicaceae family, to which cabbage, broccoli, rapeseed, and radish also belong. It was first proposed as a model organism for genetic research by Friedrich Laibach in the 1940s (1); however, its widespread adoption as a model organism occurred only in the 1980s with the advent of molecular genetics research (2,3).

Among the characteristics that make *A. thaliana* a suitable model organism are its short generation time, the feasibility of growing it under laboratory and greenhouse conditions, and the production of large amounts of progeny, i.e., seeds. A genome size of only 135 Mb arranged in 5 chromosomes, efficient transformation and mutation protocols, and a large collection of natural and mutant lines available further facilitate research in this plant. Moreover, *A. thaliana* is a selfing plant, and due to its high level of homozygosity (4), selfing produces identical sibships. This greatly facilitates genetic studies, since a line needs to be sequenced only once but can be scored for a myriad of phenotypes. Despite its selfing nature, *A. thaliana* can be manually outcrossed, resulting in a population that enables co-segregation studies of a phenotype of interest with gene markers. In addition to these intrinsic characteristics, there is a large community of *A. thaliana* researchers that generate and maintain databases and online tools covering genomics, transcriptomics, epigenomics, and more for several *A. thaliana* lines (2,3).

Even though genome sequencing and analysis of natural variation indicate that there is not a single 'reference' *A. thaliana* genotype, three *Arabidopsis* lines are generally regarded as 'reference' wild-type lines: Columbia-0 (Col-0), Landsberg *erecta* (L*er*) and Wassilewskija (Ws-1). Col-0 was the first *A. thaliana* genome sequenced, and of any land plant in fact (5). It is the most commonly used as reference genotype since, in addition to high-quality genetic and genomic data, there is an extensive collection of mutants in this genetic background, together with physiological and biochemical knowledge (2).

1.1. *Arabidopsis thaliana* genetic diversity and resources

1.1.1. Natural genetic diversity

Already in the early days of *A. thaliana* research, Friedrich Laibach and his students emphasized the use of natural variation in the study of physiological traits such as flowering time and seed dormancy (6,7). *A. thaliana* lines, known as accessions, differ not only in shape and development, but also in their physiology. The natural variation of several phenotypes has been studied in *A. thalian*a, including developmental traits such as flowering time, number of siliques or seeds, response to biotic and abiotic stress, and content of metals and oil (8).

The wide range of *A. thaliana* distribution implies that the species has successfully adapted to various environments. This adaptation should be, at least partly, reflected in the genetic variation between accessions, especially in adaptive traits such as development and resistance to biotic and abiotic stresses (8,9). The natural genetic diversity of the *A. thaliana* species is well characterized, most recently thanks to the efforts of the 1001 Genomes project (10). Cao and colleagues (11) carried out the first phase of this project, where 80 naturally-inbred lines collected from eight regions throughout the *A. thaliana* native range were sequenced, and single nucleotide polymorphisms (SNPs), indels and structural variation to the reference *A. thaliana* genome were identified. This collection was later used to investigate hybrid incompatibility (12), for which crosses of all combinations among these 80 lines were generated. The work of Cao and colleagues is of particular relevance for this thesis since the accessions included here belong to this collection.

The 1001 Genomes project sequenced 1135 *A. thaliana* naturally-inbred lines, covering both the native range of the species and recently-colonized North America. Nine admixture groups that broadly correspond to the geographical origin of the accessions were defined; one of these groups being 'relict' lines, coming mostly from the Iberian Peninsula and with extreme pairwise divergence to the other accessions (10). Data for these thousand *A. thaliana* genomes are available and can be leveraged for genetic mapping studies due to the selfing nature of the species. The cost of sequencing a large number of lines can be a limiting step for researchers, but they can benefit from this publicly-available dataset and exploit it to find the genetic basis of their trait of interest.

The study of natural variation extends our knowledge of biological mechanisms and pathways, their regulation and interactions. For example, the presence of null or weak alleles in the more traditional reference genotypes can prevent the occurrence of a particular phenotype; this can be circumvented by the inclusion of other accessions with a different genetic background (9). Moreover, certain interactions between genes may arise only in particular genetic backgrounds. By including natural genetic variation can we comprehend how a species adapts to different environments and stresses, and identify genetic traces of said adaptation (8). In addition, natural genetic variation permits the identification of genes underlying a phenotype of interest by using quantitative trait locus (QTL) and genome-wide association studies (GWAS; see also section 3.3)

1.1.2. Mutant lines

In addition to natural genetic diversity, there is a large collection of *A. thaliana* mutant lines generated experimentally. For this, a large number of seeds are treated with a mutagen, and then screened for gain or loss of a phenotype of interest. Mutagens can be physical, chemical or biological (13), common mutagens in *A. thaliana* research include ethyl methanesulfonate (EMS) and radiation. Another approach for generating mutant lines consists in transforming the plant with *Agrobacterium tumefaciens*

carrying a tumor-inducing plasmid. In this case, the T-DNA is transferred from *A. tumefaciens* to the host plant's nuclear genome, where it can disrupt a gene.

Many of these mutant lines are publicly available from stock centers (13). This is an invaluable resource that can be exploited almost infinitely. Particularly relevant for this work are immune-deficient mutants, which have been extensively used to characterize the plant immune system and the immune response against different stimuli.

1.2. Plant immunity

Disease outbreaks in plants can be caused by a variety of (micro)organisms. To fight infection, plants are equipped with two lines of defense: first, cell-surface Pattern Recognition Receptors (PRRs) that detect microbe- or pathogen-associated molecular patterns (M/PAMPs); and second, intracellular receptors encoded by Resistance (R) genes that recognize pathogen effector proteins (14). The defense response elicited upon activation of these receptors is called pattern-trigger immunity or effector-triggered immunity, respectively.

1.2.1. PTI: pattern-triggered immunity

Cell-surface PRRs recognize conserved molecular patterns of microbes, pathogens, damage and herbivory (M/P/D/HAMPs), thus activating pattern-triggered immunity (PTI; 14). PTI is characterized by cytosolic calcium influx, production of reactive oxygen species (ROS), and stomatal closure and callose deposition to prevent further pathogen penetration (14). The activation of calcium-dependent and mitogen-activated protein kinases (CPKs and MAPKs) signaling pathways activates transcription factors such as WRKY, leading to the expression of defense-related genes and to the biosynthesis of antibiotic compounds and defense-related hormones (14).

1.2.2. ETI: effector-triggered immunity

Pathogens can overcome PTI through effector proteins, resulting in effector-triggered susceptibility. In turn, plants have also evolved mechanisms to recognize effectors. In particular, resistance (R) genes encode intracellular nucleotide-binding leucine-rich repeat (NLR) proteins that recognize specific pathogen effectors (14), leading to effector-triggered immunity (ETI). NLRs are among the most diverse genes in plant genomes (8). Indeed, a recent study of *A. thaliana*'s pan-NLRome confirmed that the species-wide NLR diversity is high (15).

Pathogen effectors can be recognized directly by the host NLRs, or indirectly when the signal detected is a consequence of the effector's action on the host cell (16). ETI can be liked to a stronger PTI and results in disease resistance, usually through a hypersensitive response (HR), including cell death at the site of infection (16). Activation of helper-NLR-independent NLRs, such as ZAR1, results in cytosolic calcium influx, ROS accumulation and changes in chloroplasts and vacuoles, eventually leading to HR and cell rupture (17). However, the majority of NLRs require

a helper NLR to mediate immunity (14). Although the outcome of helper-NLR-mediated signaling is similar to that of helper-NLR-independent, e.g., calcium influx, transcriptional remodeling, and HR, this is achieved through different mechanisms.

1.2.3. Hormone-mediated immune signaling

Activation of the plant's immune response by the recognition of a pathogen results in hormone-mediated signal transduction. The main phytohormones involved in defense are salicylic acid (SA), jasmonic acid (JA) and ethylene (ET). These signaling cascades cause changes in the expression of defense-related genes and in the production of antimicrobials to limit pathogen growth (18).

Which particular hormone signaling pathway is activated depends on the lifestyle of the pathogen detected (Figure 1). The SA pathway, for which *PR1* expression is a marker, is activated to achieve resistance against biotrophic pathogens, which thrive on living plant tissue. Necrotrophic pathogens, on the other hand, feed on dead tissue and defense against them is mediated by the JA/ET signaling pathway, resulting in the expression of *VEGETATIVE STORAGE PROTEIN 1/2* (*VSP1/2*) and *PLANT DEFENSIN 1.2* (*PDF1.2*) marker genes (18; Figure 1). This pathway is also activated by insect herbivory and physical damage (18,19).

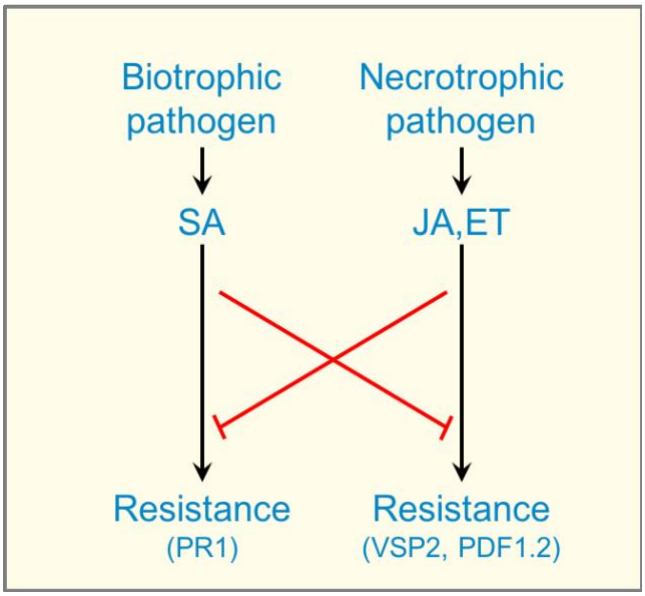

Figure 1 - Hormone-mediated defense signaling pathway.

SA-mediated defense pathway

Recognition of biotrophic pathogens results in SA synthesis and accumulation, leading to the activation of the SA signaling pathway. There are two pathways for SA production in plants: one involving an isochorismate synthase (ICS) and the other, a phenylalanine ammonia-lyase (PAL). Both use chorismate as the precursor for SA. The contribution of these pathways to SA production varies among plant species (18,20). In *Arabidopsis*, the ICS pathway is the most important, and thus the focus of

this section. Chorismate is converted to isochorismate by an isochorismate synthase (ICS), for which there are usually 1 - 2 homologs in plants (20). *A. thaliana* has two ICS homologs, *ICS1* and *ICS2*, and ICS1 is the major contributor to SA-related responses (20). *ICS1* is also known as *SID2* and *EDS16.* Isochorismate undergoes amino acid conjugation, followed by spontaneous decomposition or enzymatic conversion, resulting in SA.

SA biosynthesis and accumulation is mediated by EDS1 and PAD4 for some, but not all, stimuli. Downstream of these genes, ICS1 (SID2) and EDS5, involved in the transport of SA biosynthesis intermediates, are also required (18). SA accumulation is then transduced through NPR1, which upon an increase in SA levels enters the nucleus and interacts with TGA transcription factors (18). Together with transcription factor WRKY70, TGAs result in the expression of the marker gene *PR1* and defense responses (18). There are SA-dependent responses that are independent of NPR1, but this branch of the SA signaling pathway has not yet been characterized (Figure 2; 18).

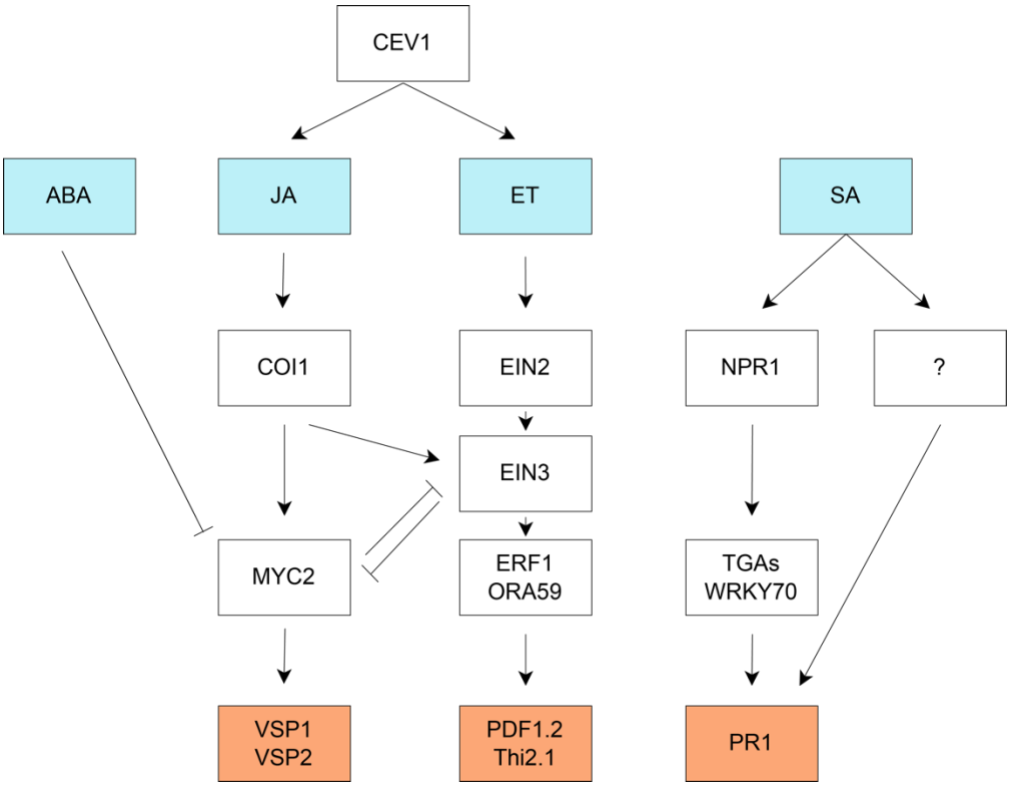

Figure 2 - Crosstalk in the hormone-mediated defense pathways.

JA/ET-mediated defense pathway

JA synthesis is promoted in response to necrotrophic pathogens. In *A. thaliana*, there are three pathways for JA biosynthesis, the two main ones being the octadecane and the hexadecane pathways (21). The biosynthesis of JA occurs in three cell

compartments. First, the unsaturated fatty acids α-linoleic acid (18:3) or hexadecatrienoic acid (16:3) are converted to 12-oxo-phytodienoic acid (OPDA) or deoxymethylated vegetable dienic acid (dinor-OPDA) respectively in the chloroplast. These compounds are transported to the peroxisome, where they are reduced and β-oxidized, resulting in JA. Finally, JA is exported to the cytoplasm, where it undergoes conjugation to achieve its active forms including JA-Ile and methyl-JA (21,22).

The other actor in this pathway is ethylene, a gaseous hormone. It is synthesized from methionine by the Yang cycle (23). In the absence of ethylene, its receptors activate CTR1, which represses downstream signaling. Binding of ethylene to its receptors releases the inhibition by CTR1, so that EIN2 can activate the transcription factors EIN3, EIL1 and EIL2, which then activate ET-responsive genes including ERF transcription factors (24; Figure 2).

JA signaling is transduced via COI1 (18). In the absence of JA-Ile, the repressor JAZ binds to the transcription factor MYC2 and prevents the expression of JA-responsive genes (25). JA-Ile binds to COI1 and promotes the ubiquitination and degradation of JAZ, thus activating the signaling pathway (25). The JA signaling pathway downstream of COI1 can be divided into two branches: the MYC and the ERF branch (23; Figure 2). The MYC branch is activated in response to wounding and insect herbivory, *MYC2/3/4* are its transcription factors (23). Activation of the MYC branch leads to the expression of JA-responsive genes such as *VSP2*. The ERF branch is activated by necrotrophic pathogens. The transcription factor *ERF1* integrates signals from both ET and JA, and its expression requires both *COI1* and *EIN2* (18). *PDF1.2* is a marker gene for the activation of this branch (23).

The MYC and ERF branches are antagonistic (23). This behavior is mediated by the interaction between MYC2 and EIN3 (22); indeed, MYC2 and ERF1 induce different sets of genes, usually in a mutually-exclusively manner (18). MYC2 negatively regulates the expression of ERF-branch genes *ERF1*, *ORA59* and *PDF1.2* and thus, resistance against necrotrophic pathogens (22).

Crosstalk between the SA and JA/ET pathways

The SA and JA/ET pathways act mainly antagonistically, meaning that upregulation of the SA pathway inhibits the JA/ET pathway and vice versa (18,23). This feature is exploited by pathogens to prevent a resistance response from the plant host (see section 2.3). Synergistic interaction between SA and JA pathways has been demonstrated as well, although this interaction is not well understood. Low concentrations of SA and JA can result in the expression of both *PR1* and *PDF1.2* marker genes (26), and both pathways can be activated simultaneously in distinct cellular spaces (27). Moreover, a recent meta-analysis identified not only genes responding exclusively to SA or JA, but also genes shared between SA and JA signaling (28). Genes induced by both SA and JA relate to broad-spectrum defense

response, while genes repressed by both hormones related to photosynthesis, auxin and gibberellin, are likely to promote defense instead of growth (28).

SA repression of the JA/ET pathway occurs downstream of JA biosynthesis, mainly at the transcriptional level (23,29). Components of the SA signaling pathway repress different components of the JA/ET pathway and its ERF branch, and this inhibition appears to occur at the *ORA59* promoter (22,23). SA can also mediate the sequestration of JA-activated transcription factors and the degradation of transcription factors that activate the JA signaling pathway (30). On the other hand, the JA/ET pathway inhibits the SA response mainly at the SA biosynthesis and accumulation level (23,31). Transcription factors from the NAM/ATAF/CUC (NAC) family that are activated by MYC2 are able to repress the expression of the SA biosynthetic gene *ICS1*, and promote the expression of the SA metabolism gene *BSMT1*, preventing SA accumulation (22,32). MYC2 regulates the expression of the transcription factor ZAT18, which directly represses *EDS1,* hence decreasing SA accumulation (31).

Other phytohormones

In addition to SA, JA and ET, other plant hormones such as auxins, abscisic acid (ABA), brassinosteroids, gibberellins, and cytokinins play a role in plant defense against pathogens. They can fine-tune the defense response by modulating the SA and JA/ET pathways (23). For their relevance for this thesis, I will focus here only on ABA and auxins.

ABA accumulates after infection and can have both positive and negative effects on defense responses (25,33). It can act antagonistically to SA and JA/ET, thus increasing plants' susceptibility to disease and herbivory. For example, ABA suppresses the expression of the SA biosynthetic gene *ICS1* (*SID2*) and promotes the degradation of NPR1 (34). ABA promotes the interaction between its receptor PYL6 and MYC2, preventing the binding of MYC2 to its target genes (22). In addition, ABA promotes resistance by inducing stomatal closure to prevent further pathogen entry to the leaf interior, and later in infection it can either promote or repress callose deposition (25). In general, ABA seems to promote defense responses and resistance in the early time points of infection, while inhibiting said responses in the later ones. However, the role of ABA on disease resistance depends not only on the stage of the defense response but also on the specific plant-pathogen interaction and the infected tissue (33).

Auxins are low-molecular-weight compounds, the predominant auxin in plants being indole acetic acid (IAA; 25). Auxins can negatively affect defense responses by interfering with other signaling pathways or with PTI (35). Exogenous auxin application promotes disease caused by biotrophic pathogens, suggesting auxins and SA act antagonistically (25,35). In line with this, some pathogens produce their own auxins or induce their synthesis by the plant (25) to increase host susceptibility to infection.

1.3. The phyllosphere

The phyllosphere refers to the above-ground portion of the plants, which constitutes around 60% of the biomass on earth (36). It is inhabited by microorganisms, including bacteria, despite it being a harsh environment due to water scarcity, temperature and UV exposure (36). In particular, the leaf harbors bacteria that inhabit its surface, known as epiphytes, while those residing in the apoplast, the space between the plant cells, are called endophytes. Commensal microorganisms are predominantly epiphytic, while pathogens enter the leaf to cause disease, i.e. they are endophytes (37).

The microbial community in the phyllosphere is not a random assembly of microorganisms, but rather a selected set that varies depending on the plant species and temporal and spatial factors (37). The phyllosphere microorganisms can originate from soil, air, water, seeds and nearby plants (38). From the members of the phyllosphere, bacteria are the most widely studied. *Proteobacteria*, in particular *Alphaproteobacteria*, constitute up to 70% of the phyllosphere of soybean (*Glycine max*), white clover (*Trifolium repens*) and rice (*Oryza sativa*) (37). In *A. thaliana*, they represent 38%, followed by *Actinobacteria* (25%) and *Bacteroidetes* (21%; 37). Of the *Gammaproteobacteria*, the genus *Pseudomonas* is the most abundant in soybean, clover and *A. thaliana*.

2. Plant-pathogenic *Pseudomonas*

Pseudomonas is one of the most diverse genera of gram-negative bacteria. It was first described in 1894 and it currently contains 254 species (39), including free-living species as well as non-pathogenic and pathogenic species able to infect not only plants but also animals, including humans (40). Its taxonomy is complex, and a recent analysis of the genus based on ten thousand *Pseudomonas* genomes suggests that *Pseudomonas* is actually a mixture of genera (39).

Multilocus sequencing analysis (MLSA) is the most widely used method for phylogenetic and taxonomic classification, since the 16S rRNA gene is not always able to differentiate between closely-related bacterial species (40). Housekeping genes commonly used for MLSA, including *atpD, gyrB, gapA, carA, cts, rpoB, recA* and *rpoD*. Another approach to comparing genetic relatedness between organisms, facilitated by the reduction of sequencing costs, is the calculation of their average nucleotide identity (ANI; 41), which correlates strongly with MLSA (40).

Most of the phytopathogens of the *Pseudomonas* genus belong to the *Pseudomonas syringae* species complex described below.

2.1. The *Pseudomonas syringae* species complex

The *Pseudomonas syringae* complex contains more than 15 species and 60 pathovars, which are defined by the ability of a given bacterial species or strain to infect a particular host (40,42,43). In 2012, *Pseudomonas syringae* pathovars were named

the top plant pathogenic bacteria in molecular plant pathology, based on its importance to basic science and its impact on food production and/or the environment (44). Indeed, diseases caused by members of the *P. syringae* complex are the most common emergent bacterial disease (45). Bacteria belonging to the *P. syringae* complex, and in fact to the *Pseudomonas* genus in general, have been isolated not only from plants, but also from environmental sources such as rivers, lakes, clouds, rain, and snow; this, in addition to its ability to promote ice enucleation, suggests that *P. syringae* is linked to the water cycle (46).

The *P. syringae* complex is a genetically diverse monophyletic group (43,45). Only 3%, around 2500 genes, of the *P. syringae* complex pan-genome, the collection of all genes identified in the *P. syringae* complex strains, is present in more than 95% of the strains (47). Furthermore, the genetic diversity of the *P. syringae* complex is not yet saturated: a new genome encodes approximately 190 new ortholog groups (47). This indicates that many ortholog groups, likely with a sparse distribution, remain to be discovered (47).

The species is divided into 13 phylogroups (PG1-13) based on MLSA of four housekeeping genes: *cts*, *gapA*, *rpoD* and *gyrB* (43). Strains within a PG have a genetic distance of less than 5% based on these genes (48). These phylogroups are classified as primary/canonical and secondary/non-canonical phylogroups (43). Y nPrimary phylogroups, PGs 1-6 and 10, include the best studied *P. syringae* strains. Most of these strains have been isolated from plant hosts and have a canonical, tripartite pathogenicity island (T-PAI; see next section) (43,49). Meanwhile, many strains from the secondary phylogroups, PGs 7-9 and 11-13, have been isolated also, or exclusively, from environmental sources (43) and contain either a T-PAI or a S-PAI, a single pathogenicity island encoding only the *hrp/hrc* cluster that was first described in *Pseudomonas* (47,49).

Primary phylogroups are late-branching in the *P. syringae* complex phylogenetic tree, while secondary phylogroups are early-branching and thus, more divergent (47). This structure is maintained in part by recombination occurring more often within primary phylogroups than between primary and secondary phylogroups (47). The *P. syringae* complex is not only genetically, but also phenotypically diverse. Phenotypes vary both among and within phylogroups, making phenotypic characterization insufficient to classify strains into phylogroups (48,50). Many of the variable phenotypes are related to virulence, the relative capacity of a strain to cause disease in a particular host.

On their plant host, *P. syringae* bacteria have two phases: an epiphytic phase that is followed by an endophytic one (43). Within the *P. syringae* complex, there are both strong and weak epiphytes (43), although disease occurs only in the endophytic phase, after *P. syringae* colonizes and divides in the apoplast inside the leaf. The host range of the *P. syringae* complex is broad, but it is not well established if strains are

generalists or specialists (45). The classification of pathovars suggests a restricted host range for particular strains, however, this is rarely assessed. Morris et al. showed recently that strains from the *P. syringae* complex form an overlapping continuum of host range potential, with individual strains having narrow, moderate or broad host ranges (45). In this light, the pathovar classification is inaccurate for the *P. syringae* complex.

2.2. Being a pathogen: virulence mechanisms

Pathogenic bacteria cause disease after proliferating in the apoplast, the intercellular space between plant cells, which they reach through wounds or natural openings such as stomata and hydathodes (16). To successfully invade, colonize and proliferate inside their host, bacteria have virulence factors such as effectors, toxins and phytohormones.

Effectors contribute to pathogen virulence by mimicking or inhibiting the host cellular functions, in particular those related to immunity (16). They are encoded by so-called avirulence (*avr*) genes and are secreted into the host cell via secretion systems. *P. syringae* strains use a type 3 secretion system (T3SS) for the injection of effectors into the host cells. The T3SS is encoded and regulated by 26 genes in the *hrp/hrc* cluster located in a pathogenicity island (51). In addition to the T3SS, pathogenicity islands encode a varying number of effectors in two loci: the conserved effector locus (CEL), and the variable exchangeable effector locus (EEL). Effectors do not only manipulate the host immune response to avoid resistance, but can also promote conditions in the plant that favor pathogen survival and growth, such as promoting an aqueous apoplast (52). As mentioned above, recognition of effectors by plant R proteins/NLRs results in ETI, an immune response to halt infection (53,54).

The bacterial effector diversity of the *P. syringae* complex is large. Dillion and colleagues identified 4636 unique effectors at the amino acid level in a collection of 494 isolates from more than 100 host and environmental sources (55). These effectors are classified in 70 families. Individual strains have between 1 and 59 putative effectors, with a mean of 29.58 ± 10 (55). The distribution of effectors varies between phylogroups. Primary PGs encode more effectors than secondary PGs in general, which is in line with secondary PGs strains having a single, non-canonical pathogenicity island (55). Strains of PG 1 encode the largest number of effectors, while PGs 2 and 10, and a subset of PG 3 strains, have the smallest (48,55). Most effector families are present only in a small subset of the isolates, highlighting the diversity in the effector repertoire size and composition within the complex. Only four effectors, AvrE, HopB/HopAC, HopM, HopAA are present in >95% of the strains (55), all but HopB/HopAC located in the CEL. Although effector content is not a good predictor of host specificity, closely related strains tend to have more similar effector repertoires (55).

In addition to effector proteins, toxins also promote virulence inside the host. They can be host-specific or be active in a wide range of hosts (56). Toxins encoded by *P. syringae* complex strains generally lack host specificity and include syringopeptin, syringomycin, tabtoxin, phaseolotoxin, and coronatine (56,57). Some toxins, such as phaseolotoxin and coronatine, are widely distributed across the species complex, while others are restricted to specific phylogroups (47). In general, most strains have the biosynthetic machinery for one or two toxins, except for PG 2 strains, which encode up to 5 toxins (47). These strains usually have a small effector repertoire, suggesting they rely more on toxins for virulence compared to other primary PGs (47,48,51,58). Indeed, there is a negative correlation between the presence of the syringomycin gene cluster and a small effector repertoire (58). Toxins have diverse modes of action: they can promote the opening of stomata to allow bacterial entry (e.g. coronatine, syringolin), manipulate the plant's immune system to avoid resistance (e.g. coronatine, phevamine A), cause membrane disruption and ion leakage (e.g. syringomycin and syringopeptin) and modify the host plant metabolism (e.g. tabtoxin, phaseolotoxin) (43,51,56,59–61). Other virulence factors of the *P. syringae* complex include the production of phytohormones or enzymes for their degradation, particularly auxin, pectate lyases that degrade plant cell walls and exopolysaccharides involved in biofilm formation and adhesion (43,51).

In the following section, I will use the widely-studied model strain *Pseudomonas syringae* pv. *tomato* DC3000 to illustrate the action of the virulence factors described in the above.

2.3. The model strain *Pseudomonas syringae* pv. *tomato* DC3000

P. syringae pv. *tomato* DC3000 (henceforth DC3000) belongs to the *P. syringae* species complex, specifically to PG 2, and is widely used in pathogenicity studies(62–64). It encodes 36 effectors, eight of which constitute its minimal effector repertoire: AvrPtoB, HopM1, AvrE, HopE1, HopG1, HopAM1-1, HopAA1-1 and HopN1 (65). These effectors suppress plant immunity and promote an aqueous apoplast. AvrPtoB inhibits PTI and ETI by targeting PPRs and immune-associated kinases for degradation or inactivation, while HopG1 and HopE1 inhibit immune responses such as callose deposition by interacting with the cytoskeleton components actin and microtubules, respectively (43). HopN1 prevents the production of ROS and callose deposition, the latter suppressed by HopAM1 as well (43). AvrE and HopM1 manipulate the ABA pathway to promote stomatal closure and create an aqueous environment suited for bacterial growth (66,67).

Upon entry into the apoplast space, DC3000 uses the toxin coronatine to overturn the closing of stomata induced by pathogen recognition (43). Effectors AvrB, HopF and HopM1 can also suppress PTI-induced stomatal closure (43). The apoplast is a hostile environment, thus, bacterial pathogens must modify it. The effectors AvrE and HopM1 create an aqueous apoplast by promoting water and nutrient release from the plant cells (52). AvrE and HopM modify the apoplast through manipulation of the ABA

signaling pathway; although functionally redundant, they are not genetically related (66,67).

The toxin coronatine is a structural homolog of JA-Ile and as such, can bind to COI1 with high affinity, resulting in JAZ degradation and de-repression of the JA signaling pathway (43). With coronatine, DC3000 takes advantage of the antagonistic nature between the SA and the JA defense signaling pathways: by activating the MYC2 branch of the JA pathway, it represses the SA pathway, which mediates host resistance against this pathogen (32). The transcription factors activated by coronatine repress *ICS1* and activate *BSMT1*, which are involved in SA biosynthesis and metabolism respectively (32). In addition, effectors AvrB, HopX and HopZ interfere with JAZs proteins, the repressors of the JA signaling pathway, leading to activation of the JA pathway (22). Together, this results in suppression of SA-mediated resistance and therefore, an increase of plant susceptibility to infection.

Other *P. syringae* strains, such as *P. syringae* pv. *syringae* B728a, encode the toxin syringomycin (68), which contributes to make the apoplast a bacteria-friendly environment by inducing pore formation on the plant membranes to release nutrients (69). It is also a surfactant that can facilitate bacterial movement on the leaf surface (69). DC3000, however, lacks the genes for the biosynthesis of known *P. syringae* toxins such as syringomycin, syringotoxin, syringostatin, syringopeptin and pseudomycin (70)

2.4. Specialized metabolites

In addition to effectors, bacteria can produce a range of specialized metabolites, also known as secondary metabolites, that may contribute to virulence. One example of specialized metabolites are the toxins described above. Specialized metabolites are not essential for bacterial growth, but facilitate the interaction of an organism with its environment; they are involved in nutrient acquisition, quorum sensing, defense and virulence. Bacterial specialized metabolites include phytohormones, antibiotics, pigments, toxins, terpenes and siderophores (57,71). They are classified based on the enzyme-class involved in the biosynthesis of the first intermediate (57).

Specialized metabolites are produced by biosynthetic gene clusters, defined as “a physically clustered group of two or more genes in a particular genome, that together encode a biosynthetic pathway for the production of a specialized metabolite (including its chemical variants)” (72). Biosynthetic gene clusters are diverse and rapid-evolving (73), but their architecture is highly conserved and enables their prediction using genome-mining approaches (section 2.4.1). They are usually considered part of the accessory genome and are prone to horizontal gene transfer, mutation, recombination and duplication or excision events, all contributing to their diversity (73).

Specialized metabolites are frequently involved in plant-microbe interactions (74). Plant defense against pathogens, as well as pathogen virulence are commonly

affected by specialized metabolites. For instance, lipopeptides produced by *Streptomyces*, *Pseudomonas*, and *Bacillus* genera can act as biosurfactants, increasing the availability of water-insoluble compounds and promoting swarming (73,75). In addition, lipopeptides can have antimicrobial and cytotoxic activity, acting as toxins and contributing to pathogen virulence (75). In contrast, non-pathogenic bacteria can produce lipopeptides that induce resistance against microbial pathogens in the host (75). Moreover, several plant-associated bacteria, including *Pseudomonas*, can produce indole-3-acetic acid, an auxin phytohormone that is used for signaling, manipulation of the host development and defense response, or induction of systemic acquired resistance (57,76,77).

2.4.1. Prediction of biosynthetic gene clusters

A simple and common way of mining a genome is doing a similarity search using, for example, BLAST (78). Here, the sequence of a gene or protein sequence of known function is used as a query to identify orthologs in other (non-annotated) genome sequences, which could have a similar function (78). In 2011, the first comprehensive pipeline for the identification of biosynthetic gene clusters (BGCs) was published: the antibiotics and specialized metabolite analysis shell (antiSMASH) (79), which integrates tools previously designed for the identification of different classes of BGCs with new methods. antiSMASH is rule-based, it uses manually curated and validated rules that define which core biosynthetic functions have to be present in a genomic region for it to be a BGC (79,80).

antiSMASH takes as input a sequence file (FASTA, GBK or EMBL files), from which genes are extracted and predicted if necessary. Then, it uses hidden Markov models of signature proteins or protein domains that are specific for each class of BGC to identify potential gene clusters in the genome of interest (79). The first version of antiSMASH was released in 2011 and could predict 19 BGCs classes (79); 10 years later, the latest version has rules for 71 classes (80). Soon after its release, antiSMASH became the most widely used tool for genome mining of bacterial specialized metabolites, and it is regarded as a gold standard.

Using an expanded version of the antiSMASH algorithm, 10724 BGCs were predicted in 1154 bacterial genomes (81). The most common BGC class is saccharides, present in 93% of the genomes analyzed. Ribosomally synthesized and posttranslationally modified peptides (RiPPs) have an abundance similar to non-ribosomal peptide synthetases, and higher than polyketides and terpenes (81). On average, bacteria devote 3.7% of their genome to BGCs, having 2.4 gene clusters per Mb (81).

Genome mining was initially performed in a small number of genomes. Due to the decreasing cost of sequencing technologies, mining is now performed at much larger scales, such as an entire genus, strain collections and microbiomes (82). These approaches result in the identification of thousands of BGCs with varying degrees of

similarity, from widely distributed to rare or unique BGCs. In order to compare the BGCs of different genomes and to identify families of BGCs that are linked to a highly-similar specialized metabolite, BiG-SCAPE was developed (82). BiG-SCAPE is a "biosynthetic gene similarity clustering and prospecting engine" that performs BGC classification and clustering based on their sequence similarity into gene cluster families (82). BiG-SCAPE combines three metrics for similarity: the Jaccard index to measure protein domain content similarity, an adjacency index of synteny conservation between gene clusters, and a domain sequence similarity index that measures both Pfam domain copy number differences and sequence identity. BiG-SCAPE and antiSMASH are integrated, as the former takes the output of the latter as input; in addition, the minimum information of a biosynthetic gene cluster (MIBiG) reference data can be included automatically in the analysis. This database contains BGCs that have been experimentally linked to their biosynthetic product/metabolite (72,83).

2.4.2. *Pseudomonas* specialized metabolite potential

As mentioned above, the genus *Pseudomonas* is one of the largest genera of Gram-negative bacteria, and it occupies widely different ecological niches. In line with this, *Pseudomonas* species are metabolically diverse and produce an extensive collection of specialized metabolites, such as siderophores, lipopeptides, terpenes, polyketides, non-ribosomal peptides and hybrids of these last two (57,73,84). Several plant-associated *Pseudomonas* also encode the auxin indole-3-acetic acid (57). Most specialized metabolites are produced only by specific *Pseudomonas* lineages, contributing to their specialization in environments and hosts (73).

Pseudomonas reference genomes encode 24 major BGCs classes (85), the most abundant being NRPS, RiPPs and redox-cofactors (57,85). *Pseudomonas* encode between 6 and 18 BGCs per genome, although one genome rarely has more than 15 BGCs (57,85). This diversity is also observed at the intra-species level: *P. fluorescens* strains have between 7 to 18 BGCs belonging to 20 BGCs classes (85,86). Model strain *P. syringae* pv. *tomato* DC3000 encodes 8 - 10 BGCs, while *P. syringae* pv. *syringae* B728a has 11 (85,87). These numbers could change as the tools for predicting BGCs improve or new methods are developed.

Examples of the specialized metabolites identified in *Pseudomonas*, in particular *P. syringae*, are the NRPS siderophore pyoverdine and toxins syringomycin and syringopeptide, PKS mupirocin and DAPG (2,4-Diacetylphloroglucinol), NRPS-PKS hybrids coronatine and syringolin (73). In addition, bioinformatics analyses have identified orphan BGCs, for which the product is not yet known (73,85).

3. How to study infections

Host infection by a pathogen is a complex process. The host and pathogen genotypes have a role, in addition to environmental variables including temperature, humidity and the presence of other microorganisms. Disease can only occur when the

proper combination of these three factors occurs, as recapitulated by the disease triangle (Figure 3). The most common disease outcome analyzed is symptom development by the host.

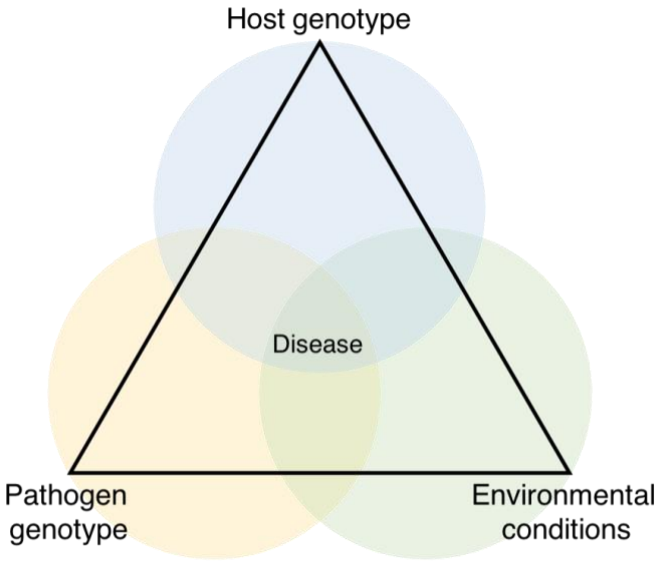

Figure 3 - The disease triangle.

When studying infections under laboratory conditions, as done in this thesis, the choice of infection method can have an impact on the infection process. Furthermore, different (molecular) phenotypes can be analyzed. Unfortunately, the complexity of the infection process results in high variability among replicates within and between plants upon infection (88).

3.1. Infections methods: syringe, flood, drip inoculation

One crucial step for studying infections in the lab is the infection method. The most common method used in phytopathology studies is syringe inoculation, whereby an inoculum of bacterial solution is injected into the leaf using a needleless syringe. This method provides reproducible results (88). It is important to note, however, that syringe infiltration brings the pathogen directly inside the leaf, to the apoplastic space. Thus, the first physical barrier that pathogens encounter in the wild, entering the apoplast, is absent when using this method.

In contrast, flood inoculation allows to recapitulate the penetration of pathogens into the leaf. It consists of the flooding of the compartment where the plant is growing with bacterial inoculum, which is later removed or discarded (89). Flood inoculation is performed on plate-grown plants, where wells can be easily filled with the bacterial solution. Similar to flood inoculation, dipping and spray inoculation cover the leaves with bacterial suspension, and mimic the natural entry of pathogens via wounds or natural openings (90). The addition of surfactant reagents such as Silwet is common in these infection methods. Surfactants decrease the surface tension on the leaf

surface, ensuring homogeneous covering of the leaf and thus resulting in reliable symptom development (90).

Drip inoculation consists of distributing a given volume of bacterial inoculum on the leaves' surface, usually in a drop-by-drop manner. All infections included in this thesis were performed using drip inoculation. Although the results with this method may be more variable than with more traditional syringe inoculation, it allows the inclusion of an essential step of bacteria pathogenicity: the ability to get inside the leaf and colonize the apoplast, without the use of surfactants that can modify this process.

3.2. Phenotypes: disease symptoms and bacterial growth

The most common phenotype assessed in phytopathogenicity studies is the development of visible symptoms in the host, particularly in the leaves. Symptoms include chlorosis, i.e. the appearance of yellow spots, and necrosis due to cell death. Other phenotypes include plant weight or size. In fitness studies, the number of fruits and/or seeds is used. Pathogen load or growth is often measured as well. For bacterial pathogens, this is usually done by colony counting, where infected host tissue is macerated in buffer to release bacterial cells from inside the leaves. The bacterial solution is then serially diluted and plated on agar to determine the number of colony forming units, a proxy for viable bacterial cells. Previous studies have found that the number of leaves displaying symptoms of infection and plant size correlate positively with bacterial growth (88,91). However, the amount of bacteria resulting in evident symptoms of infection varies among host genotypes, suggesting different thresholds for successful infection even within a species (92–95).

Traditional phenotypes are time-consuming to measure and hinder high-throughput studies. Fortunately, techniques that allow higher processivity have been developed in recent years. One such method is the determination of plant size from pictures taken before and after infection, since plant size or growth can be used as a proxy for host susceptibility to infection (91,96,97) and plant size has been shown to correlate well with plant fresh weight and bacterial load (91,96). Regarding bacterial pathogens, the addition of genetic barcodes has been used for tracking bacterial abundance during infection (98). This is especially useful when studying pair-wise bacterial interactions or communities, since each bacterial isolate can be tagged with a unique barcode and their abundance can be determined by sequencing. One approach that does not require sequencing is tagging bacteria with a luminescence operon. A complete luminescence operon can be transformed into the bacterial genome, so that no addition of substrate is necessary for the reaction (99,100). In this case, luminescence is used as a proxy for bacterial growth.

3.3. Genetics and genomics approaches: gene mapping and GWAS

Genetic mapping aims to identify loci in the host or pathogen that are relevant for infection, with a strong focus on genes conferring resistance to the host. Two common mapping strategies to identify candidate loci are quantitative trait locus (QTL) and

genome-wide association studies (GWAS). QTL takes advantage of the recombination blocks in the progeny of crosses between susceptible and resistant hosts. Thus, it is necessary to first identify genotypes with opposite phenotypes, to then cross them and obtain an F_2 population where the phenotype of interest, i.e. resistance to infection, segregates. The F_2 individuals are phenotyped and then genotyped to identify genetic regions associated with the phenotype of interest.

In *A. thaliana,* F_2 segregating populations have allowed the identification of genes underlying several phenotypes using QTL mapping (3). However, the success of genetic mapping is contingent upon factors such as the heritability of the phenotype of interest, the size and type of the mapping population, the coverage of the molecular genetic map and the statistical methods used (9). The genetic architecture of the trait of interest also has a role: qualitative or discrete phenotypes are usually mediated by a single locus with large effects, which are easier to identify than for quantitative or continuous phenotypes, where several loci with small effects are involved (9).

The outcome of infection depends on the interactions between both host and pathogen genotypes. QTL and GWAS are limited in that they focus only on one partner of this interaction. Computational methods for mapping that integrate both host and pathogen genetic information have been developed (101–103), although they are not widely used, in part because they require enough genetic diversity and knowledge of both host and pathogen, which is not always easy to achieve.

3.4. Functional approaches: transcriptomics and metabolomics

Beyond genetics and genomics, transcriptomics and metabolomics have gained traction in recent years for the study of plant-microbe interactions. Most transcriptomics and metabolomics studies focus on either the host or the pathogen, although dual RNA-seq, where host and pathogen are profiled simultaneously, is now possible yet not widely used due to the challenges it poses (104).

In particular for the *A. thaliana - P. syringae* pathosystem, transcriptional changes initiated by pathogen recognition or effector delivery have been studied using transcriptomics (29,105). Lewis et al. found that more than one-third of the *A. thaliana* transcriptome was differentially expressed after DC3000 infection, with most changes being observed in the early phase of infection (105). Even though the transcriptome response is similar in resistant and susceptible plants, remodeling occurs faster in the former (29). Transcriptomic changes in the pathogen have also been described, highlighting the type 3 secretion system, effectors, siderophores, response to oxidative stress, among others (106). Genetic variation in the pathogen and the host influence the pathogen transcriptome strategy upon infection and the host defense system (107,108), underscoring the importance of studying more than one host and/or pathogen genotype.

Metabolites play a key role in plant defense responses and in pathogen virulence, hence metabolomics is a great tool to study plant-pathogen interactions (109). *A. thaliana* accessions have distinct metabolic phenotypes, both qualitative and quantitative (110), which can influence the susceptibility of a particular accession to certain pathogen genotypes. It has been reported that infection with DC3000 results in a metabolomic transition that includes changes in the abundance of amino acids and nitrogenous compounds, but also of glucosinolates and indolic compounds involved in defense responses (111). Metabolomics has been used to study plant-pathogen interactions in other systems as well (112–114). Moreover, transcriptomics and proteomics/metabolomics data can be integrated to provide a better understanding of the regulatory networks and their intricated layers of connection, further strengthening findings and hypotheses (115,116).

4. The *Arabidopsis thaliana - Pseudomonas viridiflava* pathosystem

Already in the early 1990s, *A. thaliana* was proposed as a model host for studying plant-pathogen interactions due to its well-defined molecular biology and genetics (117). Before that, *A. thaliana* had been mainly used for developmental biology, cell biology and metabolism research. This was in part because no *A. thaliana* pathogen had been described in detail (117). However, after it was demonstrated that *P. syringae* pv. *tomato* DC3000, a bacterial pathogen isolated from tomato, was able to infect *A. thaliana* and cause disease symptoms (90,117), a new pathosystem was developed for the study of plant-pathogen interactions, where manipulation of both host and pathogen was possible. Although DC3000 infection of *A. thaliana* does not occur naturally, and artificial methods are required in the lab for pathogen inoculation (90), the study of the *A. thaliana* - *P. syringae* system has provided very valuable knowledge. In this thesis, I use instead a natural pathogen of *A. thaliana*: *Pseudomonas viridiflava*.

4.1. *Pseudomonas viridiflava*: an understudied opportunistic pathogen

Pseudomonas viridiflava is a globally-distributed natural pathogen of *A. thaliana* (88,92,118). It belongs to the *P. syringae* species complex, specifically to phylogroups 7 and 8 (119). Nevertheless, it is genetically distinct from the well-characterized strain DC3000. *P. viridiflava* was first described as a natural pathogen of *A. thaliana* in 2002 (88), and since then it has been isolated from populations in the Midwestern United States (92), Germany (118) and several other European locations (120), where it is one of the most, if not the most, common *Pseudomonas* growing endophytically. Interestingly, the plants from which *P. viridiflava* has been isolated range from lacking evident symptoms of disease to displaying chlorotic, necrotic and tan-colored lesions (88,118). Moreover, *P. viridiflava* has been shown to be pathogenic on *A. thaliana* under lab conditions on both soil and culture media (88,98,118).

4.1.1. *Pseudomonas viridiflava* virulence mechanisms

Multiple virulence factors have been described in *P. viridiflava*: pectate lyase enzyme, phase variation capability, ice nucleation and the presence of the effector

AvrE (119). Some of these factors, such as pectate lyases and phase variation, are absent in DC3000. AvrE is an effector that helps establish an aqueous apoplast by manipulating the abscisic acid signaling pathway, resulting in increased pathogen virulence (66,67,121). In *P. viridiflava* isolated from the Midwestern USA, AvrE absence correlates with the absence of pathogenicity (50). AvrE, together with three other effectors, is part of the soft core genome of the *P. syringae* complex, defined as being present in more than 95% of the genomes (55). Phase variation refers to the ability of some *P. viridiflava* isolates to produce two types of colonies which differ in virulence (50). Finally, pectate lyases degrade plant cell walls thereby releasing nutrients into the apoplast, and their activity correlates with virulence in isolates that have a single pathogenicity island (122). Furthermore, bacterial growth *in planta* has been shown to positively correlate with the pectate lyase activity for some *P. viridiflava* isolates (122). DC3000, on the other hand, is not able to degrade pectate, in agreement with the lack of putative pectate lyase genes in its genome (122).

P. viridiflava encodes a type 3 secretion system in a pathogenicity island (PAI). Two PAIs have been described in *P. viridiflava*, a tripartite PAI (T-PAI) that contains the three elements found in DC3000 (*hrp/hrc* cluster, EEL and CEL), as described in section 2.2; and a single PAI (S-PAI) encoding only the *hrp/hrc* cluster with the secretion machinery and effectors *avrE* and *avrF* (49,123). These PAIs are located in two different chromosome positions and are mutually exclusive, constituting a presence/absence polymorphism (49). The S-PAI was identified for the first time in *P. viridiflava* and, while present in most *P. viridiflava* isolates, it is not present in other *P. syringae sensu stricto* isolates (119). This suggests differences in the virulence mechanisms between *P. viridiflava* and more-studied *P. syringae* isolates such as DC3000. Indeed, *P. viridiflava* isolates belonging to PG7 have not been shown to produce toxins, while isolates in PG8 produce an antifungal toxin (119).

4.1.2. *Pseudomonas viridiflava* ATUE5

A multi-year survey of *Pseudomonas* in six *A. thaliana* populations from Southwest Germany found that a single operational taxonomic unit (OTU) was the most abundant throughout the study (118). This OTU was named OTU5 and was recently renamed to ATUE5, from Around Tuebingen 5 (96,98); all *P. viridiflava* isolates used in this thesis belong to this clade.

The pathogenicity of ATUE5 has been tested on axenic and soil systems by measuring the impact of infection on plant growth. Most ATUE5 isolates reduce plant growth under axenic conditions and on soil on the local *A. thaliana* accession Ey15-2 (97,118), and a synthetic community of seven ATUE5 isolates reduces the fresh weight of this and five additional local host genotypes grown on soil (98). Interestingly, while most ATUE5 isolates are pathogenic and reduce plant growth, the impact of individual isolates on the host, i.e. their virulence, varies (118). Karasov and colleagues (118) analyzed the genomes of more than a thousand ATUE5 isolates, and did not find the genes required for the biosynthesis of toxins coronatine, syringomycin, syringopeptide,

mangotoxin, phaseolotoxin nor tabtoxin. Moreover, the only conserved effector gene among isolates is *avrE* which, as mentioned above, has been associated with pathogenicity in *P. viridiflava* (50). Given that *avrE* is present in all ATUE5 isolates, its presence does not explain the differences in virulence observed among isolates.

4.2. Previous studies in this pathosystem

Despite *P. viridiflava* being the most abundant *Pseudomonas* species in *A. thaliana* populations, little is known about the mechanisms of bacterial virulence and plant resistance in this pathosystem. As mentioned above, most of the plant-pathogen studies carried out so far in *A. thaliana* exploit pathogens that are not natural to this plant, DC3000 in particular.

Only a handful of scientific papers have looked into the *A. thaliana* - *P. viridiflava* pathosystem (88,92,93,122,124–126). Overall, they report genetic variation on both bacterial virulence and plant resistance. Recently, our lab has expanded the knowledge of this pathosystem by characterizing the distribution of *P. viridiflava* in European *A. thaliana* population and their virulence in axenic- and soil-grown plants (91,98,118). Although many *A. thaliana* accessions have an intermediate phenotype, both highly susceptible and resistant hosts have been identified, as well as very virulent bacterial isolates (88,93). There is no evidence of local adaptation in this pathosystem (93), supporting the previous finding that *P. viridiflava* is a generalist pathogen (119).

The interactions between *A. thaliana* and *P. viridiflava* are quantitative rather than qualitative, with the outcome of infection dependent on fitness trade-offs and temporal and spatial variations in the environment (93). Indeed, development of disease symptoms in leaves and pathogen growth *in planta* are influenced not only by host and bacteria genotype but by their interaction, in addition to environmental variables (88,92,93,98). The contribution of each of these factors varies between studies, reflecting the complex nature of their interactions. Thus, there is a range of continuous phenotypes of infection, characteristic of quantitative resistance (127). Infection of *A. thaliana* with *P. viridiflava* not only leads to diseased leaves, it can also reduce plant fitness in terms of seed production (92). Interestingly, infection can have both positive and negative effects on fitness on specific plant genotypes, again underlying the diversity of response to infection within a single host species.

Defense against *P. viridiflava* isolated from the Midwestern USA is mediated mainly by the JA signaling pathway, with a smaller contribution from the SA pathway (122). This contrasts with the defense response against DC3000, where SA is the main player. The host immune response is independent of the pathogenicity island, as both T-PAI and S-PAI *P. viridiflava* isolates induce the SA and JA defense pathways to a similar extent (122). Furthermore, there is no difference in the virulence of T-PAI and S-PAI isolates (50). In addition to JA, the polyamines spermine and thermospermine modulate *A. thaliana* resistance to *P. viridiflava* (124,125).

The virulence of *P. viridiflava* isolates varies on the same host (118), and the response of *A. thaliana* to infection, including symptom severity and seed production, varies among accessions infected with the same bacterial strain (92,93). Nevertheless, it is not yet clear how this variation arises, nor if the same variation is observed for *P. viridiflava* from different geographic locations.

5. Thesis overview

In this thesis, I characterized the natural *Arabidopsis thaliana* - *Pseudomonas viridiflava* pathosystem following an integrative approach. I combined genetics, transcriptomics and metabolomics to identify mechanisms underlying resistance to infection, and used a computational framework to describe the diversity of specialized metabolites in the pathogen, which can have a role in virulence.

In the first chapter, my goal was to investigate how *P. viridiflava* interacts with *A. thaliana*, and to compare these interactions with the model pathogen *P. syringae* pv. *tomato* DC3000. For this purpose, I performed a large-scale screen using a collection of host and pathogen genotypes; based on the plant phenotype and the bacterial growth upon infection, I identified susceptible and resistant hosts. Using these host genotypes, I delved into a functional characterization of the infection process using transcriptomics and metabolomics. I found that the ET branch of the JA/ET defense pathway is involved in defense against *P. viridiflava*, and that susceptible and resistant hosts have a similar transcriptomic response to infection that occurs faster in the resistant host.

In the second chapter, my goal was to identify gene clusters involved in the synthesis of specialized metabolites encoded by *P. viridiflava* that might be associated with differences in their virulence on *A. thaliana*. I first described the large biosynthetic potential of a collection of *Pseudomonas* genomes from the *A. thaliana* phyllosphere, which was dominated by NRPS metabolites. Then, I identified gene cluster families with a putative role in virulence, one of them related to the siderophore pyoverdine.

Finally, I present an overall summary of the main finding of this thesis and reflect on their contribution to the field of plant-pathogen interactions. Based on my results, I propose future studies to advance our understanding of the mechanisms underlying the interaction between hosts and their pathogens.

Chapter 1: *Arabidopsis thaliana* resistance to *Pseudomonas viridiflava* is recessive and due to an increased defense response

ABSTRACT

During evolution of host-pathogen interactions, pathogens improve their virulence tools while hosts advance their defenses. *Pseudomonas viridiflava*, an opportunistic pathogen, is the most common *Pseudomonas* clade in the phyllosphere of European *Arabidopsis thaliana* populations. Belonging to the *P. syringae* complex, it is genetically and phenotypically distinct from well-characterized *P. syringae sensu stricto*. Despite its broad range we lack knowledge of how *A. thaliana* responds to diverse co-occurring *P. viridiflava* strains and how they colonize *A. thaliana.* Here, we characterized the host response in a *A. thaliana* - *P. viridiflava* pathosystem. We measured host and pathogen growth in infections to ascertain gene x gene interactions and used immune mutants, transcriptomics and metabolomics to determine defense pathways influencing infection. We found a large effect of host genotype on infection outcome and evidence of host x pathogen genotype interactions. We found that *A. thaliana* uses jasmonic acid (JA)/ethylene (ET) signaling to defend itself against *P. viridiflava*, more so than salicylic acid. Our results suggest JA/ET is important for suppression of *P. viridiflava*, yet suppression capacity varies between accessions due to still unknown mechanisms. Our results shed light on how *A. thaliana* suppress the ever-present *P. viridiflava*, but further studies are needed to understand how this interaction contributes to persistence of *P. viridiflava* in *A. thaliana* populations.

AUTHOR CONTRIBUTIONS

Author are listed in the order to be used for publication.

1. Alejandra Duque-Jaramillo, Max Planck Institute for Biology Tübingen, Germany.
   Devised the study, performed or supervised all experiments, performed all data analyses, wrote the manuscript.
2. Nina Ulmer, Max Planck Institute for Biology Tübingen, Germany.
   Performed the infection of immune mutants.
3. Saleh Alseekh, Max Planck Institute of Molecular Plant Physiology, Germany.
   Measured primary metabolites and phytohormones.
4. Ilja Bezrukov, Max Planck Institute for Biology Tübingen, Germany.
   Developed the image analysis scripts
5. Alisdair R. Fernie, Max Planck Institute of Molecular Plant Physiology, Germany.
   Measured primary metabolites and phytohormones.
6. Aleksandra Skirycz, Max Planck Institute of Molecular Plant Physiology, Germany; Boyce Thompson Institute, Cornell University, USA.
   Measured secondary metabolites
7. Talia L. Karasov, Max Planck Institute for Biology Tübingen, Germany; School of Biological Sciences, University of Utah, USA.
   Devised the study, provided input for manuscript. Co-corresponding author.
8. Detlef Weigel, Max Planck Institute for Biology Tübingen, Germany.
   Devised the study, provided input for manuscript. Co-corresponding author.

STATUS IN PUBLICATION PROCESS

Advanced manuscript, awaiting for submission to target journal.

INTRODUCTION

Pseudomonas are ubiquitous bacteria of the phyllosphere, the above-ground parts of plants whose main component is the leaves (1,2). There, *Pseudomonas* can have a wide range of roles, from serious pathogens to biocontrol agents (3,4). *Pseudomonas viridiflav*a is a widespread opportunistic pathogen of plants (5–7), and it is found in a variety of environments, such as leaf litter, rain, and snow (8). Its effects range from no disease symptoms to growth impairment and obvious disease in soil-grown plants (9,10), while in axenic conditions it can be deadly to plants (11).

P. viridiflava is a natural pathogen of the model plant *Arabidopsis thaliana*, and it has been isolated from populations in the USA, Japan and across Europe (6,7,9,12,13). In a recent survey of *Pseudomonas* in natural *A. thaliana* populations in Southwest Germany, we found *P. viridiflava* to be the most prevalent *Pseudomonas* clade (11). It is an opportunistic pathogen under lab conditions, with the same isolate having different effects on plants grown on soil versus axenically on agar (6,10,11). Due to its prevalence and potential to cause disease, *P. viridiflava* likely constitutes a major selective pressure for *A. thaliana* populations (12).

Despite its prevalence in *A. thaliana* and other plant populations (6,7,9,11,13), *P. viridiflava* remains a poorly characterized microbial species. Most of our knowledge regarding *A. thaliana* and *Pseudomonas* interactions comes from *P. syringae sensu stricto* isolates, particularly from the model pathogen *P. syringae* pv. *tomato* DC3000 (henceforth DC3000). This strain has been widely used as a model pathogen due to its ability to infect *A. thaliana*, despite not being a natural pathogen of this species (14). DC3000 uses effector proteins and toxins to manipulate the host defense responses and promote bacterial growth. Effector proteins are translocated into the host cells via the type III secretion system (15), where they act to suppress host immunity and/or promote disease (15,16). However, effectors can be recognized by the plant immune system, eliciting an effector-triggered immune response (16). The recognition of DC3000's effectors by the host plant activates the salicylic acid (SA) defense pathway to achieve resistance (17). The SA signaling pathway is antagonistic to the jasmonic acid (JA) signaling pathway (17,18), a property exploited by DC3000. It produces the toxin coronatine that activates the JA defense pathway, consequently decreasing resistance through downregulation of the SA pathway and promoting bacterial growth (14,19).

Both DC3000 and *P. viridiflava* belong to the *P. syringae* complex, which includes many agriculturally relevant pathogens, yet they are genetically different (20,21). In particular, they possess different virulence factors (22) and thus, the aforementioned mechanisms characterized for DC3000 and related *P. syringae sensu stricto* isolates are not necessarily generalizable to *P. viridiflava.* Indeed, one study suggests that the defense response of *A. thaliana* to *P. viridiflava* is mediated mainly by the JA defense signaling pathway, with a minor contribution of the SA pathway (23). This is in contrast with the defense response against DC3000 relying heavily on SA. Moreover, we have

observed that DC3000 is a more serious pathogen than *P. viridiflava* isolates from Southwest Germany on soil-grown plants while, in axenic conditions, many *P. viridiflava* isolates are deadlier than DC3000 (6,11). Regarding virulence factors, most *P. viridiflava* isolates have a reduced effector repertoire, encoding only *avrE* (11,24). AvrE contributes to the establishment of an aqueous apoplast by manipulating abscisic acid (ABA) signaling, thus resulting in increased pathogen growth (25–27). In addition, *P. viridiflava* isolated in Southwest Germany lack known genes for synthesis of the toxins coronatine, syringomycin, syringopeptin, mangotoxin, phaseolotoxin and tabtoxin (11). These differences in the virulence factor repertoire between *P. viridiflava* and DC3000 suggest that different defense mechanisms are elicited in the host plant upon infection.

While *P. viridiflava* is likely to be among the most, if not the most, abundant bacterial pathogens of *A. thaliana* globally, little is known about the mechanisms of virulence in *P. viridiflava* and the corresponding mechanisms of resistance in *A. thaliana*. It has been proposed that the interaction between *A. thaliana* and *P. viridiflava* isolates from the Midwestern USA is not mediated by gene-for-gene recognition due to the absence of a clear hypersensitive response and the role of JA in resistance (23,28). The few studies investigating these interactions have focused on these bacteria isolates from Midwest USA (9,12,23,28), but due to the great intraspecific variation within *P. viridiflava* (7,8,11), we do not yet know whether the virulence mechanisms identified in USA isolates are conserved for other *P. viridiflava*.

Here, we investigated how *P. viridiflava* interacts with *A. thaliana*, and compared the plant response to *P. viridiflava* with a widely-studied model pathogen. We used *P. viridiflava* isolated from Southwest Germany in axenic conditions to i) determine the presence of genotype-by-genotype interactions in this pathosystem, ii) compare the host immune response with what has been described for USA *P. viridiflava* isolates and iii) compare *A. thaliana* response to *P. viridiflava* and the model *P. syringae* pathogen DC3000. Through infection trials with a genetically-diverse set of plants and pathogens to assess the ability of *P. viridiflava* isolates to infect *A. thaliana* genotypes, we identified susceptible and resistant *A. thaliana* genotypes to *P. viridiflava* infection, and characterized the plant transcriptome and metabolome response upon infection. We found that *P. viridiflava* infection upregulates the JA/ethylene (ET) defense signaling pathway, extending the results obtained with USA isolates to the most prevalent *A. thaliana* pathogen in Europe. Moreover, we found that *P. viridiflava* and DC3000 activate different branches of the JA/ET pathway, underscoring their different pathogenicity mechanisms. Finally, we posit that the difference in resistance between two closely-related host genotypes is due to a primed defense status and an earlier establishment of a defense response in the resistance host.

RESULTS

Host genotype is the best predictor for infection outcome, and there is a significant host x pathogen interaction

We previously reported that a single diverse *Pseudomonas* clade, OTU5, was prevalent in natural *Arabidopsis thaliana* populations in Southwest Germany (11). This clade, recently renamed ATUE5 (around Tuebingen 5; (13), was classified as *P. viridiflava* (11). We found that ATUE5 isolates differed in pathogenicity on a local *A. thaliana* accession (11). The extensive genetic and phenotypic response variation of the *P. viridiflava* ATUE5 clade could suggest that virulence differs in different host backgrounds, namely *A. thaliana* genotypes. If that is the case, isolates that are highly virulent on one accession might be less virulent on others, and vice versa.

To test whether there are such genotype-by-genotype interactions, we characterized how different *A. thaliana* genotypes interact with closely-related *P. viridiflava* ATUE5 isolates. We selected 21 *A. thaliana* genotypes representing the species global diversity (Table S1, Figure S1), and drip-inoculated them with 12 luminescence-tagged (29,30) *P. viridiflava* ATUE5 isolates. We used an axenic system and measured bacterial load and plant growth three days post-infection (dpi; Figure 1). To facilitate high-throughput screening of bacterial load, we integrated the *luxCDABE* into the *P. viridiflava* isolates and measured luminescence as a proxy of bacterial load (31,32). We confirmed that luminescence signal positively correlated with colony counts (Spearman's correlation coefficient 0.71, p-value < 2.2×10^{-16}), and thus was a valid proxy of bacterial load (Figure S2A).

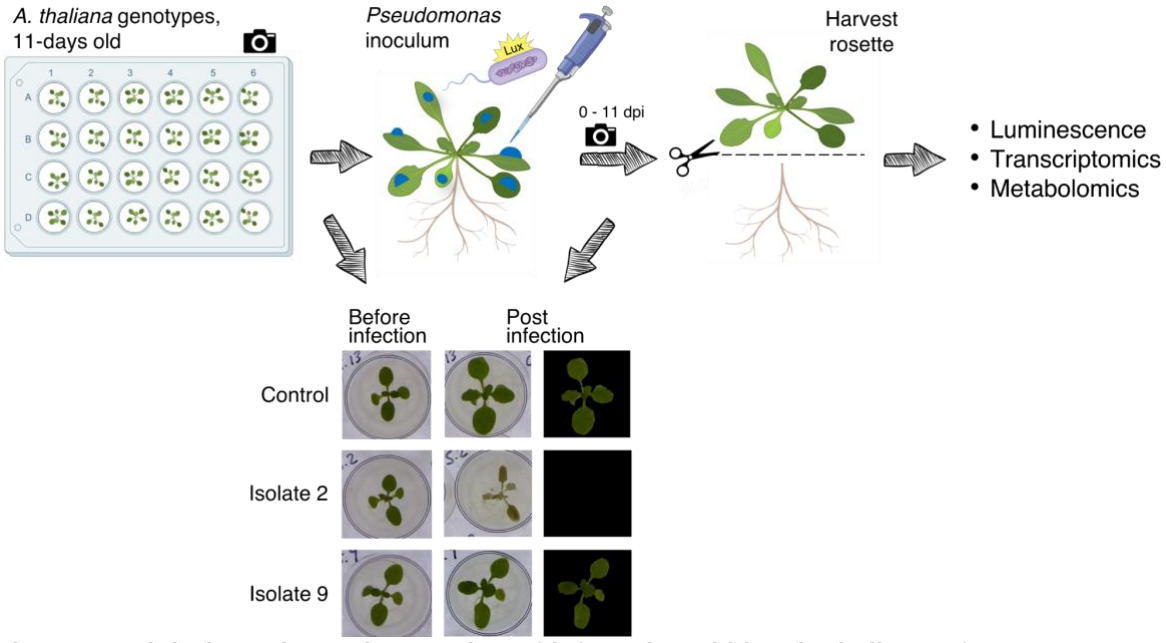

Figure 1 - High-throughput phenotyping of infected *Arabidopsis thaliana*. Plants were grown on 24-well plates. At 11 days-old, they were drip-inoculated with luminescence-labeled *Pseudomonas* inoculum at OD_{600} = 0.01. Then, at selected timepoints, the aerial part of the plant, i.e. the rosette, was harvested to measure luminescence as a proxy of bacterial growth, for RNA extraction and sequencing or for quantification of primary and secondary metabolites and phytohormones. Pictures of the plates

were taken before infection and before harvesting the plants. The number of green pixels was obtained from the pictures and used as a proxy for non-diseased tissue.

Hierarchical clustering of luminescence measurements divided host genotypes into two groups based on bacterial load: either susceptible or resistant to all *P. viridiflava* (Figure 2A). Hosts classified as susceptible were Sha, Col-0, Yeg-1, Monte-1, Koch-0, Toufl-1, Qui-0, HKT2.4, Rovero-1, Star-8, Mammo-1 and Jablo-1; while those classified as resistant were Aitba-2, Sij-4, Apost-1, Shigu-1, Ciste-2, Fei-0, TuWa1-2, Ey15-2 and Slavi-1. There were significant differences in the bacterial load between host genotypes (ANOVA p-value < $2x10^{-16}$). We identified Sha as the most susceptible *A. thaliana* genotype, with a mean $\log_{10}$ luminescence (3.600, SD = 0.541) significantly different from that of all other host genotypes (Tukey HSD p-value < 0.05 for all comparisons; Figure S3A). Aitba-2 and Sij-4 were the most resistant host genotypes (mean $\log_{10}$ luminescence across all bacterial genotypes = 2.312 and 2.464, SD = 0.569 and 0.657 respectively). Despite having opposite susceptibility to *P. viridiflava* ATUE5 isolates, Sij-4 and Sha were genetically the most closely related host genotypes included in this study (pairwise genetic distance = 0.018, Figure S1), whereas Aitba-2 was the most distant to all other host genotypes (range of pairwise genetic distance to other genotypes = 0.047 - 0.053, Figure S1), in line with its classification as a relict by the 1001G project (33). These results suggest that host genotype is an important determinant of the outcome of infection, and genetically-close hosts can have different susceptibility to ATUE5.

There were statistically significant differences in the bacterial load each pathogen genotype reached across all host genotypes after three days (ANOVA, p-value < $2x10^{-16}$; Figure S3B). We identified p13.G4 as the most virulent pathogen (mean $\log_{10}$ luminescence = 3.094, SD = 0.660) and isolate p13.C1 as the least virulent one (mean $\log_{10}$ luminescence = 2.646, SD = 0.531; Figure S3B). Although the distribution of luminescence by pathogen genotype was narrower than that from host genotype, the bacterial load was significantly different between isolates p13.G4 and p13.C1 (Tukey HSD p-value < 0.05). The phylogenetic relationships between ATUE5 isolates were not recapitulated by the hierarchical clustering of luminescence signal, with the exception of isolates p13.C1 and p1.D2 (Figure 2, Figure S4). Furthermore, we found that closely-related isolates reached different bacterial loads in the same host (Figure 2A), again supporting the existence of genotype-by-genotype interactions in this pathosystem.

In addition to pathogen load, host-pathogen interactions can be assessed by host growth and/or the development of symptoms in leaves, i.e. chlorosis and necrosis. To estimate plant growth as a measure of susceptibility to infection, we took pictures of the plants before and after infection, and quantified the number of green pixels as a proxy for plant size (Figure 1). There was a significant negative correlation between

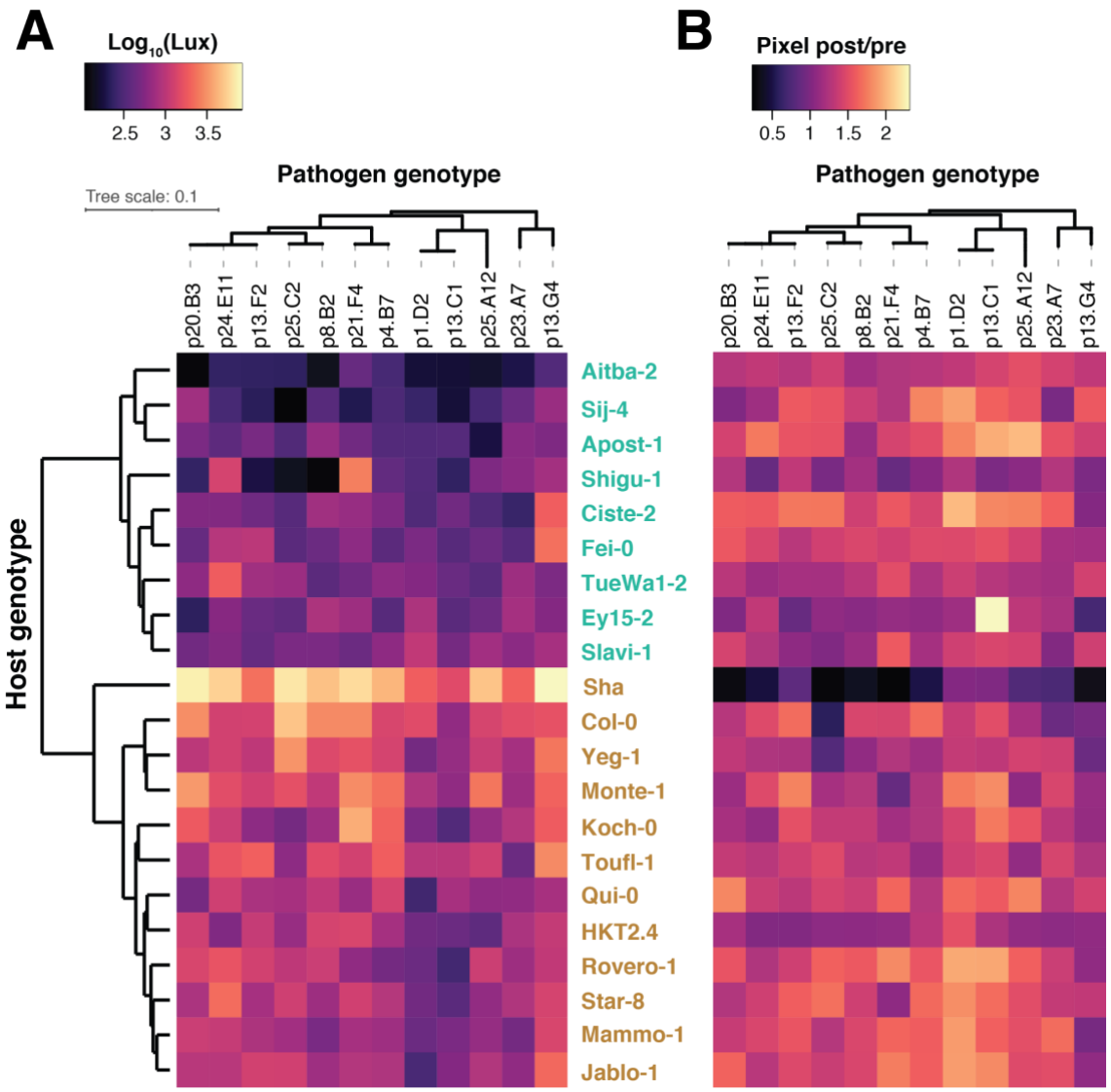

Figure 2 - *A. thaliana* genotypes differ in the *P. viridiflava* load they support. (A) Pathogen load measured as luminescence 3 days post-infection in all host x pathogen genotype combinations. In total, 21 host genotypes were drip-inoculated with 12 luminescence-labeled pathogen genotypes. Bacterial isolates are ordered based on their phylogenetic relationship, host genotypes are ordered by hierarchical clustering based on the luminescence data displayed in the heatmap. **(B)** Ratio of green pixels post to before infection as a proxy for plant growth in all host x pathogen genotype combinations. Host and pathogen genotypes are ordered as in A. Phenotype based on bacteria load is indicated by the color of the host genotype: susceptible (ocre) and resistant (aquamarine). n = 8 - 15 for each host x pathogen genotype combination. We repeated the experiment with a subset of seven host genotypes and seven pathogen genotypes with similar results.

the delta of green pixels at 3 dpi (3 dpi - before infection) and the luminescence signal (Spearman's correlation coefficient -0.484, p-value < $2.2x10^{-16}$; Figure S2B) even though visually, the clustering of host genotypes was weaker with this phenotype compared to bacterial load (Figure 2B). This was also the case for the ratio of green pixels 3 dpi to before infection (Spearman's correlation coefficient -0.457, p-value < $2.2x10^{-16}$).

Our results supported the existence of genotype-by-genotype interactions between *A. thaliana* and *P. viridiflava* ATUE5. To determine the individual contribution of host and pathogen genotype to bacterial load and to quantify genotype-by-genotype interactions, we performed a variance decomposition analysis (34,35). We found that host genotype explained 16.8% of the variance in luminescence signal, while pathogen genotype explained 3.4%. There was a significant interaction between host and pathogen genotypes (p-value = 0.0076), which explained 7% of the luminescence signal, more than pathogen genotype alone. Figure S2C shows an example of genotype-by-genotype interactions: on the one hand, isolate 12 grew to high abundance on host genotype Yeg-1 while isolate 4 reached a lower load on the same host. On the other hand, the abundance of isolate 4 was higher than isolate 12 in host genotype Ey15-2. These results confirm that there are genotype-by-genotype interactions in this pathosystem, which affect the outcome of infection. While host genotype, rather than pathogen genotype, is the best predictor of the load a determined pathogen reaches in a certain host, this is also affected by the interaction between host and pathogen genotype, such that both are important for the outcome of infection.

Bacterial growth is slower in the resistant host

We selected susceptible Sha and resistant Sij-4 host genotypes for further investigation of the *A. thaliana* response against *P. viridiflava* infection. How infection progresses in this pathosystem has not been described yet. To address this gap in knowledge, we followed infection with a two-fold aim: first, to characterize its progression over time in terms of bacterial load and host symptom development; and second, to evaluate if Sij-4 resistant phenotype observed at 3 dpi persisted over time. We drip-inoculated Sha and Sij-4 host genotypes with *P. viridiflava* ATUE5::p13.G4, the most virulent pathogen from our axenic screen (Figure S3B), and followed infection from 0 to 14 dpi by both luminescence and green pixels quantification.

We found bacterial load to be highest in Sha (susceptible genotype) at 3 dpi, while it took 6 days for bacteria in Sij-4 (resistant genotype) to reach its maximum (Figure 3A). The maximum bacterial load was comparable between Sha and Sij-4, yet Sij-4 infected-plants did not decrease in size despite this bacterial load being associated with tissue collapse in Sha. We observed that infected Sij-4 plants had a lower growth rate compared to mock-infected plants (Figure 3B). These results suggest that resistance of Sij-4 is, at least partly, due to slower pathogen growth compared to susceptible Sha.

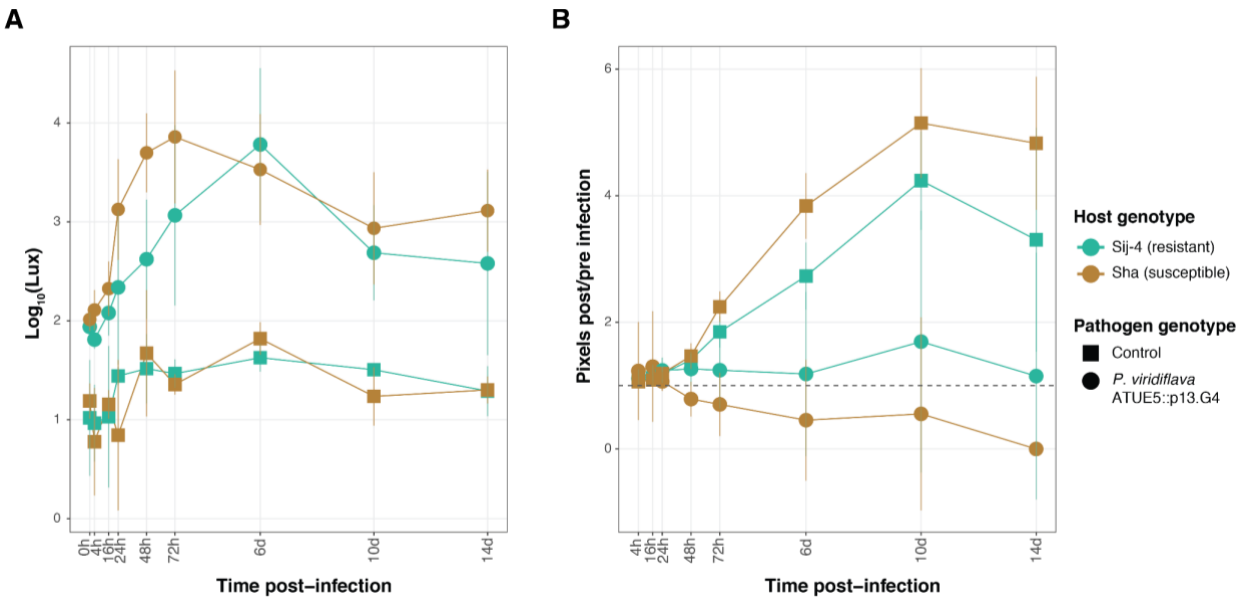

Figure 3 – Bacterial growth is slower in the resistant host and does not result in plant death. Mean luminescence **(A)** and mean ratio of green pixels 3 days post infection to before infection **(B)** of susceptible Sha and resistant Sij-4 host genotypes infected with *P. viridiflava* ATUE5::p13.G4. Plants were drip-inoculated and phenotyped 3 days post infection. The dashed line at 1 indicates the ratio of green pixels with neither gain nor loss of pixels. Error bars represent standard deviation. Ocre: susceptible Sha, aquamarine: resistant Sij-4. Square: mock-infected plants, circle: plants infected with *P. viridiflava* ATUE5::p13.G4. n = 3 - 6 replicates per host x pathogen x time point for A, n = 4 - 24 per host x pathogen x time point for B.

Resistance is a recessive trait

Having confirmed the resistant phenotype of Sij-4, we set to determine the genetic basis of this trait. We performed Mendelian segregation analysis using an F_2 population derived from a cross between resistant Sij-4 and susceptible Sha generated previously (36). We infected 427 F_2 individuals and both parental lines in two independent experiments, and measured plant size before and 3 dpi.

We first evaluated if the size of host genotypes Sha and Sij-4 differed. A genetically-determined difference in size of the parental lines could affect the outcome of infection in the F_2 plants, independently of resistance-associated genes. There was no statistically significant difference in the size of Sha and Sij-4 plants before infection when controlling for variation between experiments (p-value = 0.181). Similarly, there was no significant difference in their growth rate (p-value = 0.316). We also observed that the size distribution of the F_2 plants before infection was homogeneous (Figure S5). These results indicate that size-related genes do not confound the outcome of infection in this F_2 population.

We observed a wide distribution of the ratio of green pixels 3 dpi to before infection for the F_2 plants, indicating resistance to *P. viridiflava* is a quantitative trait. We estimated the segregation ratio of resistance as a trait. For this, we classified F_2 plants as susceptible or resistant based on the green pixels' ratio of mock-infected F_2 plants

(Figure S5). The smallest ratio was taken as a threshold (experiment 1 = 1.57, experiment 2 = 1.12), and infected F_2 plants with a green pixels ratio equal to or higher than the threshold were considered resistant. Resistance was observed less frequently than susceptibility: 31% of F_2 infected plants in experiment 1 (95% confidence interval [0.257, 0.372]), and 40%, in experiment 2 (95% confidence interval [0.331 - 0.473]) were classified as resistant to *P. viridiflava* ATUE5::p13.G4 infection (Table 1). Resistance was the recessive trait in our experiments, with a 1:1.5-2 resistant to susceptible segregation that is significantly different from the 3:1 ratio expected from a single locus (χ^2 goodness-of-fit test, p-value < 0.05). It could, on the contrary, be explained by a two locus model, in which being recessive at one and only one of the loci results in resistance to *P. viridiflava* ATUE5::p13.G4 infection. Thus, our results are compatible with a simple genetic architecture where few genes with strong effects underlie resistance to *P. viridiflava*.

Table 1. Segregation of resistant phenotype in Sha x Sij-4 F_2 individuals infected with *P. viridiflava*

Experiment (threshold)	Host genotype	n resistant (n total)	Proportion resistant	χ^2 test p-value (3:1 segregation)
1 (1.57)	F2	77 (247)	0.31	0.025
	Sij-4	4 (4)	1.00	
	Sha	0 (7)	0.00	
2 (1.12)	F2	72 (180)	0.40	3.359×10^{-6}
	Sij-4	12 (14)	0.86	
	Sha	2 (10)	0.20	

P. viridiflava infection results in a large transcriptome remodeling, which occurs earlier in the resistant host

To test whether *A. thaliana* defense mechanisms against *P. viridiflava* are conserved between USA and European isolates, we performed transcriptomic analysis of susceptible Sha and resistant Sij-4 infected with *P. viridiflava* ATUE5::p13.G4. We included DC3000 for comparison against ATUE5 (Figure S6). Transcriptome remodeling as early as 4 hours post-infection (hpi) has been reported in *A. thaliana* infected with avirulent DC3000 expressing AvrRpt2 or AvrRpm1, and from 9 hpi for virulent DC3000 (37). These results come from soil-grown plants that were syringe-infiltrated, while our experiments are on axenic plants grown on agar and drip-inoculated. To ensure we could capture transcriptome remodeling in our system, we collected samples in a time series from 0 to 72 hpi, from infection to plant death in susceptible Sha.

We performed principal component analysis and estimated the number of differentially expressed genes in each combination of host x pathogen x time point, compared to its

corresponding mock-infected plants (Figure 4). Principal component 1 (PC1, 35.9% of variance explained) recapitulated the time point when plants were harvested and it separated samples infected with *P. viridiflava* from those infected with DC3000 and control, while PC2 (14.6%) separated samples based on the day/night cycle stage: time points 16 and 42 hpi were at night, while time points 0, 4, 24 and 72 hpi were during the day (Figure S7A). Finally, PC3 (13.4%) separated samples based on host genotype (Figure 4A, Figure S7B). These results suggest that time point and host genotype have a strong effect on gene expression both at the moment of infection and during its course.

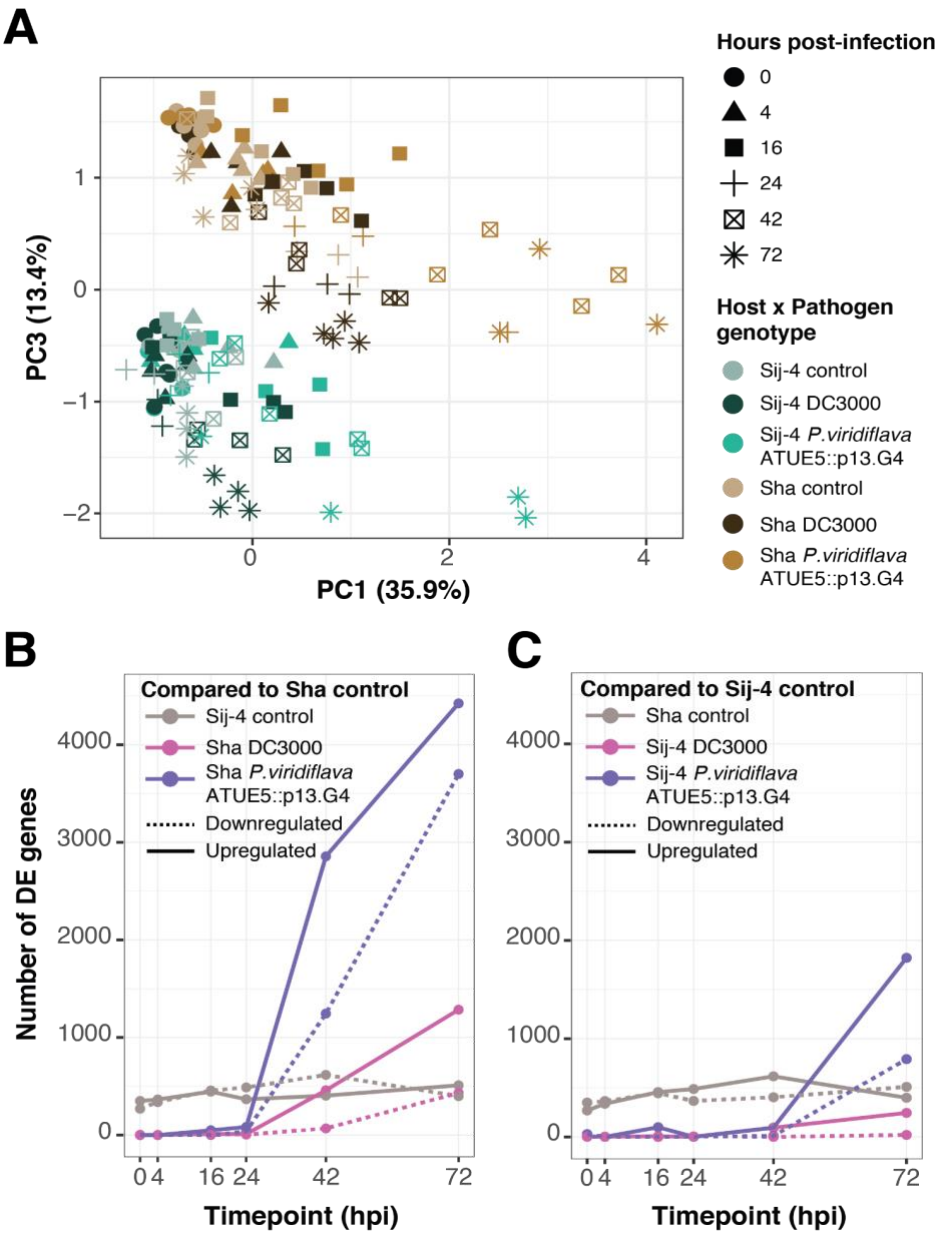

Figure 4 - Transcriptome remodeling upon *P. viridiflava* infection is larger in susceptible Sha. (A) Principal component analysis of variance-stabilized counts for all transcripts in Sha and Sij-4 plants infected with *P. viridiflava*, DC3000 or mock-infected (control). Plants were harvested at the indicated time point between 0 and 72 hours post-infection. Symbols represent time post-infection at collection and colors represent the host x pathogen genotype combination. **(B)** Number

of differentially expressed genes over time in infected Sha plants compared to mock-infected plants at the same time point. **(C)** Number of differentially expressed genes over time in infected Sij-4 plants compared to mock-infected plants at the same time point. Gray: mock-infected, pink: DC3000, purple: ATUE5 p13.G4. n = 3 - 5 replicates per host x genotype x time point combination, except for Sij-4 x DC3000 x 24 hpi where n = 2.

The principal component analysis indicated susceptible Sha underwent a large transcriptome remodeling upon *P. viridiflava*, more evident at the later time points. Indeed, we observed the largest number of differentially expressed genes (DEGs) in Sha plants infected with *P. viridiflava*: more than 4000 and 8000 genes were up or downregulated at 42 and 72 hpi, respectively (Figure 4B). This was approximately 8 and 5 times more DEGs than in DC3000-infected Sha plants. In *P. viridiflava*-infected Sij-4 plants, the largest number of DEGs was also observed at 72 hpi: 2614 DEGs. Most DEGs were upregulated in infected plants compared to mock-infected controls in all host and pathogen genotypes. There were more DEGs in Sij-4 than in Sha infected plants at early time points, particularly downregulated genes at 0 hpi and upregulated genes at 16 hpi. From 24 hpi onwards, the number of differentially expressed genes was consistently higher in infected Sha compared to Sij-4.

An earlier transcriptional response in Sij-4 upon infection compared to Sha could underlie the resistant phenotype of the former. To evaluate this hypothesis, we performed GO enrichment analysis of the differentially expressed genes between infected and mock-infected Sij-4 plants at 0 (35 DEGs) and 16 hpi (99 DEGs). At 4 hpi there were only 3 DEGs, so we excluded this time point from the analysis. While no significant GO term was enriched at 0 hpi, upregulated genes in infected Sij-4 plants at 16 hpi were enriched in GO terms related to jasmonic acid (JA biosynthetic process, response to fatty acid, JA and wounding) and immune and defense responses (Table S3). In contrast, upregulated genes in infected Sha plants at 16 hpi (49 DEGs) had no enrichment of JA-related GO biological processes (Table S3). Instead, top enriched GO categories included detoxification/response to toxic substances and response to hypoxia. There were no enriched GO categories related to detoxification/response to toxic substances in the top 20 for Sij-4. GO categories enriched in both Sij-4 and Sha after infection compared to control were related to response to biotic stimulus and fungus. Taken together, these results indicate distinct transcriptional responses between resistant and susceptible hosts to *P. viridiflava* infection, with the resistant host establishing a JA-mediated immune response already at 16 hpi.

Susceptible and resistant host genotypes differ in defense potential before infection

So far, we observed enrichment of biological processes related to JA in the resistant host Sij-4 at 16 hpi. Since basal differences between host genotypes could contribute to differences in susceptibility, we identified genes with differential expression between mock-infected Sha and Sij-4, and performed a GO enrichment analysis.

We found that between 620 and 1021 are differentially expressed between mock-infected Sha and Sij-4 plants (Figure 4B,C). A total of 252 genes were differentially expressed across all five timepoints (Figure S8). To test if these basal differences in transcriptome were related to the defense response, we focused on time point 0 hpi, before any strong transcriptome response to infection would take place. At 0 hpi there were 270 genes with higher expression in Sha and 350 with higher expression in Sij-4. The top enriched GO terms among genes with higher expression in Sha were related to iron transport, homeostasis and starvation (Table S4). The only GO term significantly enriched in Sij-4 was defense response. These results suggest that the resistant genotype Sij-4 has a primed defense response that could contribute to its resistant phenotype.

P. viridiflava activates the ET branch of the JA/ET defense pathway

Given that our analysis of enriched GO biological processes in the resistant host pointed to the involvement of JA in the establishment of resistance against *P. viridiflava*, we focused on the JA/ET signaling pathway. JA is known to mediate defense against necrotrophic pathogens and upon herbivory (17,38), and a previous study found that it is important for defense against *P. viridiflava* (23).

We assess the expression of the JA/ET pathway in both susceptible Sha and resistant Sij-4 upon infection with *P. viridiflava* ATUE5::p13.G4 and with DC3000. We selected *COI1, MYC2* and*VSP2* as marker genes for the JA pathway, *EIN2* for the ET pathway, and *ERF1 and PDF1.2* as genes receiving input from both pathways (Figure 5A). We observed that infection with *P. viridiflava* ATUE5 upregulated different marker genes of the JA/ET defense pathway compared to infection with DC3000. Infection with *P. viridiflava* led to higher levels of *ERF1* and *PDF1.2*; while DC3000 infection increased *MYC2* and *VSP2* (Figure 5B). This differential upregulation of the MYC2 and the ET branches of the JA/ET defense pathway was observed in both the susceptible and the resistant host. In general, transcript levels were higher in susceptible Sha compared to resistant Sij-4. Infection with *P. viridiflava* ATUE5 resulted in more transcripts of *COI1* in susceptible Sha, whereas the effect of infection on *EIN2* was less prominent. Our results suggest that upregulation of the ET branch of the JA/ET defense pathway by *P. viridiflava* ATUE5 is achieved via JA and *COI1,* not via ET and *EIN2* signaling. Taken together, these results show that DC3000 and *P. viridiflava* activate different branches of the JA/ET defense pathway in *A. thaliana*, and underscore the role of the JA/ET pathway in defense against *P. viridiflava*, as previously reported by Jakob *et al.* using Midwest USA isolates (23).

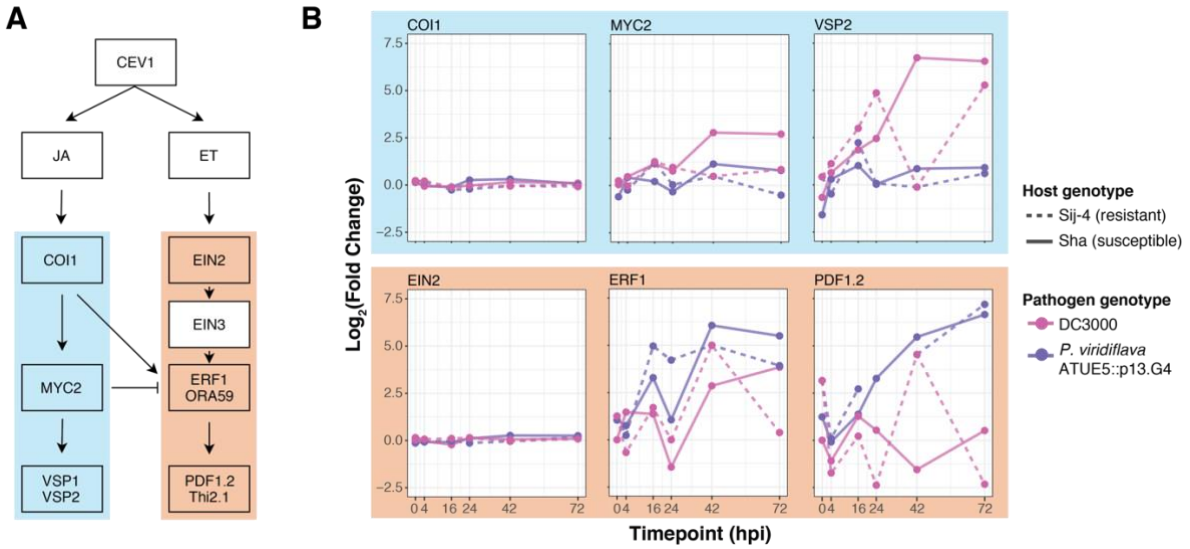

Figure 5 - *P. viridiflava* ATUE5 and DC3000 upregulate different branches of the JA/ET defense pathway. (A) Schematic representation of the JA/ET defense signaling pathway. **(B)** Log_2 fold change of marker genes of the JA/ET defense signaling pathway in *A. thaliana* susceptible Sha and resistant Sij-4 infected with *P. viridiflava* ATUE5, DC3000 and mock. The fold change corresponds to the contrast of the infected plants vs. mock-infected plants harvested at the same time point. Solid line: Sij-4, dashed line: Sha. Gray: mock-infected, pink: DC3000, purple: ATUE5 p13.G4. n = 3 - 5 replicates per host x genotype x time point combination, except for Sij-4 x DC3000 x 24 hpi were n = 2.

P. viridiflava infection correlates with increased abundance of JA precursors

Metabolomics is an emerging tool to study plant-pathogen interactions (39), given that metabolites play a key role in plant defense and in pathogen virulence. To assess the impact of *P. viridiflava* ATUE5 infection on susceptible and resistant *A. thaliana* genotypes, we measured primary and secondary metabolites in Sha and Sij-4 plants. Since hormone defense pathways are known to play a key role in establishing resistance against pathogens, we also measured six major plant hormones. We drip-inoculated Sha and Sij-4 plants with *P. viridiflava* ATUE5::p13.G4, DC3000 and mock buffer solution, and harvested them 2 dpi. We chose 2 dpi because at this time the susceptible host Sha did not display disease symptoms. Each sample corresponded to a pool of 3 to 5 plants. We measured phytohormones JA, JA-Ile, SA, abscisic acid (ABA), and auxins indole-3-acetic acid (IAA) and indole-3-carboxylic acid (ICA), and identified 67 primary and 534 secondary metabolites (268 lipids and 269 polar metabolites) after annotation and removal of metabolites that did not vary among samples.

We first validated our experimental approach by leveraging two bacterial metabolites present in our in-house library of reference compounds, toxins coronatine and phevamine. Both coronatine and phevamine contribute to DC3000 virulence (14,19,40), but the genes required for coronatine and other toxins biosynthesis were not found in *P. viridiflava* ATUE5 (11). In line with this, we detected these toxins only in DC3000-infected samples (Figure S9). Both toxins had higher abundance in

susceptible Sha compared to resistant Sij-4 plants, with phevamine being detected only in one Sij-4 sample. This demonstrates that our experimental approach allowed us to detect metabolites involved in plant-pathogen interactions.

To get an overview of the relationship between samples based on their primary and secondary metabolomic profile, we performed principal coordinate analysis (PCoA) of infected and control plants on the Euclidean distance matrix between samples. We found that the metabolomic profile of infected plants differed from that of mock-infected plants (Figure 6A,B, Figure S10). Metabolic change appeared greater in the susceptible host Sha than in the resistant host Sij-4, and upon *P. viridiflava* infection compared to DC3000 infection. Sha plants infected with *P. viridiflava* had a primary and secondary metabolomic profile distinct from that of mock- and DC3000-infected plants (Figure 6A,B, Figure S10). Resistant host genotype Sij-4 had a smaller metabolic remodeling upon infection: only *P. viridiflava* infected plants clustered separately from mock-infected plants.

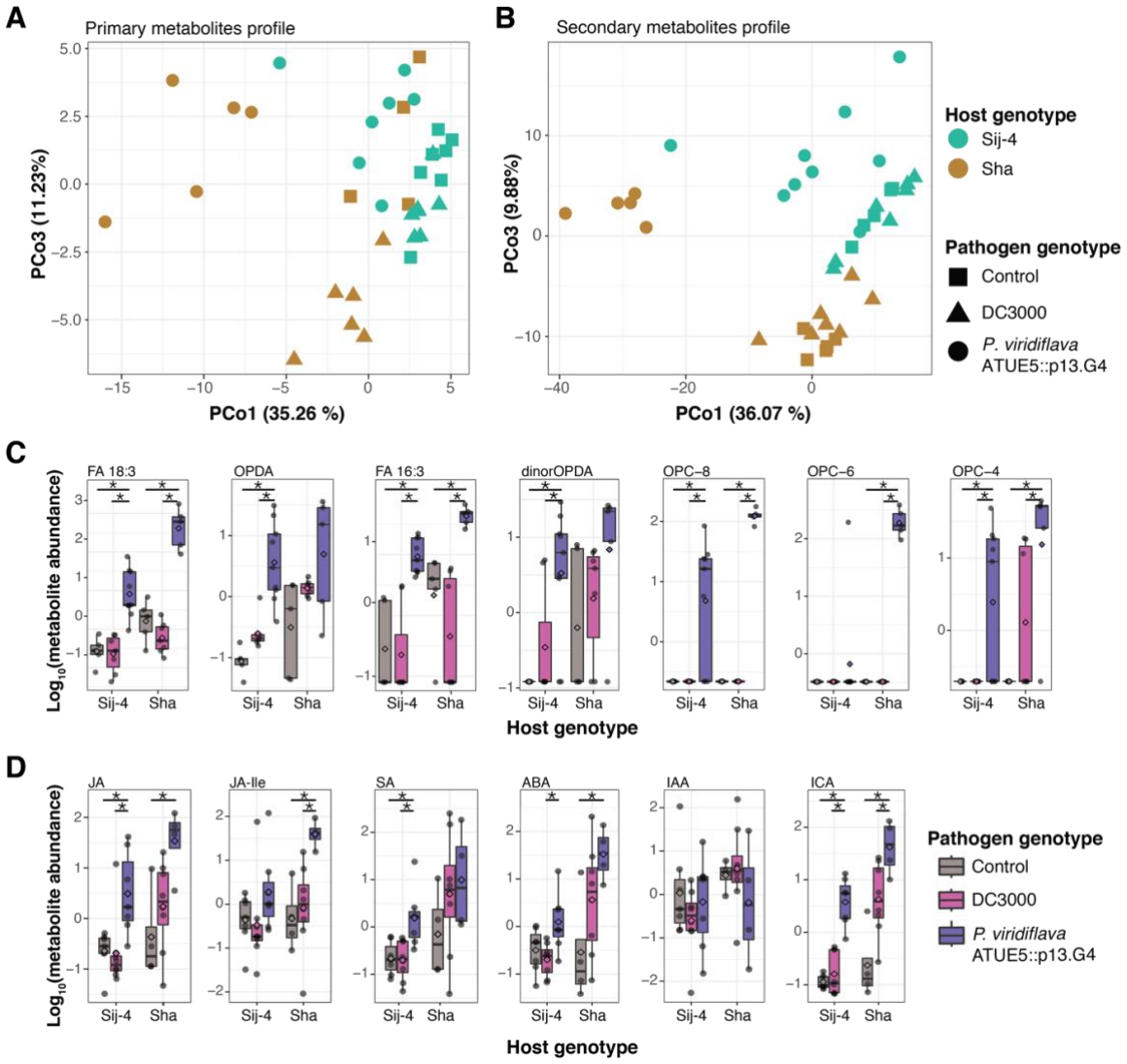

Figure 6 - *P. viridiflava* infection increases the abundance of JA. (A) Principal coordinate analysis of the Euclidean distance based on 72 primary metabolites identified in Sij-4 and Sha

infected and control plants. **(B)** Principal coordinate analysis of the Euclidean distance based on 537 secondary metabolites identified in Sha and Sij-4 infected and control plants. **(C)** Abundance of intermediates of the jasmonic acid biosynthesis pathway in Sij-4 and Sha infected and control plants. **(D)** Abundance of phytohormones in Sij-4 and Sha infected and control plants. FA: fatty acid, OPDA: oxophytodienoic acid, OPC-8: oxo-pentenyl-cyclopentane octanoic acid, OPC-6: oxo-pentenyl-cyclopentane hexanoic acid, OPC-4: oxo-pentenyl-cyclopentane butanoic acid. JA: jasmonic acid, JA-Ile: jasmonic acid - isoleucine conjugate, SA: salicylic acid, ABA: abscisic acid, IAA: indole-3-acetic-acid, ICA: indole-3-carboxylic acid. Aquamarine: Sij-4, ocre: Sha. Square and gray: mock-infected control, triangle and pink: DC3000, circle and purple: *P. viridiflava* ATUE5::p13.G4. n = 4 - 9 replicates per host x pathogen genotype combination. Each replicate was a pool of 3 to 5 individual plants. Asterisks indicate statistically significant differences between two groups (Dunn's test, adjusted p-value < 0.025)

To evaluate the influence of host and pathogen genotype on the metabolic profile we performed PERMANOVA on the Euclidean distance between samples. Regarding primary metabolites, host genotype explained 18.6% of the variance of the metabolomic profile (p-value = 0.0001) and pathogen genotype explained 23.3% (p-value = 0.0001). We found a significant genotype-by-genotype interaction, explaining 12.1% of the primary metabolomic profile (p-value = 0.0001). Comparably, for the secondary metabolites, host genotype, pathogen genotype and their interaction explained 18.0%, 22.0% and 9.78% of the variance (p-value = 0.0001, 0.0001 and 0.098 respectively). These results indicate that genotype-by-genotype interactions influence the outcome of infection not only in terms of the bacterial growth, but also of the host metabolome.

The distribution of control samples in the PCoA suggested Sij-4 and Sha have different metabolomic profiles before infection (Figure 6A,B, Figure S10), similar to what we observed in the transcriptome analysis (Figure 4A, Figure S5A,B). We found statistically significant differences between mock-infected Sha and Sij-4 in both primary (PERMANOVA, R^2 = 0.334, p-value = 0.0055) and secondary (PERMANOVA, R^2 = 0.365, p-value = 0.0075) metabolomic profiles. Our PCoA showed that Principal Coordinate (PCo) 1 explained roughly 35% of the variance in primary and secondary metabolites profiles. In both cases PCo1 was driven by the separation of Sha plants infected with *P. viridiflava* from the rest (Figure 6A,B), recapitulating what we observed with transcriptomics data (Figure 4A). Taken together, this indicates that the metabolite composition of Sij-4 and Sha is different before infection, and that *P. viridiflava* leads to metabolomic changes which are larger in the susceptible host Sha.

Given that *P. viridiflava*-infected Sha were clearly distinct on PC1, we identified specific metabolites underlying the separation of samples on this axis. For this, we calculated the Spearman's correlation between each sample's PC1 loading and their secondary metabolite abundance. A total of 359 (67%) secondary metabolites were significantly correlated with PCo1 (adj. p-value < 0.05). The Spearman's correlation coefficient ranged from -0.93 to 0.77, and most metabolites were negatively correlated

with PCo1 (284/359, 79.1%), indicating higher abundance in Sha infected with *P. viridiflava*. We selected the top 10% of the significantly correlated secondary metabolites based on their correlation coefficient, after removal of dipeptides. The abundance of all but one of these 25 secondary metabolites was negatively correlated to PCo1 (Table S5). Among these 25 metabolites was linoleic acid (FA 18:3), a precursor of JA.

Since our transcriptomics analysis suggested infection with *P. viridiflava* ATUE5 increased transcription of marker genes for the JA/ET signaling pathway, we examined the correlation between JA precursors' abundance and PCo1. We consistently found a strong negative correlation (Table 2), indicating *P. viridiflava* infection led to higher abundance of JA precursors in susceptible Sha at 2 dpi. The abundance of FA18:3, FA16:3 and OPC-4/6/8 was significantly different between Sha and Sij-4 (Kruskal-Wallis, p-value < 0.05). Furthermore, within each host genotype, the abundance of several JA precursors was significantly higher in *P. viridiflava* ATUE5-infected plants compared to mock- and DC3000-infected plants (Dunn's test, adjusted p-value < 0.025; Figure 6C). The last intermediates before JA synthesis, OPC-8,6,4, were detected almost exclusively in *P. viridiflava* ATUE5-infected plants. Taken together, these analyses indicate that infection of *A. thaliana* with *P. viridiflava* remodels both the primary and the secondary metabolite profile, and that this remodeling is larger with *P. viridiflava* than with the model pathogen DC3000. Susceptible Sha metabolome experienced the largest remodeling upon infection, driven mainly by an increased abundance of metabolites, including JA intermediates. Moreover, the differences in the metabolomic profile before infection between susceptible and resistant hosts could underlie the different outcome of infection, also suggested by our transcriptomics results.

Table 2. Spearman correlation between jasmonic acid precursors abundance and principal component 1.

Metabolite	Spearman's rho	p adj
FA 18:3	-0.780	1.33E-06
FA 16:3	-0.710	3.11E-06
OPC-4	-0.657	2.85E-05
OPC-6	-0.627	8.00E-05
OPC-8	-0.625	8.48E-05
dinorOPDA	-0.520	1.78E-03
OPDA	-0.515	2.32E-03

p adj.: adjusted p-value for multiple comparisons using the Benjamini-Hochberg method.

P. viridiflava infection increases the abundance of several defense-related hormones

To confirm that higher JA precursor levels in plants infected with *P. viridiflava* led to an increase in the hormone abundance, we measured JA and JA-Ile levels. JA-Ile results from the conjugation of JA and the amino acid isoleucine, and it is one of the active forms of JA. It promotes the interaction between COI1 and JAZ proteins, leading to the degradation of JAZ proteins and thus to the activation of transcription factors and JA-responsive genes (41).

We found no significant differences in the hormone abundance between mock-infected Sha and Sij-4 samples (Kruskal-Wallis, p-value > 0.25). We observed higher abundance of JA and JA-Ile in *P. viridiflava*-infected plants compared to mock- and DC3000-infected plants, in agreement with the increased abundance of JA precursors in these samples (Figure 6D). The abundance of JA in *P. viridiflava*-infected Sha was significantly higher than in infected Sij-4 (Kruskal-Wallis p-value = 0.038). Since we found no significant difference in JA and JA-Ile abundance between mock-infected Sha and Sij-4 samples (Kruskal-Wallis, p-value = 0.71 and 0.85, respectively), the differences in abundance are due to *P. viridiflava* infection and not to baseline differences between the host genotypes. Compared to DC3000-infected plants, JA was higher in both Sij-4 and Sha infected with *P. viridiflava* (Dunn's test, adj. p-value = 0.014 and 0.022, respectively), while JA-Ile was higher only in Sha infected with *P. viridiflava* compared to DC3000 (Dunn's test, adj. p-value = 0.006).

In addition to JA, other hormones play important roles in the defense response of *A. thaliana*. To determine the effect of *P. viridiflava* infection on said hormones and their potential role in resistance, we measured SA, ABA, and auxins IAA and ICA. Similarly, to JA and JA-Ile, infection with *P. viridiflava* led to an increase of SA, ABA and ICA in both susceptible and resistant host genotypes; while it had no effect on auxin IAA (Figure 6D). The abundance of ABA and ICA was not significantly different between mock-infected Sha and Sij-4 (p value = 0.500 and 0.571, respectively). In resistant Sij-4, ICA abundance after *P. viridiflava* infection was higher than in mock-infected plants (Dunn's test, p-value value = 0.001), but this was not the case for ABA (Dunn's test, p-value value = 0.104). In susceptible Sha, the abundance of both ABA and ICA was significantly higher in *P. viridiflava*-infected plants compared to mock-infected samples (Dunn's test, p-value value = 0.008 and 0.004, respectively). Moreover, ABA and ICA abundance was significantly higher in infected Sha compared to Sij-4 (Kruskal-Wallis, p-value = 0.0140 and 0.023, respectively). Unlike the hormones above, the abundance of SA was not significantly different between mock- and *P. viridiflava*-infected Sha plants (Kruskal-Wallis, p-value = 0.469; Figure 6D). In contrast, SA abundance was significantly different upon *P. viridiflava* infection in resistant Sij-4 compared to control and DC3000-infected plants (Kruskal-Wallis, p-value = 0.011 and 0.008, respectively).

In summary, while JA and ICA changed similarly in the susceptible and the resistant host genotype, JA-Ile and ABA increased significantly only in susceptible Sha, while

SA increased significantly only in resistant Sij-4 when compared to mock-infected plants. These differences could relate to the resistance against *P. viridiflava* infection. Together, our results show that *P. viridiflava* infection increases the abundance of the major phytohormones JA, JA-Ile, SA, ABA and ICA. Infection with DC3000 had a smaller impact on hormone levels compared to *P. viridiflava*.

JA-deficient mutants are more susceptible to *P. viridiflava* but not to DC3000 infection, and activation of the JA pathway decreases susceptibility

Our transcriptomics and metabolomics results strongly support a role for the ET branch of the JA/ET defense pathway in *A. thaliana* response against *P. viridiflava*. To confirm this, we took advantage of mutant lines deficient in the SA (*eds1-12*, *sid2-2)*, JA (*jar1-1*, *coi1-16*) and ET (*ein2-1*) defense pathways. We infected these mutants with *P. viridiflava* ATUE5 and DC3000, and measured both bacterial load and plant growth. All immune mutants, except for *coi1-16,* are in Col-0 background. *coi1-16* is in Col-5 background, which we determined to be as susceptible as Col-0 (Figure S11), and thus is shown compared with Col-0 here.

All SA and JA mutants tested supported significantly higher *P. viridiflava* load than Col-0 WT (Figure 7A). Additionally, plant growth was reduced upon infection, with infected plants having a significantly lower ratio of green pixels 3 dpi to before infection compared to Col-0 WT (Dunn's test, p-value adj. < 0.05; Figure 7B). The reduction in plant growth was most striking in the *coi1*-16 mutant. The susceptibility of *ein2-1* to *P. viridiflava* was similar to that of Col-0 WT, consistent with the JA/ET signaling via COI1 and ERF1, and not via EIN2. In contrast, only the SA mutant *sid2-2* showed significantly increased bacterial load upon DC3000 infection (Figure 7A), although both *sid2-2* and *eds1-12* had significantly lower plant growth (Dunn's test, p-value adj. < 0.05; Figure 7B).

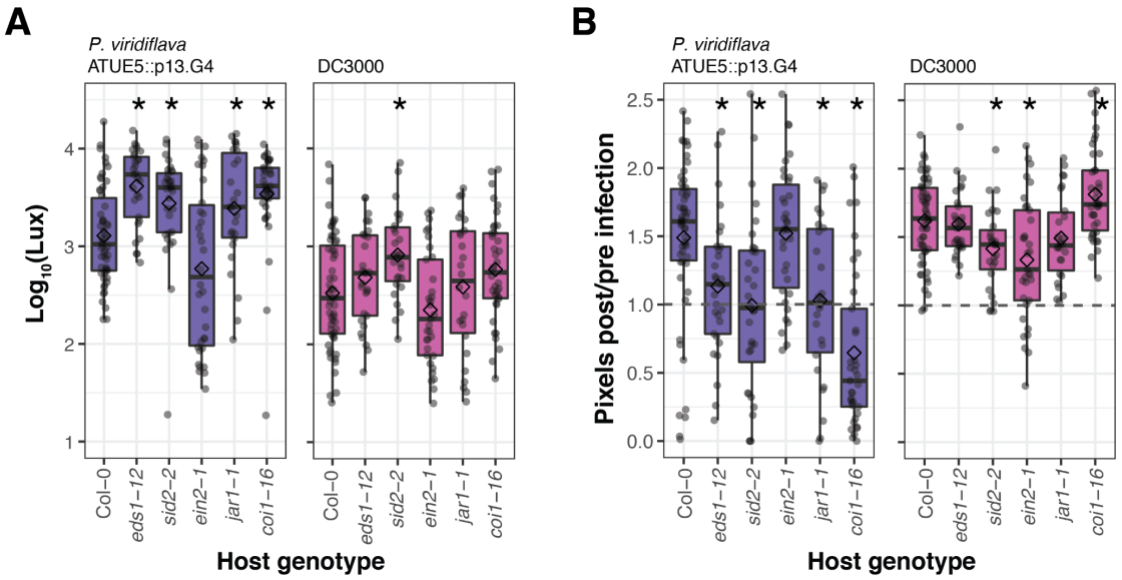

Figure 7 - JA mutants are more susceptible to *P. viridiflava* but not to DC3000 infection. (A) Bacterial load measured as luminescence 3 days post-infection in Col-0 WT and immunocompromised host genotypes infected with *P. viridiflava* ATUE5::p13.G4 or DC30000. **(B)**

Plant growth measured as the ratio of green pixels 3 days post-infection in Col-0 WT and immunocompromised host genotypes infected with *P. viridiflava* ATUE5::p13.G4 or DC30000. Asterisks indicate statistical significance of the Dunn's test at adj. p-value < 0.05 of each mutant compared to Col-0 WT. p-value adjusted for multiple comparisons using the Benjamini-Hochberg method. n = 24 - 55 plants per host x pathogen genotype combination. Purple: *P. viridiflava* ATUE5::p13.G4, pink: DC30000

Incidentally, we found that Sha susceptibility to *P. viridiflava* decreased when plants were infected after mechanical injury. This type of damage has been shown to increase JA and its precursors OPDA and dinor-OPDA levels, and to induce changes in defense-related gene expression as early as 15 minutes after injury (42). We cut one leaf of Sha and Sij-4 plants immediately before *P. viridiflava* infection, and observed in Sha a partial rescue of susceptibility based on the ratio of pixels 3 dpi compared to non-injured plants (Figure S8). Cutting a leaf had no effect on resistant host Sij-4. These results provide further support for JA/ET defense pathway in resistance against *P. viridiflava*.

DISCUSSION

Pseudomonas viridiflava is a highly prevalent bacterium in *Arabidopsis thaliana* populations around the globe (9,11,13) and an opportunistic pathogen. Despite its abundance and likely importance for *A. thaliana* fitness, little is known about the mechanisms of interaction and pathogenesis. In this work, we characterized the interaction between several *A. thaliana* genotypes and closely-related *P. viridiflava* isolates from Southwest Germany and compared these results to the infection process with the well-characterized *P. syringae* strain DC3000. We demonstrated that resistance against *P. viridiflava* infection is mediated by the JA/ET defense signaling pathway, specifically by the ET branch, a result consistent with defense mechanisms against Midwest USA isolates of *P. viridiflava*. Resistance to *P. viridiflava* was due to a primed immune response that allowed the plants to respond faster upon infection. Our results confirm the importance of the JA defense signaling pathway in the infection process of *A. thaliana* with *P. viridiflava*, and underlie the importance of an early immune response to achieve resistance against this ever-present opportunistic pathogen.

Our past work indicated that *P. viridiflava* ATUE5 isolates had varying pathogenicity on a single *A. thaliana* genotype (11). We have now expanded these results to a set of 21 hosts infected with 12 pathogen genotypes, aiming to elucidate the interaction mechanisms in this understudied natural pathosystem. We found that, while host genotype was the main determinant of bacterial load, there was a significant genotype-by-genotype interaction. In wild populations, there is extensive genetic variation in the genes underlying resistance and pathogenicity, and thus is it likely that genotype-by-genotype interactions play a role in determining the outcome of infection. In particular, the genotype-by-genotype interactions we identified may prevent a single bacterial

isolate from taking over in this pathosystem, thus explaining the vast genetic diversity of ATUE5 seen in Southwest Germany (11).

Sij-4 was able to withstand similar bacterial load as susceptible Sha without dying, which could indicate tolerance to *P. viridiflava.* Tolerance, the host's ability to reduce the negative effects of infection (43), is another mechanism of plant defense against pathogens. Although less studied than resistance, tolerance has a genetic basis and is relevant for plant-pathogen interactions (43). Previous studies have reported tolerance of *A. thaliana* to *P. syringae*, *P. viridiflava* and *Xanthomonas campestris* bacterial pathogens (12,28,44,45). However, the genetic basis for neither tolerance nor resistance in the *A. thaliana* - *P. viridiflava* pathosystem have been identified, and more studies are needed to address this question.

We found that resistance to *P. viridiflava* ATUE5::p13.G4 is a recessive trait segregating approximately at a 1:2 ratio. This segregation ratio could mean that two loci underlie resistance, which would be established when only one of the loci are recessive. Indeed, recessive resistance usually arises from recessive mutations in genes that encode factors critical for pathogen infection or in negative regulators of defense signaling pathways (46,47), resulting in loss-of-susceptibility. Recessive resistance has been mostly studied in viruses, with around half of the known host resistance genes being recessive (48). The most common genes identified as recessive resistance genes against viruses are translation initiation factors eIF4E and eIF4G (46). For bacterial pathogens, the best studied system for recessive resistance is rice - *Xanthomonas oryzae* pv. *oryzae*, where one third of the genes identified are recessive (47,49). The characterized genes encode a subunit of transcriptional factor IIAγ and SWEET proteins (47,50). Two recessive resistance genes with additive effects have been reported against *Xanthomonas euvesicatoria* in pepper (51,52). Recessive resistance is not associated with a (typical) hypersensitive response (47,52). Accordingly, we did not observe loss of green pixels in resistant Sij-4 or F_2 plants, further supporting a recessive genetic architecture. So far, no recessive resistance to bacterial pathogens has been reported in *A. thaliana.*

Compared to model pathogen DC3000, both transcriptome and metabolomic remodeling were larger upon *P. viridiflava* infection in both Sha and Sij-4 host genotypes. Consistent with previous reports, we found that DC3000 activates the MYC2 branch of the JA/ET pathway, and that only SA mutants were more susceptible to DC3000 infection (14,17,19). Even though DC3000 and *P. viridiflava* belong to the *P. syringae* species complex, our results underscore the differences in plant response to each of them. Much is still unknown about the interactions in the *A. thaliana* - *P. viridiflava* pathosystem, not only on the host side, but also on the pathogen side. For instance, the virulence mechanisms of *P. viridiflava* isolates and how they contribute to pathogen establishment and growth *in planta* are not yet established.

We suggest that resistant host genotype Sij-4 has a primed immune system that is able to respond faster than susceptible Sha to *P. viridiflava*, leading to resistance to infection. Sij-4 had higher expression of defense-response related genes compared to Sha before infection. Moreover, Sij-4 infected plants displayed JA-related responses at 16 hpi, while similar patterns were not observed at any time point in infected Sha plants. Although we observed upregulation of JA/ET marker genes and increased JA in Sha, this was only at later time points. We propose this response occurred too late to achieve resistance. In line with this, Mine *et al.* reported that plants susceptible to DC3000 infection had an almost identical transcriptome response to resistant plants but with several hours delay (37).

Hormone defense pathways play a central role in establishing resistance against pathogens (17,18,53), and there is extensive cross-talk among hormone-mediated defense responses, with both synergistic and antagonistic interactions (18,53,54). Hence, in addition to JA, other hormones could be related to resistance or susceptibility. SA is involved in resistance to *P. viridiflava*, although to a lesser extent than JA (23). Consistent with this, we found that SA-deficient mutants were more susceptible to *P. viridiflava* and that SA was significantly increased upon *P. viridiflava* infection only in the resistant host. Conversely, we detected an increase in ABA levels with *P. viridiflava* infection only in susceptible Sha. Recently, increased ABA abundance and ABA signaling induced by DC3000 were associated with stomatal closure and the establishment of an aqueous apoplast, favoring bacterial growth (26,27). DC3000 manipulation of ABA was mediated by functionally-redundant effectors AvrE and HopM1. AvrE was the only effector identified in *P. viridiflava* isolated from Southwest Germany (11), thus *P. viridiflava* could be using similar mechanisms as DC3000 to manipulate ABA and increase its virulence on susceptible Sha.

To the best of our knowledge, ours is the first study to use transcriptomics and metabolomics to study the natural *A. thaliana* - *P. viridiflava* pathosystem. This study provides evidence for the importance of the JA/ET defense pathway and the timing of the defense response for *A. thaliana* resistance against *P. viridiflava.* Further studies are needed to identify the genetic basis of resistance in this pathosystem, which we propose is due to a recessive mechanism.

METHODS

Plant material and growth conditions

A. thaliana natural accessions, i.e. host genotypes, Aitba-2, Apost-1, Ciste-2, Col-0, Ey15-2, Fei-0, HKT2.4, Jablo-1, Koch-1, Mammo-1, Monte-1, Qui-0, Rovero-1, Sha, Shigu-1, Sij-4, Slavi-1, Star-8, Toufl-1, TueWa1-2, Yeg-1 were included in this study (Table S1). F_2 seeds from a Sha x Sij-4 cross were used for segregation analysis, as described below (36). For immune mutant screening, *sid2-2*, *eds1-12*, *ein2-1*, *coi1-16* and *jar1-1* were used (Table S2).

Seeds were sterilized by overnight incubation at -80 ºC followed by at least 4 hours of vapor-phase sterilization with chlorine gas. Seeds were sowed on Petri dishes with ½ MS medium with vitamins and MES buffer (Duchefa M0255.0050) and stratified for seven days at 4°C in the dark. They were then grown under long-day (16 h) at 23°C in a growth chamber (Percival, model CU-36L5). After 3-5 days, seedlings were transferred to 24-well plates with the same medium, one seedling per well. 11 days-old plants were infected with single bacterial isolates.

To calculate the genetic distance between host genotypes included in this study, their single nucleotide (SNP) and insertion/deletion polymorphisms were obtained from the data generated by the 1001 Genomes project data (33,55). Host genotypes Ey15-2, Koch-1 and Toufl-1 were not present in this dataset. Genetic distance was calculated using PLINK (56,57) and plotted in R (58,59).

Axenic infection

Pseudomonas isolates were transformed to express the *luxCDABE* operon via electroporation or mating using pUC18-mini-Tn7T-Gm-lux, a gift from Herbert Schweizer (Addgene plasmid # 64963; (31), or a modified version including an oriT. Lux-transformed *Pseudomonas* isolates were grown overnight at 28°C in Luria-Bertani (LB) medium with 100 ng/mL of nitrofurantoin (Sigma), diluted the following morning 1:10 in 5 mL selective medium and grown for 3 - 4 additional hours. Bacteria were then pelleted at 3,500 g and brought to an OD_{600} of 0.01 in 10 mM $MgSO_4$. 100 µL of this bacterial suspension were used to drip-inoculate 11 days-old plants on 24-well plates, distributing the volume over the whole rosette. Plants were mock infected with 10 mM $MgSO_4$ as control. Plates with plants were returned to the growth chamber, and whole rosettes were cut for analysis between 0 and 72 hours post infection.

Luminescence and green pixels quantification

For luminescence quantification whole rosettes harvested 3 days post-infection were transferred to 96 deep-well plates (2.2 mL, Axygen), containing two 5 ± 0.03 mm glass beads (Roth) and 400 µL of 10 mM $MgSO_4$, and ground for two minutes at 20 Hz in a TissueLyser II (QIAGEN). Then, 10 mM $MgSO_4$ was added to a final volume of 1 mL, and 200 µL were transferred to a 96-well Lumitrac white plate. Luminescence was measured in a TECAN Spark multiplate reader with 2000 ms of integration time. Each well was measured three times, and the mean was calculated for further analysis. The signal of 10 mM $MgSO_4$ blanks was subtracted from the samples' signal before analysis. A constant equal to the lowest value was added to all samples before log_{10}-transformation, and these log_{10}-transformed data were used for the analysis.

Statistical analyses were performed using R v4.0.2 (59) and Rstudio (58). ANOVA was used to identify significant differences between host genotype, bacterial genotype, and to determine the proportion of variance explained by each of these factors and their interaction. The model used for the ANOVA was log_{10}(luminescence) ~ pathogen genotype * host genotype + day + edge; where day corresponds to the day of the experiment and edge to whether the plant was on the edge wells of the 24-well plate

or not. Variance decomposition was performed by dividing the ANOVA sum of squares of each variable by the total sum of squares. Significant differences among host or pathogen genotypes were determined using Tukey's Honestly Significant Differences implemented in the function HSD.test of the agricolae package v1.3-5 (60). An example of the commands used: agricolae::HSD.test(final.lm, "host_genotype", group = T, unbalanced = F. Differences were considered statistically significant when p-value < 0.05.

Green pixel quantification was performed as described before. Briefly, the number of green pixels was determined before and after infection for each plant. Plates with plants were photographed with a tripod-mounted Lumix DMC-TZ61 digital camera. Plates were illuminated from below to prevent light reflection on the lid. Individual plants were extracted from whole-plate images by applying thresholds in Lab color space, followed by a series of morphological operations to remove noise and non-plant objects. Finally, a GrabCut-based postprocessing was applied and csv files with plant IDs and green pixel counts were created. The workflow was implemented in Python 3.6 and bash using OpenCV 3.1.0 and scikit-image 0.13.0 for image processing operations. The ratio of green pixels was calculated by dividing the count of green pixels at a given time after infection by the count of green pixels before infection. A ratio > 1 meant plants were gaining green pixels, i.e. growing; while a ratio < 1 indicated a loss in green pixels due to plant death.

Segregation analysis

F_2 individuals derived from a Sha x Sij-4 cross generated previously (36), Sha and Sij-4 individuals were grown and infected as described above. Plant size was extracted from pictures taken before and 3 days post-infection. We classified plants as resistant or susceptible based on the minimum green pixel ratio 3 dpi of mock-infected F_2 plants. Statistical analyses were performed in R v4.0.2 (58,59). Plant size before infection and green pixels ratio 3 dpi were compared using a linear regression pixels ~ host genotype, with experiment as a covariate when indicated. The binomial proportion confidence interval at 95% for the proportion of resistant plants was estimated using Wilson method as implemented in the package binom (61). The experimental segregation ratio was compared to that expected from a 3:1 segregation using the Chi-squared goodness-of-fit test (chisq.test) implemented in the stats package from base R (59).

Transcriptomics

Single rosettes were harvested at 0, 4, 16, 42 and 72 hours post infection in 2 mL tubes containing two 5 ± 0.03 mm glass beads (Roth). Samples were snap-frozen in liquid nitrogen and stored at -80ºC until RNA extraction. Plants were ground for 30 seconds at 25 Hz in a TissueLyser II (QIAGEN) and RNA was extracted as described by Yaffe et al., 2012 (62) using EconoSpin plate columns (Epoch Life Science). mRNA libraries were prepared following an in-house protocol (63). Multiplexed libraries were sequenced single-end on a HiSeq3000 instrument (Illumina). Reads were mapped

against *A. thaliana* TAIR10 reference and transcript abundance was calculated using RSEM v1.2.31 (64) and bowtie2 v2.2.3 (65). Default parameters were chosen unless mentioned otherwise. Differential gene expression analyses were performed in R v4.1.0 (59) using the DESeq2 package v1.34.0 (66).

Seven samples with more than 13 million reads were subsampled to obtain 13 million reads before analysis. We then removed all samples with less than 3.4 million mapped reads from further analysis. One read was added to all read counts to avoid plotting -INF values in genes with read count 0 ($\log_{10}(0 + 1) = 0$; (67). Genes with less than ten counts over all samples were removed from downstream analyses. For exploratory data analysis, variance stabilizing transformed data were used and are referred here as normalized transcript counts. A gene was called as differentially expressed between two conditions when $|\log_2\text{FoldChange}| > 1$ and adj. p-value value < 0.01. Plots were generated using the R package ggplot2 (68). Gene ontology enrichment analysis was performed with ShinyGO v0.76, available at http://bioinformatics.sdstate.edu/go/ (69). A list of differentially expressed genes was used as input for enrichment, and the 25324 genes identified in this experiment were used as reference background. *Arabidopsis thaliana* was selected as species and 'GO biological process' as pathway database. Minimum (2) and maximum (2000) pathway size were left as default, and the remove redundancy option was selected.

Metabolomics

Rosettes were harvested two days post infection, before evident symptoms were visible. Between 3 and 5 plants were pooled to reach a sample weight of 30 mg ±10%. Samples were collected in 1.5 mL tubes containing three steel beads of approx. 4 mm diameter (KGM KU 4.000 G28 1.3505 StrG), snap-frozen in liquid nitrogen, and then ground for 30 seconds at 30 Hz in a TissueLyser II (QIAGEN). A methyl tert-butyl ether/methanol (MTBE/MeOH) solvent system was used to separate the hormone-containing fraction, metabolite-containing fraction and protein/starch/cell wall pellet from the ground plant material (70). This allowed us to measure hormones, primary and secondary metabolites from a single sample.

Measurement of primary metabolites (GC-MS)

The polar fraction was dried under vacuum, and the residue was derived for 120 min at 37°C (in 40 µL of 20 mg ml−1 methoxyamine hydrochloride (Sigma-Aldrich) in pyridine followed by a 30 min treatment at 37°C with 70 µL of N-methyl-N (trimethylsilyl)trifluoroacetamide (MSTFA reagent, Macherey-Nagel). Metabolites were measured according to Lisec et al. (71). The GC–MS system used was a gas chromatograph coupled to a time-of-flight mass spectrometer (Leco Pegasus HT TOF-MS). An auto sampler Gerstel MultiPurpose system injected the samples. Chromatograms and mass spectra were evaluated by using Chroma TOF 4.5 (Leco) and Xcalibur 2.1 software, peak area was normalized by comparison to an internal standard (Ribitol) and the fresh weight of the sample used for extraction (71).

Measurement of secondary metabolites (LC-MS)

The dried aqueous phase was measured using ultra-performance liquid chromatography coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific) in positive and negative ionization modes, as described earlier (72). Expressionist Refiner MS 12.0 (Genedata AG, Basel, Switzerland) was used for processing the LC-MS data with the following settings: chromatogram alignment (RT search interval 0.5 min), peak detection (minimum peak size 0.03 min, gap/peak ratio 50%, smoothing window 5 points, center computation by intensity-weighted method with intensity threshold at 70%, boundary determination using inflection points), isotope clustering (RT tolerance at 0.02 min, m/z tolerance 5 ppm, allowed charges 1–4), filtering for a single peak not assigned to an isotope cluster, charge and adduct grouping (RT tolerance 0.02 min, m/z tolerance 5 ppm). In-house library of authentic reference compounds was used to identify molecular features allowing 10 ppm mass deviation and dynamic retention time deviation (maximum 0.15 min).

Measurement of phytohormones (UPLC-ESI-MS)

A fixed volume (0.250 ml) of upper supernatant (MTBE phase) was transferred to a fresh 1.5-ml microcentrifuge tube and dried down using a SpeedVac concentrator at 25°C room temperature (RT) (concentration to dryness take up to 1 h at 30°C). The dried pellets were resuspended in 100 µl of water:methanol (50:50) solution and the resuspended samples were immediately subjected to UPLC-ESI-MS/MS hormonal analysis. Hormones were measured according to Salem et al., (70)) using a quadrupole/linear ion trap (QLIT) mass analyzer (e.g. 4000 QTRAP; AB Sciex Germany GmbH, https://sciex.com) with a multiple-reaction monitoring (MRM) scan type equipped with an electrospray ionization (ESI) source (e.g. Turbo V™ Ion Source; AB Sciex) and attached to the UPLC system. analyst 1.6.2 (AB Sciex) was used for instrument control, data acquisition, processing and analysis (70).

Data analysis

Median-normalized primary and secondary metabolite abundances were $\log_{10}$-transformed, scaled and centered before further analysis. Hormone levels were $\log_{10}$-transformed, scaled and centered before further analysis. Metabolites that did not vary among samples were removed. Statistical tests and plotting were performed using R (59) and Rstudio (58). To detect patterns between treatments, principal coordinate analysis (PCoA) was performed using the cmdscale function. Permutational multivariate analysis of variance (PERMANOVA) was conducted on Euclidean distance between samples, using the function adonis2 of the vegan package (73). To detect statistically significant differences in metabolites' abundance between treatments within each host genotype, a Kruskal-Wallis test was performed, followed by Dunn's test with Benjamini-Hochberg multiple comparison correction, as implemented in the dunn.test package (74). Differences were called statistically significant when adj. p-value < 0.025 as recommended by the dunn.test function.

ACKNOWLEDGEMENTS

We thank Manuela Neumann and Peter Laurie for their help in the axenic screen. We thank Wei Yuan, Max Collenberg and Thanvi Srikant for their help with transcriptomics. We thank Andy Gloss for providing six luminescence-tagged *Pseudomonas* isolates. We thank Jacobo de la Cuesta-Zuluaga for help with statistical analyses. This work was supported by the Max Planck Society.

REFERENCES

1. Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012 Dec;10(12):828–40.
2. Almario J, Mahmoudi M, Kroll S, Agler M, Placzek A, Mari A, et al. The Leaf Microbiome of Arabidopsis Displays Reproducible Dynamics and Patterns throughout the Growing Season. MBio. 2022 Apr 14;e0282521.
3. Hirano SS, Upper CD. Bacteria in the leaf ecosystem with emphasis on Pseudomonas syringae-a pathogen, ice nucleus, and epiphyte. Microbiol Mol Biol Rev. 2000 Sep;64(3):624–53.
4. Legein M, Smets W, Vandenheuvel D, Eilers T, Muyshondt B, Prinsen E, et al. Modes of Action of Microbial Biocontrol in the Phyllosphere. Front Microbiol. 2020 Jul 14;11:1619.
5. Wilkie JP, Dye DW, Watson DRW. Further hosts of Pseudomonas viridiflava. New Zealand Journal of Agricultural Research. 1973 Aug 1;16(3):315–23.
6. Lundberg DS, de Pedro Jové R, Ayutthaya PPN, Karasov TL, Shalev O, Poersch K, et al. Contrasting patterns of microbial dominance in the Arabidopsis thaliana phyllosphere. bioRxiv. 2021. p. 2021.04.06.438366. doi: 10.1101/2021.04.06.438366v2.full
7. Goss EM, Kreitman M, Bergelson J. Genetic diversity, recombination and cryptic clades in Pseudomonas viridiflava infecting natural populations of Arabidopsis thaliana. Genetics. 2005 Jan;169(1):21–35.
8. Bartoli C, Berge O, Monteil CL, Guilbaud C, Balestra GM, Varvaro L, et al. The Pseudomonas viridiflava phylogroups in the P. syringae species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. Environ Microbiol. 2014 Jul;16(7):2301–15.
9. Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J. Pseudomonas viridiflava and P. syringae--natural pathogens of Arabidopsis thaliana. Mol Plant Microbe Interact. 2002 Dec;15(12):1195–203.
10. Shalev O, Karasov TL, Lundberg DS, Ashkenazy H, Pramoj Na Ayutthaya P, Weigel D. Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant. Nat Ecol Evol. 2022 Apr;6(4):383–96.
11. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe. 2018 Jul 11;24(1):168–79.e4.
12. Goss EM, Bergelson J. Fitness consequences of infection of Arabidopsis thaliana with its natural bacterial pathogen Pseudomonas viridiflava. Oecologia. 2007 May;152(1):71–81.
13. Karasov TL, Neumann M, Shirsekar G, Monroe G, PATHODOPSIS Team, Weigel

D, et al. Drought selection on Arabidopsis populations and their microbiomes. bioRxiv. 2022. doi: 10.1101/2022.04.08.487684v1

14. Xin XF, He SY. Pseudomonas syringae pv. tomato DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants. Annu Rev Phytopathol. 2013 May 31;51:473–98.
15. Xin XF, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. Nat Rev Microbiol. 2018 May;16(5):316–28.
16. Jones JDG, Dangl JL. The plant immune system. Nature. 2006 Nov 16;444(7117):323–9.
17. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu Rev Phytopathol. 2005;43:205–27.
18. Li N, Han X, Feng D, Yuan D, Huang LJ. Signaling Crosstalk between Salicylic Acid and Ethylene/Jasmonate in Plant Defense: Do We Understand What They Are Whispering? Int J Mol Sci. 2019 Feb 4;20(3).
19. Zheng XY, Spivey NW, Zeng W, Liu PP, Fu ZQ, Klessig DF, et al. Coronatine promotes Pseudomonas syringae virulence in plants by activating a signaling cascade that inhibits salicylic acid accumulation. Cell Host Microbe. 2012 Jun 14;11(6):587–96.
20. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol. 2012 Aug;13(6):614–29.
21. Gomila M, Busquets A, Mulet M, García-Valdés E, Lalucat J. Clarification of Taxonomic Status within the Pseudomonas syringae Species Group Based on a Phylogenomic Analysis. Front Microbiol. 2017 Dec 7;8:2422.
22. Lipps SM, Samac DA. Pseudomonas viridiflava: An internal outsider of the Pseudomonas syringae species complex. Mol Plant Pathol. 2022 Jan;23(1):3–15.
23. Jakob K, Kniskern JM, Bergelson J. The Role of Pectate Lyase and the Jasmonic Acid Defense Response in Pseudomonas viridiflava Virulence. Mol Plant Microbe Interact. 2007 Feb 1;20(2):146–58.
24. Araki H, Tian D, Goss EM, Jakob K, Halldorsdottir SS, Kreitman M, et al. Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. Proc Natl Acad Sci U S A. 2006 Apr 11;103(15):5887–92.
25. Xin XF, Nomura K, Ding X, Chen X, Wang K, Aung K, et al. Pseudomonas syringae Effector Avirulence Protein E Localizes to the Host Plasma Membrane and Down-Regulates the Expression of the NONRACE-SPECIFIC DISEASE RESISTANCE1/HARPIN-INDUCED1-LIKE13 Gene Required for Antibacterial Immunity in Arabidopsis. Plant Physiol. 2015 Sep;169(1):793–802.
26. Hu Y, Ding Y, Cai B, Qin X, Wu J, Yuan M, et al. Bacterial effectors manipulate plant abscisic acid signaling for creation of an aqueous apoplast. Cell Host Microbe. 2022 Apr 13;30(4):518–29.e6.
27. Roussin-Léveillée C, Lajeunesse G, St-Amand M, Veerapen VP, Silva-Martins G, Nomura K, et al. Evolutionarily conserved bacterial effectors hijack abscisic acid signaling to induce an aqueous environment in the apoplast. Cell Host Microbe. 2022 Apr 13;30(4):489–501.e4.
28. Goss EM, Bergelson J. Variation in resistance and virulence in the interaction

between Arabidopsis thaliana and a bacterial pathogen. Evolution. 2006 Aug;60(8):1562–73.

29. Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, et al. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods. 2005 Jun;2(6):443–8.
30. pUC18-mini-Tn7T-Gm-lux. Addgene plasmid # 64963.
31. Choi KH, Schweizer HP. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. Nat Protoc. 2006;1(1):153–61.
32. Matsumoto A, Schlüter T, Melkonian K, Takeda A, Nakagami H, Mine A. A versatile Tn7 transposon-based bioluminescence tagging tool for quantitative and spatial detection of bacteria in plants. Plant Commun. 2022 Jan 10;3(1):100227.
33. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016 Jul 14;166(2):481–91.
34. Lindeman, Harold R, Merenda, Francis P, Gold, Z. R. Introduction to bivariate and multivariate analysis. Glenview, Ill.: Scott, Foresman; 1980.
35. Grömping U. Variable importance in regression models. WIREs Comput Stat. 2015 Mar;7(2):137–52.
36. Chae E, Bomblies K, Kim ST, Karelina D, Zaidem M, Ossowski S, et al. Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell. 2014 Dec 4;159(6):1341–51.
37. Mine A, Seyfferth C, Kracher B, Berens ML, Becker D, Tsuda K. The Defense Phytohormone Signaling Network Enables Rapid, High-Amplitude Transcriptional Reprogramming during Effector-Triggered Immunity. Plant Cell. 2018 Jun;30(6):1199–219.
38. Mengiste T. Plant immunity to necrotrophs. Annu Rev Phytopathol. 2012 Jun 15;50:267–94.
39. Castro-Moretti FR, Gentzel IN, Mackey D, Alonso AP. Metabolomics as an Emerging Tool for the Study of Plant-Pathogen Interactions. Metabolites. 2020 Jan 29;10(2).
40. O'Neill EM, Mucyn TS, Patteson JB, Finkel OM, Chung EH, Baccile JA, et al. Phevamine A, a small molecule that suppresses plant immune responses. Proc Natl Acad Sci U S A. 2018 Oct 9;115(41):E9514–22.
41. Ruan J, Zhou Y, Zhou M, Yan J, Khurshid M, Weng W, et al. Jasmonic Acid Signaling Pathway in Plants. Int J Mol Sci. 2019 May 20;20(10).
42. Reymond P, Weber H, Damond M, Farmer EE. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell. 2000 May;12(5):707–20.
43. Pagán I, García-Arenal F. Tolerance of Plants to Pathogens: A Unifying View. Annu Rev Phytopathol. 2020 Aug 25;58:77–96.
44. Kover PX, Schaal BA. Genetic variation for disease resistance and tolerance among Arabidopsis thaliana accessions. Proc Natl Acad Sci U S A. 2002 Aug 20;99(17):11270–4.
45. Tsuji J, Somerville SC, Hammerschmidt R. Identification of a gene in Arabidopsis thaliana that controls resistance to Xanthomonas campestris pv. campestris. Physiol Mol Plant Pathol. 1991 Jan 1;38(1):57–65.

46. Hashimoto M, Neriya Y, Yamaji Y, Namba S. Recessive Resistance to Plant Viruses: Potential Resistance Genes Beyond Translation Initiation Factors. Front Microbiol. 2016 Oct 26;7:1695.
47. Iyer-Pascuzzi AS, McCouch SR. Recessive resistance genes and the Oryza sativa-Xanthomonas oryzae pv. oryzae pathosystem. Mol Plant Microbe Interact. 2007 Jul;20(7):731–9.
48. Truniger V, Aranda MA. Recessive resistance to plant viruses. Adv Virus Res. 2009;75:119–59.
49. Cao J, Zhang M, Xiao J, Li X, Yuan M, Wang S. Dominant and Recessive Major R Genes Lead to Different Types of Host Cell Death During Resistance to Xanthomonas oryzae in Rice. Front Plant Sci. 2018 Nov 21;9:1711.
50. Yuan M, Wang S. Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. Mol Plant. 2013 May;6(3):665–74.
51. Vallejos CE, Jones V, Stall RE, Jones JB, Minsavage GV, Schultz DC, et al. Characterization of two recessive genes controlling resistance to all races of bacterial spot in peppers. Theor Appl Genet. 2010 Jun;121(1):37–46.
52. Jones JB, Minsavage GV, Roberts PD, Johnson RR, Kousik CS, Subramanian S, et al. A non-hypersensitive resistance in pepper to the bacterial spot pathogen is associated with two recessive genes. Phytopathology. 2002 Mar;92(3):273–7.
53. Checker VG, Kushwaha HR, Kumari P, Yadav S. Role of Phytohormones in Plant Defense: Signaling and Cross Talk. In: Singh A, Singh IK, editors. Molecular Aspects of Plant-Pathogen Interaction. Singapore: Springer Singapore; 2018. p. 159–84.
54. Hou S, Tsuda K. Salicylic acid and jasmonic acid crosstalk in plant immunity. Essays Biochem. 2022 Jun 14.
55. 1001Genomes. Available from: https://1001genomes.org/
56. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.
57. Purcell S. PLINK. 2017. Available from: http://pngu.mgh.harvard.edu/purcell/plink/
58. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc.; 2019. Available from: http://www.rstudio.com/
59. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
60. de Mendiburu F. agricolae: Statistical Procedures for Agricultural Research. 2021. Available from: https://CRAN.R-project.org/package=agricolae
61. Dorai-Raj S. binom: Binomial Confidence Intervals for Several Parameterizations. 2022.
62. Yaffe H, Buxdorf K, Shapira I, Ein-Gedi S, Moyal-Ben Zvi M, Fridman E, et al. LogSpin: a simple, economical and fast method for RNA isolation from infected or healthy plants and other eukaryotic tissues. BMC Res Notes. 2012 Jan 19;5(1):1–8.
63. Cambiagno DA, Giudicatti AJ, Arce AL, Gagliardi D, Li L, Yuan W, et al. HASTY modulates miRNA biogenesis by linking pri-miRNA transcription and processing.

Mol Plant. 2021 Mar 1;14(3):426–39.

64. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug 4;12:323.
65. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.
66. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
67. Barragan AC, Collenberg M, Wang J, Lee RRQ, Cher WY, Rabanal FA, et al. A Truncated Singleton NLR Causes Hybrid Necrosis in Arabidopsis thaliana. Mol Biol Evol. 2021 Jan 23;38(2):557–74.
68. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org
69. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020 Apr 15;36(8):2628–9.
70. Salem MA, Yoshida T, Perez de Souza L, Alseekh S, Bajdzienko K, Fernie AR, et al. An improved extraction method enables the comprehensive analysis of lipids, proteins, metabolites and phytohormones from a single sample of leaf tissue under water-deficit stress. Plant J. 2020 Aug;103(4):1614–32.
71. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc. 2006;1(1):387–96.
72. Giavalisco P, Li Y, Matthes A, Eckhardt A, Hubberten HM, Hesse H, et al. Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. Plant J. 2011 Oct;68(2):364–76.
73. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2020. Available from: https://CRAN.R-project.org/package=vegan
74. Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. 2017. Available from: https://CRAN.R-project.org/package=dunn.test

SUPPLEMENTARY MATERIAL

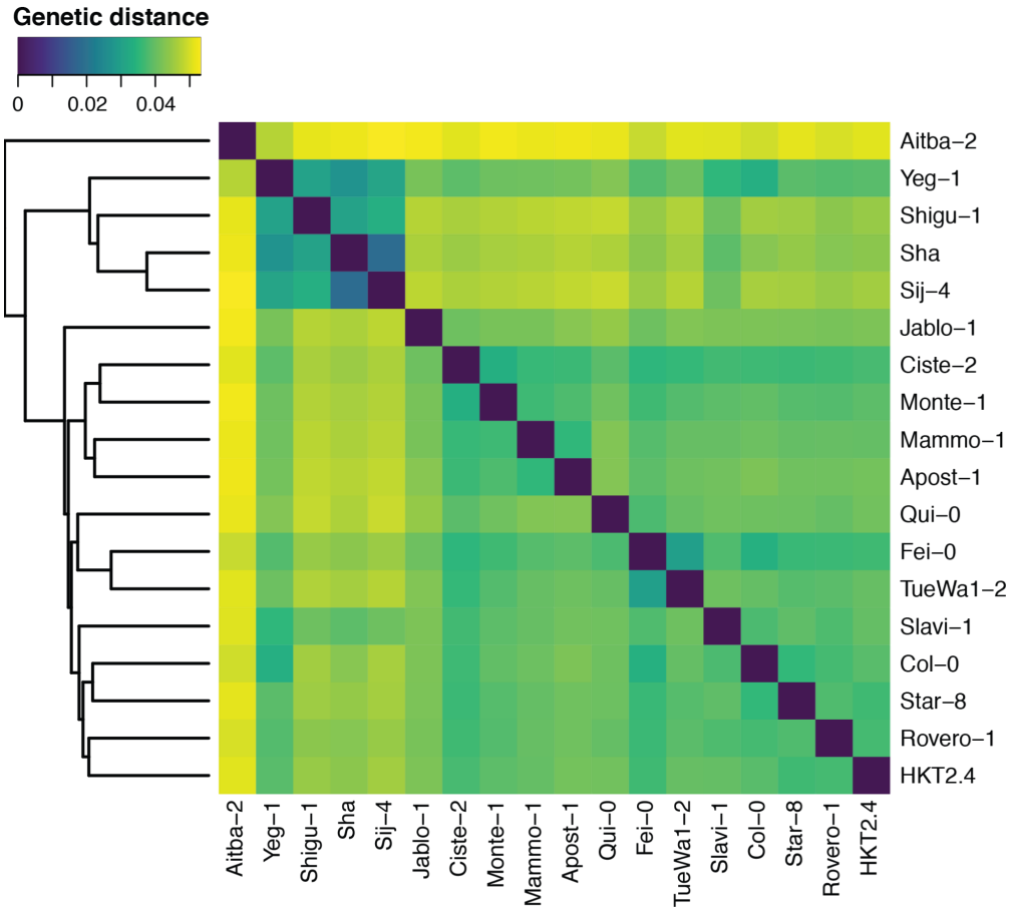

Figure S1 - Pairwise genetic distance between *A. thaliana* genotypes included in this study. Pairwise genetic distance between 18 *A. thaliana* genotypes for which single nucleotide polymorphisms data was available from the 1001 Genomes project (33). No data was available for Ey15-2, Koch-1 and Toufl-1.

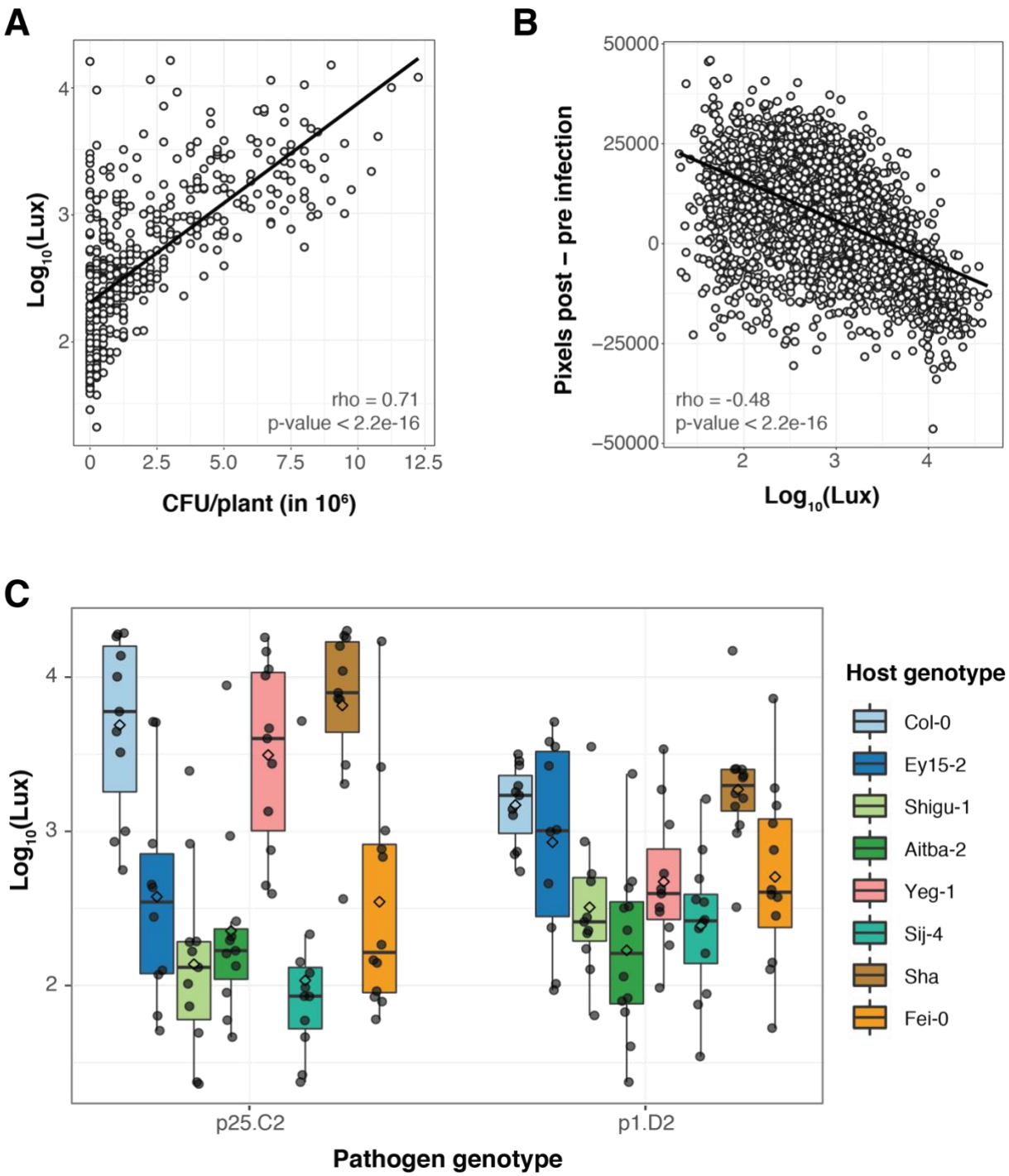

Figure S2 - Genotype-by-genotype interactions between *A. thaliana* and *P. viridiflava* ATUE5. **(A)** Strong positive correlation between the bacterial load measured as luminescence and the total number of colony forming units (CFU) per plant in a selection of infected plants. **(B)** Strong negative correlation between luminescence and the ratio of green pixels 3 days post infection (green pixels 3 days post infection - green pixels before infection). **(C)** Luminescence 3 days post-infection of two *P. viridiflava* isolates in six selected host genotypes. n = 440 plants for A, n = 2841 plants for B and n = 8 - 15 replicates per host x pathogen genotype combination for C.

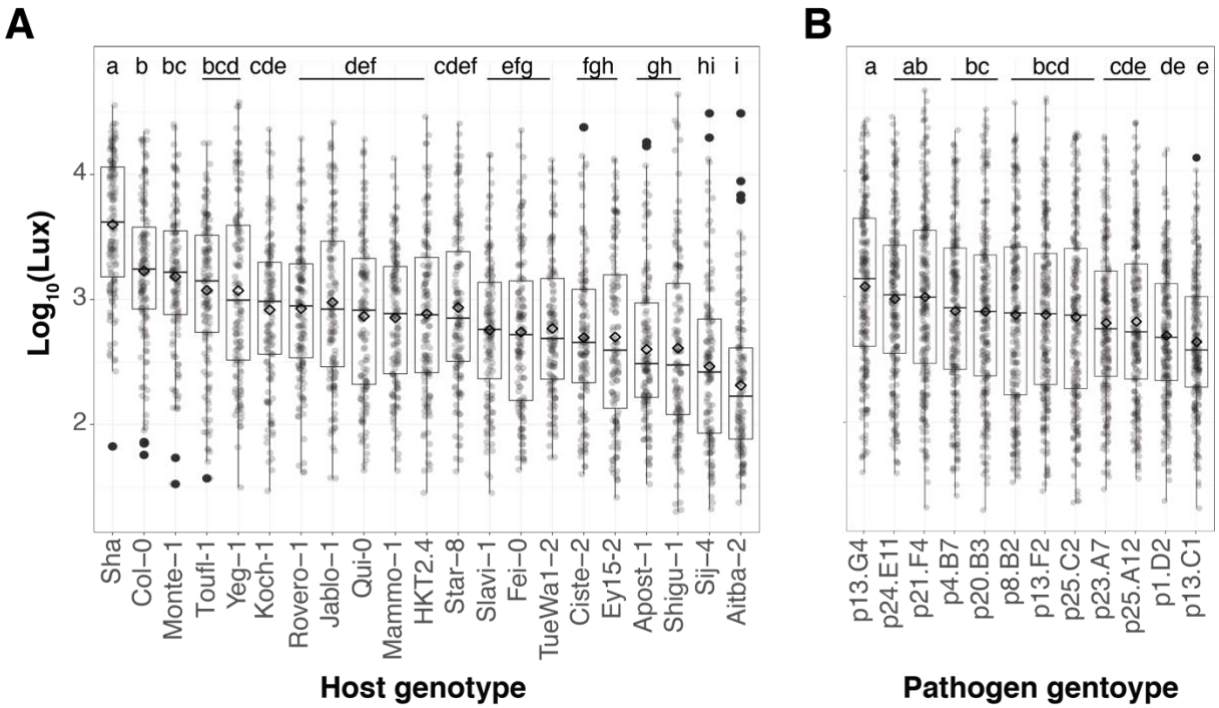

Figure S3 – *P. viridiflava* growth *in planta* is influenced by host genotype and pathogen genotype to different extents. Distribution of bacterial load measured as luminescence for all pathogen genotypes on each host genotype **(A)** and of each pathogen genotype across all host genotypes **(B)**. Diamonds represent the mean $\log_{10}$luminescence. Letters above boxplot indicate Tukey's HSD results, genotypes with the same letter were not significantly different from each other at a p-value < 0.05. n = 231 - 244. We repeated the experiment with a subset of seven host genotypes and seven pathogen genotypes with similar results.

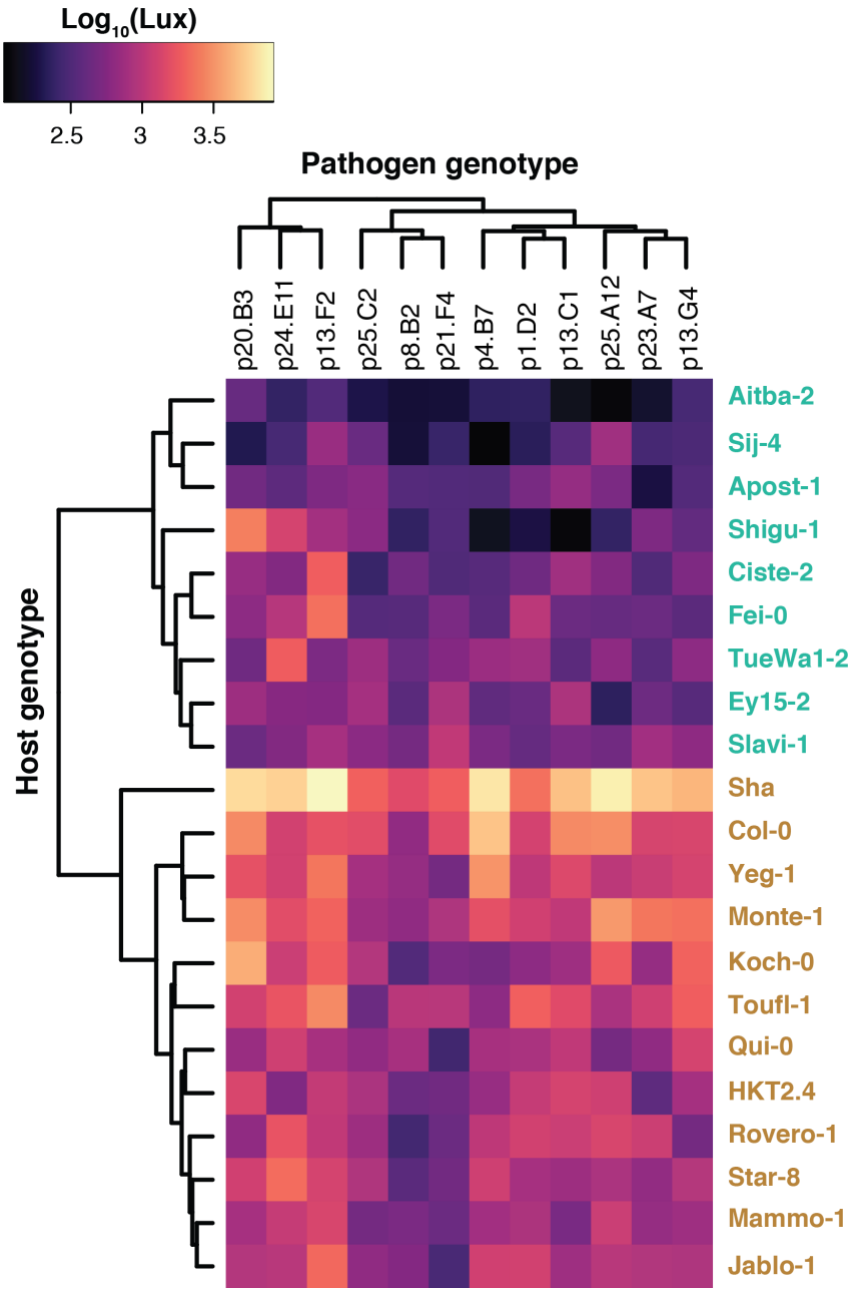

Figure S4 - Clustering of pathogen genotype based on luminescence signal does not recapitulate phylogenetic relationships. Pathogen load measured as luminescence 3 days post-infection in all host x pathogen genotype combinations. In total, 21 host genotypes were drip-inoculated with 12 luminescence-labeled pathogen genotypes. Host and pathogen genotypes are ordered according to hierarchical clustering based on the luminescence data displayed in the heatmap. The data presented in the same as in Figure 2a, but note the different order of pathogen genotype: here, hierarchical clustering based on luminescence signal instead of phylogenetic relationship between isolates. n = 8 - 15 for each host x pathogen genotype combination. We repeated the experiment with a subset of seven host genotypes and seven pathogen genotypes with similar results.

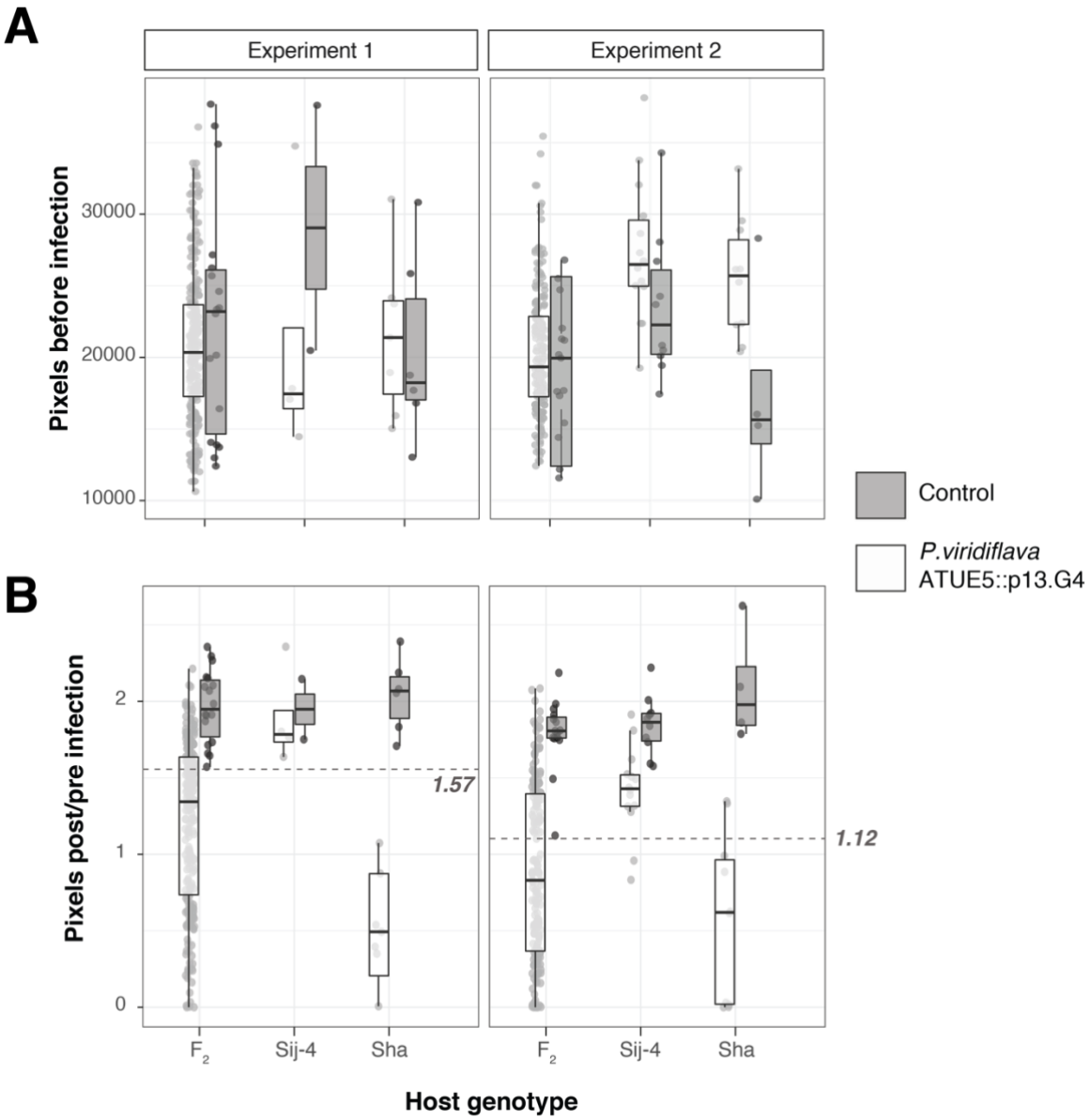

Figure S5 - Resistance to *P. viridiflava* is a recessive trait. The segregation of resistance to *P. viridiflava* ATUE5 isolate p13.G4 was determined in a F_2 population resulting from the cross of susceptible Sha and resistant Sij-4 parents. Two independent experiments (Experiment 1 and Experiment 2) were performed. Green pixels were used as a proxy for plant size before **(A)** and 3 days post-infection with *P. viridiflava* ATUE5 isolate p13.G4 and control. **(B)** Ratio of green pixels 3 days post-infection to before infection. Plants with a ratio > 1 gained green pixels, i.e., were growing, while plants with a ratio < 1 lost green pixels, i.e., were dying. A threshold for classifying plants as susceptible or resistant was determined for each experiment based on the ratio of pixels post/pre-infection of control plants, indicated by the dashed line. Plants above said threshold were classified as resistant, and those below the threshold, as susceptible (see Table1). Gray: mock-infected plants, white: plants infected with *P. viridiflava* ATUE5::p13.G4.

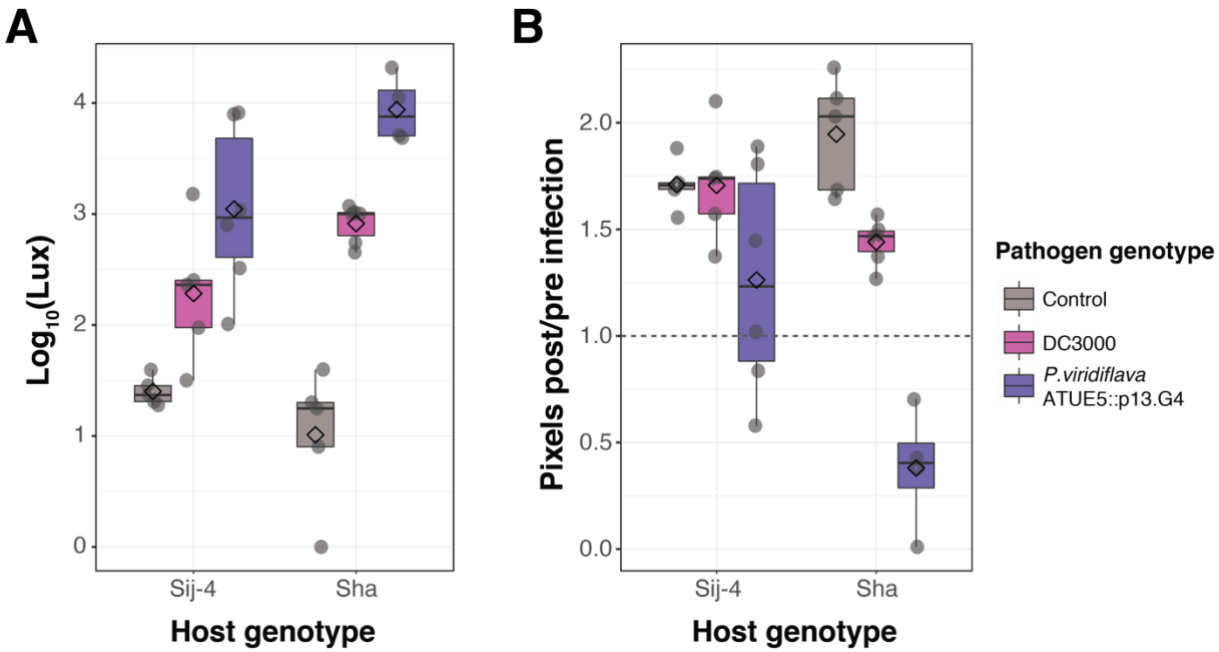

Figure S6 – *P. viridiflava* is more virulent in *A. thaliana* compared to DC3000. (A) Bacterial load measured as luminescence 3 days post-infection with *P. viridiflava* ATUE5::p13.G4 and DC30000 and **(B)** plant growth measured as the ratio of green pixels 3 days post-infection in Sij-4 (resistant) and Sha (susceptible) *A. thaliana* genotypes. n = 27 - 55 plants per host x pathogen genotype combination.

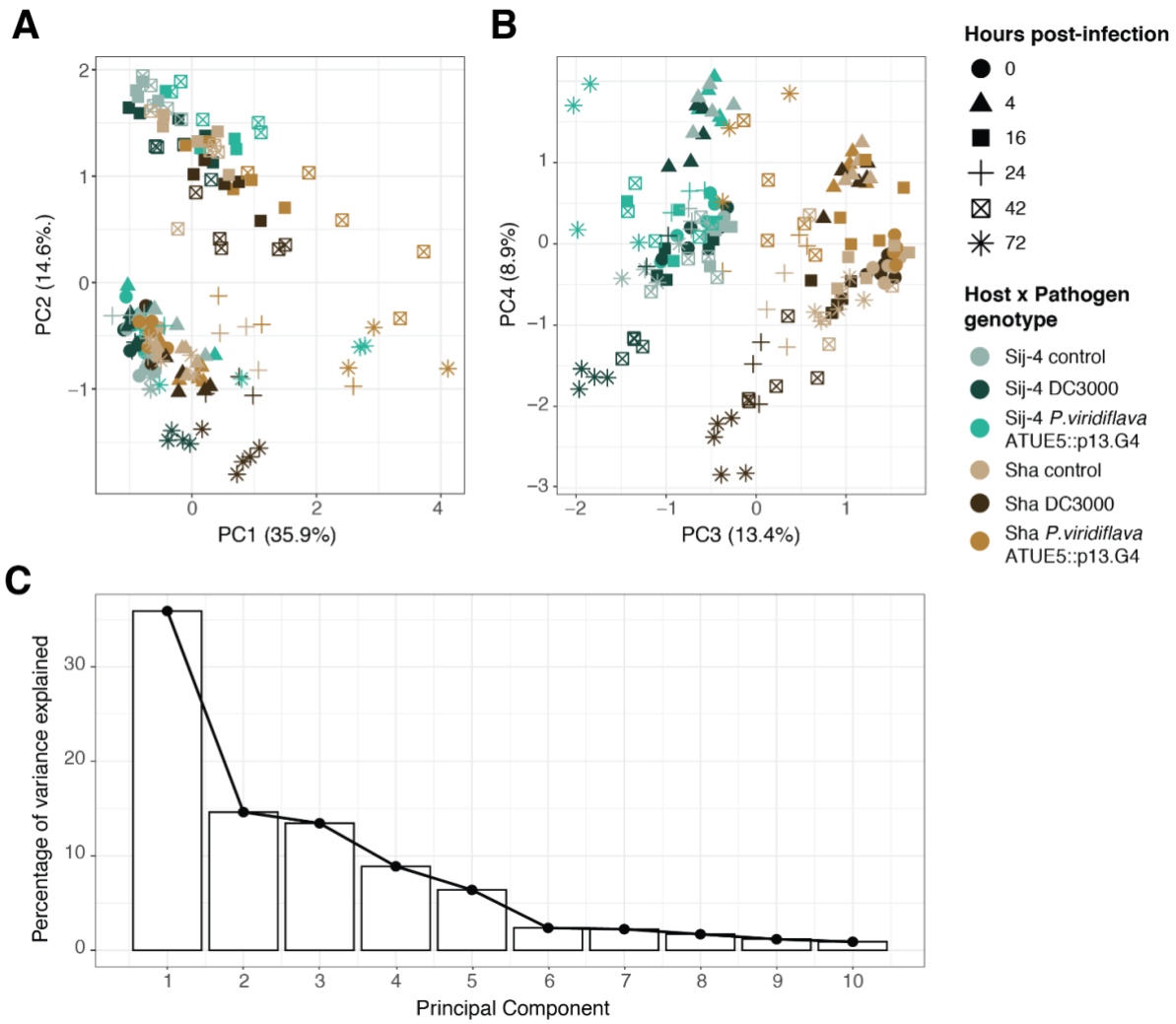

Figure S7 - Principal component analysis of the transcriptome of susceptible and resistant *A. thaliana* genotypes infected with *P. viridiflava* ATUE5 and DC3000. (A, B) Principal component analysis. Symbols represent time post-infection at collection, colors represent the host x pathogen genotype combination. **(C)** Scree plot of the first 10 principal components.

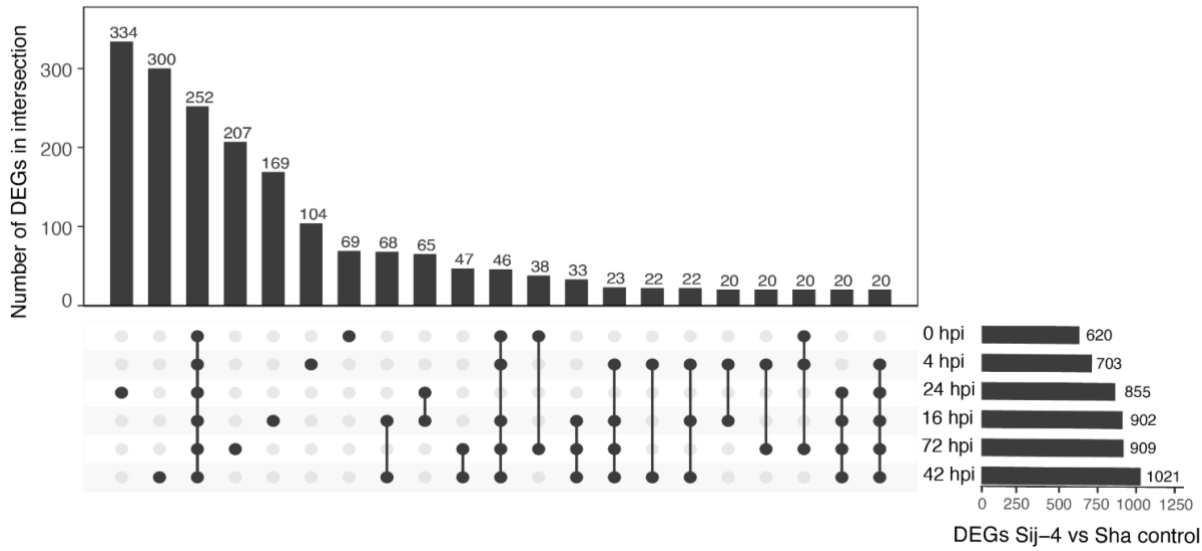

Figure S8 - Overlap in differentially expressed genes between mock-infected Sha and Sij-4. Intersection of differentially expressed genes between Sij-4 and Sha mock-infected plants at different timepoints. DEGs = differentially expressed genes. n = 3 - 5 replicates per host x genotype x time point combination, except for Sij-4 x DC3000 x 24 hpi were n = 2.

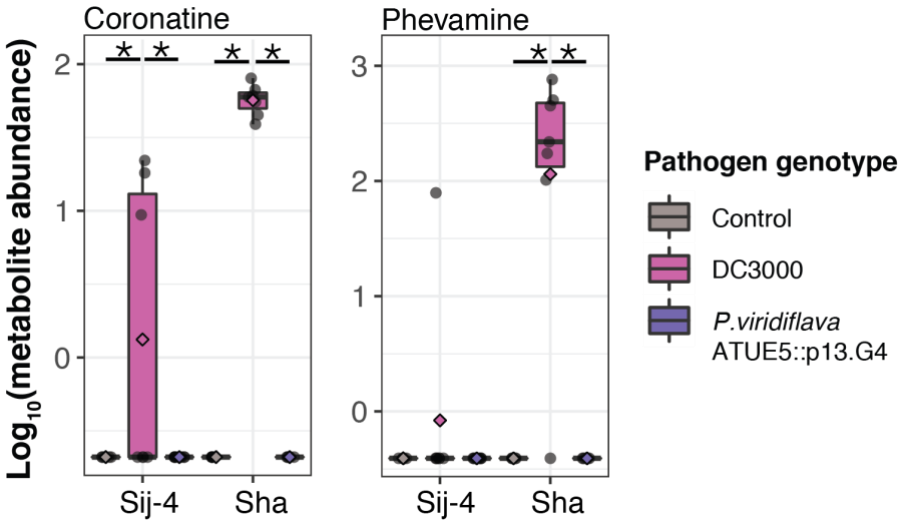

Figure S9 - Toxins coronatine and phevamine are only detected in DC3000-infected plants. Log_{10} abundance of coronatine and phevamine in susceptible and resistant *A. thaliana* genotypes infected with *P. viridiflava* ATUE5 and DC3000. DC3000 is known to produce these toxins, while the genes required for coronatine biosynthesis were not found in *P. viridiflava* ATUE5 (11). n = 5 - 9 replicates per host x pathogen genotype combination. Each replicate was a pool of 3 to 5 individual plants. Asterisks indicate statistically significant differences between two groups (Dunn's test, adjusted p-value < 0.025)

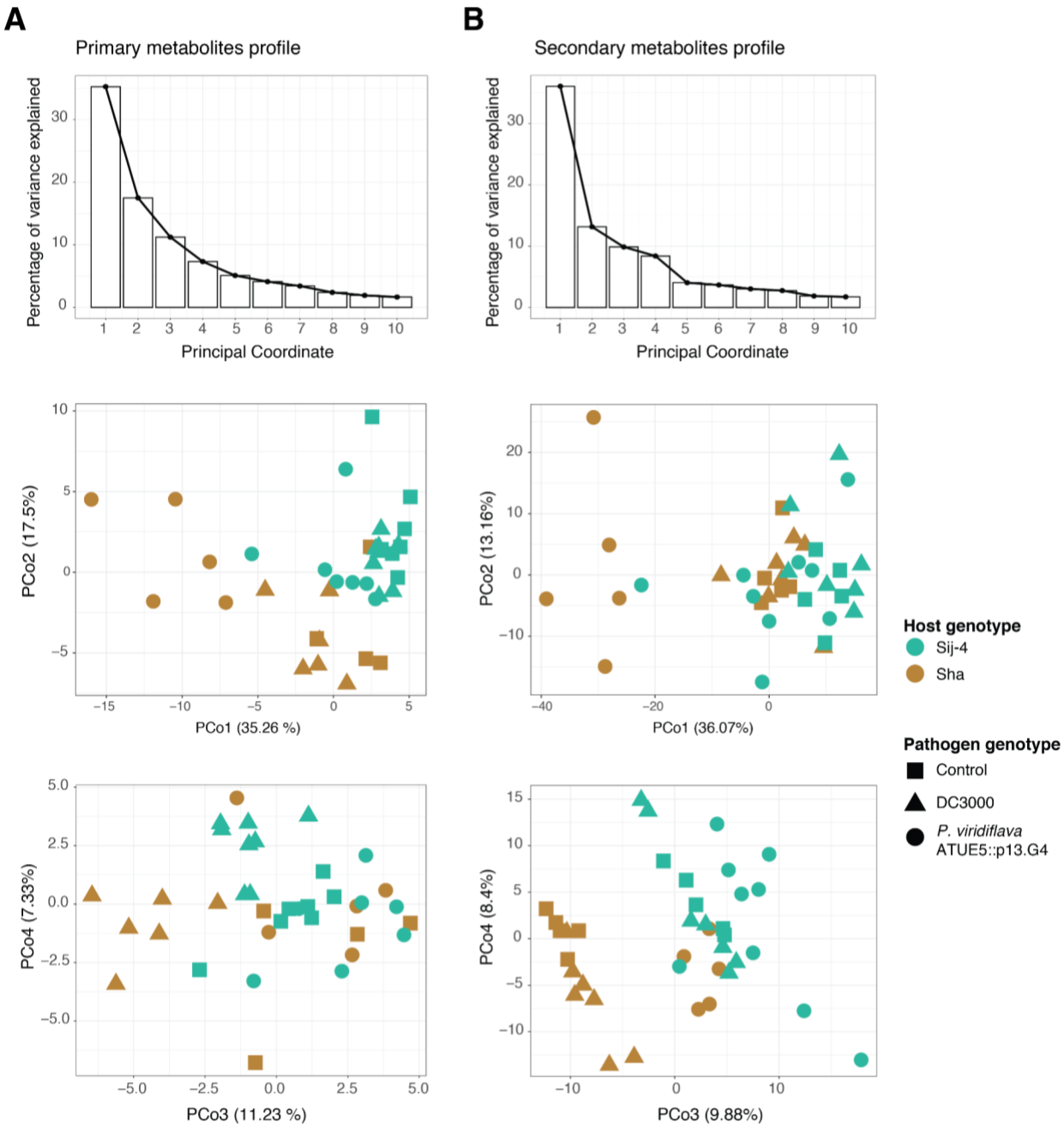

Figure S10 - Principal coordinate analysis of primary and secondary metabolites profile of susceptible and resistant *A. thaliana* genotypes infected with *P. viridiflava* ATUE5 and DC3000. (A) Principal coordinate analysis of the Euclidean distance based on 72 primary metabolites identified in Sij-4 and Sha infected and control plants. n = 4 - 8 replicates per host x pathogen genotype combination. **(B)** Principal coordinate analysis of the Euclidean distance based on 537 secondary metabolites identified in Sha and Sij-4 infected and control plants. n = 5 - 9 replicates per host x pathogen genotype combination. Colors indicate host genotypes and shapes indicate pathogen genotype. Each replicate was a pool of 3 to 5 individual plants.

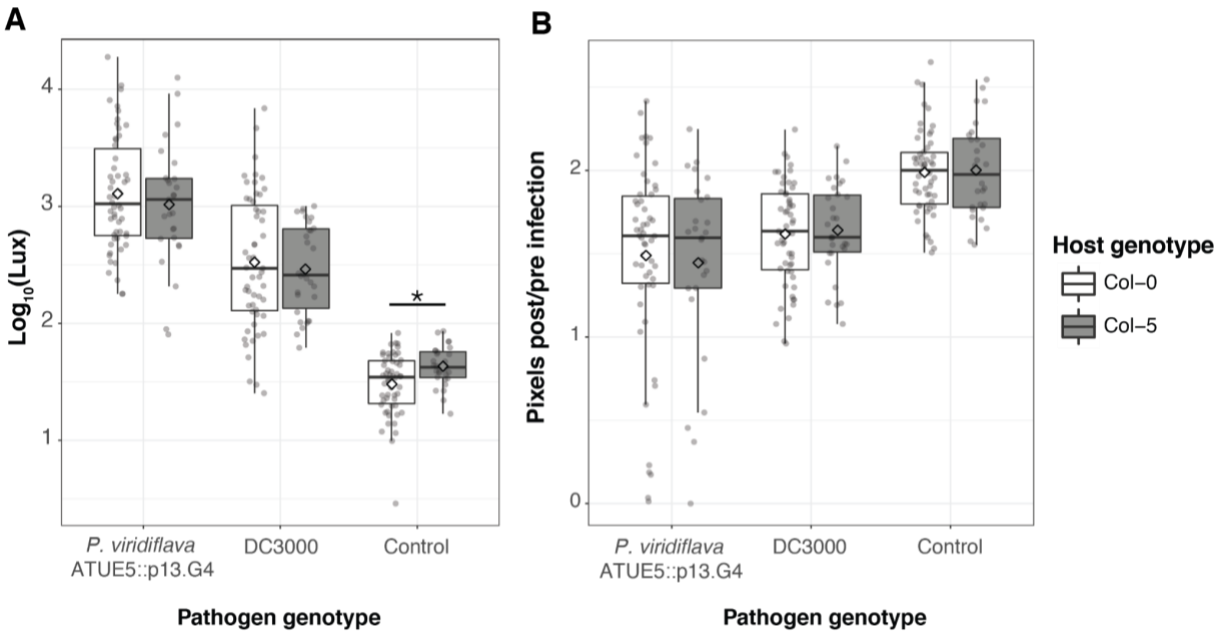

Figure S11 - Col-0 and Col-5 are equally susceptible to *P. viridiflava* ATUE5 and DC3000. (A) Bacterial load measured as luminescence 3 days post-infection with *P. viridiflava* ATUE5::p13.G4 and DC30000 and **(B)** plant growth measured as the ratio of green pixels 3 days post-infection in Col-0 WT (white) and Col-5 (grey) host genotypes infected with *P. viridiflava* ATUE5::p13.G4 or DC30000. Asterisks indicate statistical significance of the Wilcoxon test at adj. p-value < 0.05. p-value adjusted for multiple comparisons using the Benjamini-Hochberg method. n = 27 - 55 plants per host x pathogen genotype combination.

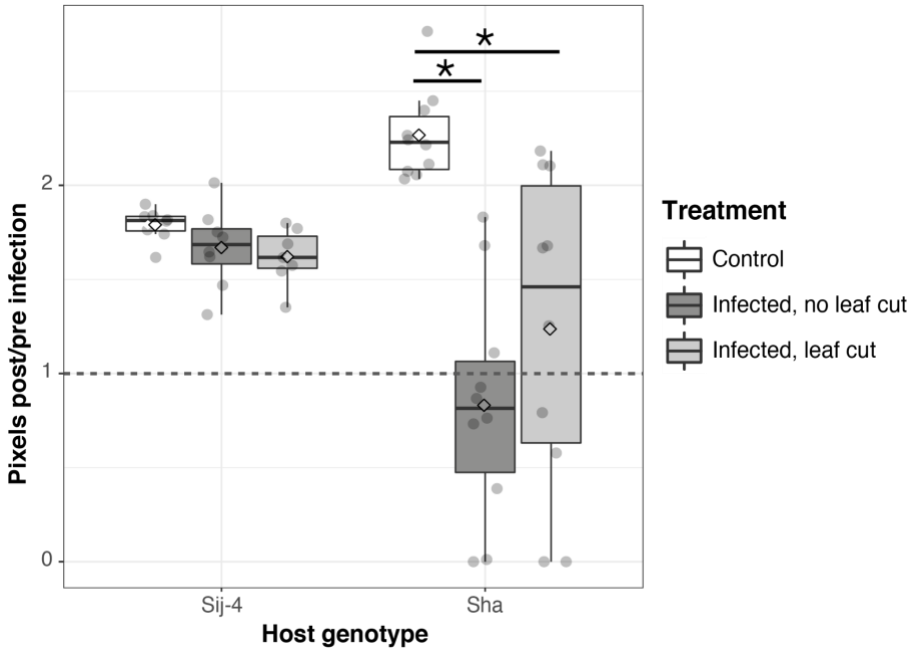

Figure S12 - Sha resistance to *P. viridiflava* increases after mechanical injury. Plant growth measured as the ratio of green pixels 3 days post infection to before infection of Sij-4 and Sha plants infected with *P. viridiflava* ATUE5::p13.G4 or mock. Infected plants had one leaf cut before infection (light gray) or not (dark gray). Asterisks indicate statistically significant differences (adj. p-value < 0.05) between treatments within each host genotype according to a two-tailed t-test. p-value adjusted for multiple comparisons using the Benjamini-Hochberg method. n = 7 - 10 plants per host x treatment combination.

Table S1. Natural A. thaliana accessions used in this study

1001ID	Accession	Country	Latitude	Longitude	80s Region from Cao et al. (1)	1001 cluster from the 1001 Genomes Consortium (2)
9939	Aitba-2	MAR	31,48	-7,45	N-Africa Spain	Relict
9982	Apost-1	ITA	39,01	16,47	Southern Italy	Italy Balkan Caucasus
9984	Ciste-2	ITA	41,62	12,87	Southern Italy	Admixed
6909	Col-0	USA	38,3	-92,3	ND	Germany
9994	Ey15-2	GER	48,4345	8,7678	Tuebingen	ND
9941	Fei-0	POR	40,92	-8,54	N-Africa Spain	Western Europe
9995	HKT2.4	GER	48,14	9,4	Tuebingen	Central Europe
9986	Jablo-1	BUL	41,59	25,2	East Europe	Italy Balkan Caucasus
10007	Koch-1	UKR	50,3553	29,3244	East Europe	ND
9964	Mammo-1	ITA	38,36	16,23	Southern Italy	Italy Balkan Caucasus
9966	Monte-1	ITA	40,28	15,65	Southern Italy	Italy Balkan Caucasus
9949	Qui-0	ESP	42,69	-6,93	N-Africa Spain	Western Europe
9976	Rovero-1	ITA	46,25	11,17	Southern Tyrol	Central Europe
10015	Sha	AFG	37,29	71,3	Central Asia	Asia
9958	Shigu-1	RUS	53,33	49,48	Russia	Asia
10010	Sij-4	UZB	41,45	70,05	Central Asia	Asia
9985	Slavi-1	BUL	41,43	23,65	East Europe	Italy Balkan Caucasus
9998	Star-8	GER	48,43	8,82	Tuebingen	Admixed
9940	Toufl-1	MAR	31,4687	-7,4166	N-Africa Spain	ND
10002	TueWa1-2	GER	48,53	9,04	Tuebingen	Western Europe
10011	Yeg-1	ARM	39,869	45,362	Kaukasus	Asia

ND: not determined

1. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet. 2011 Aug 28;43(10):956–63.
2. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016 Jul 14;166(2):481–91.

Table S2. *A. thaliana* immune mutants included in this study

Mutant	Gene	Function	Reference
sid2-2	*SALICYLIC ACID INDUCTION DEFICIENT 2*	Encodes isochorismate synthase 1, required for SA biosynthesis	(1)
eds1-12	*ENHANCED DISEASE SUSCEPTIBILITY 1*	Associates with PAD4 to promote SA biosynthesis and accumulation	(2)
ein2-1	*ETHYLENE INSENSITIVE 2*	Positively regulates ET response	(3)
coi1-16	*CORONATINE INSENSITIVE 1*	Part of the JA/COR receptor	(4)
jar1-1	*JASMONATE RESISTANT 1*	Conjugates JA and isoleucine to form biologically active JA-Ile	(5)
Col-5 (*gl1*)	*GLABRA 1*		

* Lacks trichomes on leaves and stems. Genetic background of *coi1-16*

1. Nawrath C, Métraux JP. Salicylic acid induction-deficient mutants of Arabidopsis express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. Plant Cell. 1999 Aug;11(8):1393–404.
2. Ordon J, Gantner J, Kemna J, Schwalgun L, Reschke M, Streubel J, et al. Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. Plant J. 2017 Jan;89(1):155–68.
3. Guzmán P, Ecker JR. Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. Plant Cell. 1990 Jun;2(6):513–23.
4. Ellis C, Turner JG. A conditionally fertile coi1 allele indicates cross-talk between plant hormone signalling pathways in Arabidopsis thaliana seeds and young seedlings. Planta. 2002 Aug;215(4):549–56.
5. Staswick PE, Su W, Howell SH. Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an Arabidopsis thaliana mutant. Proc Natl Acad Sci U S A. 1992 Aug 1;89(15):6837–40.

Table S3. Top 20 enriched GO biological processes based on differentially expressed genes 16 hpi in Sij-4 and Sha infected with *P. viridiflava*

Sij-4 (resistant)

Enrichment FDR	n Genes	Pathway Genes	Fold Enrichment	Pathway
4,22E-20	37	1230	7,1	Response to biotic stimulus
4,22E-20	37	1207	7,2	Response to external biotic stimulus
4,22E-20	37	1220	7,2	Biological process involved in interspecies interaction between organisms
4,22E-20	37	1207	7,2	Response to other organism
7,03E-17	38	1645	5,4	Response to external stimulus
2,44E-15	34	1416	5,7	Defense response
5,25E-15	21	418	11,9	Response to fungus
2,28E-14	28	967	6,8	Immune system process
1,26E-13	27	951	6,7	Defense response to other organism
1,54E-13	27	963	6,6	Immune response
2,92E-13	21	523	9,5	Response to bacterium
2,68E-09	12	198	14,3	Response to fatty acid
3,38E-09	16	447	8,4	Defense response to bacterium
5,58E-09	28	1646	4,0	Response to oxygen-containing compound
5,77E-09	14	331	10,0	Defense response to fungus
2,74E-08	11	194	13,4	Response to jasmonic acid
8,44E-08	11	217	12,0	Response to wounding
1,46E-06	12	358	7,9	Secondary metabolic process
1,46E-06	7	78	21,2	Indole-containing compound metabolic process
1,67E-06	5	24	49,1	Jasmonic acid biosynthetic process

Sha (susceptible)

Enrichment FDR	n Genes	Pathway Genes	Fold Enrichment	Pathway
1,60E-06	15	1230	5,75	Response to biotic stimulus
1,60E-06	15	1207	5,86	Response to external biotic stimulus
1,60E-06	15	1220	5,80	Biological process involved in interspecies interaction between organisms
1,60E-06	15	1207	5,86	Response to other organism
7,44E-06	16	1645	4,59	Response to external stimulus
7,44E-06	16	1634	4,62	Cellular response to chemical stimulus
8,17E-05	7	282	11,71	Response to toxic substance
8,17E-05	6	175	16,18	Cellular oxidant detoxification
8,53E-05	8	418	9,03	Response to fungus
8,53E-05	6	190	14,90	Cellular response to toxic substance
8,53E-05	6	189	14,98	Cellular detoxification
0,00013	3	16	88,47	Indole glucosinolate metabolic process
0,00014	7	331	9,98	Defense response to fungus
0,00022	5	135	17,47	Response to chitin

0,00022	6	241	11,75	Cellular response to decreased oxygen levels
0,00022	6	241	11,75	Cellular response to oxygen levels
0,00022	6	239	11,84	Cellular response to hypoxia
0,00022	6	244	11,60	Detoxification
0,00022	2	3	314,54	Sulfate reduction
0,00031	6	265	10,68	Response to hypoxia

Table S4. Enriched gene ontology (GO) categories between control Sha and Sij-4 at 0hpi

GO enriched in Sha (n = 270)				
Enrichment FDR	**n Genes**	**Pathway Genes**	**Fold Enrichment**	**Pathway**
0,00000265	5	8	70,14	Cellular response to iron ion starvation
0,00024	3	3	112,23	Positive regulation of iron ion transport
0,00024	11	192	6,43	Response to starvation
0,00024	7	62	12,67	Iron ion homeostasis
0,00024	6	39	17,27	Cellular iron ion homeostasis
0,00079	11	222	5,56	Response to nutrient levels
0,00202	6	60	11,22	Cellular transition metal ion homeostasis
0,00254	11	263	4,69	Response to extracellular stimulus
0,00254	8	133	6,75	Cellular cation homeostasis
0,00254	7	98	8,02	Transition metal ion homeostasis
0,00311	3	8	42,09	Regulation of iron ion transport
0,00322	8	144	6,23	Cellular ion homeostasis
0,00322	7	105	7,48	Cellular metal ion homeostasis
0,00364	3	9	37,41	Positive regulation of ion transport
0,0047	8	155	5,79	Cellular response to starvation
0,0055	8	160	5,61	Metal ion homeostasis
0,00632	5	54	10,39	Iron ion transport
0,00705	8	170	5,28	Cellular response to nutrient levels
0,00705	3	12	28,06	Response to iron ion starvation
0,00755	8	173	5,19	Cellular chemical homeostasis
GO enriched in Sij-4 (n = 350)				
Enrichment FDR	**n Genes**	**Pathway Genes**	**Fold Enrichment**	**Pathway**
0,00108	40	1416	2,28	Defense response

Table S5. Top 10% secondary metabolites with the largest Spearman correlation with principal coordinate 1

Metabolite	Spearman's rho	p adj.
Cer t18:0/c22:0	-0,891	0,00E+00
Cer t18:0/c24:0	-0,841	3,96E-08
Gluconic acid	-0,834	1,55E-07
Cer t18:0/c16:0	-0,827	2,66E-07
Cer t18:1/c16:0	-0,827	2,66E-07
Cer t18:0/h22:0	-0,826	2,87E-07
Citric acid	-0,822	3,46E-07
Citramalic acid	-0,820	3,95E-07
Lyso-DGDG 16:0 (2)	-0,819	5,86E-09
Adenosine 2',3'-cyclic monophosphate	-0,817	4,54E-07
DAG 34:4 (3)	-0,812	9,64E-09
Serotonine	-0,810	1,09E-08
Cer t18:1/c22:0	-0,795	9,24E-07
Gluconic acid lactone	-0,791	1,01E-06
2 deoxyadenosine	-0,785	7,05E-08
FA 18:3	-0,780	1,33E-06
3-Deoxy-D-manno-2-octulosonic acid	-0,780	1,33E-06
Methionine sulfoxide	-0,779	1,33E-06
Thymidine	-0,779	1,33E-06
Cer t18:1/h22:0	-0,775	1,45E-06
TAG 54:5	-0,773	1,47E-06
Guanosine	-0,770	1,57E-06
FA 18:2	-0,766	1,67E-06
TAG 50:5 (1)	-0,763	1,80E-06
Glucoraphanin	0,766	1,67E-06

p adj.: p-value adjusted for multiple comparisons using the Benjamini-Hochberg method.

Chapter 2: The specialized-metabolite potential of phyllosphere *Pseudomonas* and its link to virulence

ABSTRACT

The bacterium *Pseudomonas viridiflava* is an agricultural pest and a natural pathogen of the model plant *Arabidopsis thaliana.* It is globally distributed, having been isolated from plants in Europe and the USA, as well as from environmental sources. *P. viridiflava* isolates are pathogenic on *A. thaliana* under lab conditions, even though they encode only one effector, AvrE, and none of the toxins associated with virulence in other *Pseudomonas* strains. Moreover, closely-related isolates vary in their virulence in the same host, so additional unidentified virulence factors must be contributing to this. Specialized metabolites, which facilitate the interaction between an organism and its environment, can be virulence factors. Here, we predicted the specialized metabolite potential of 284 *Pseudomonas* genomes from the *A. thaliana* phyllosphere, most of which were *P. viridiflava*. Combined, these genomes encoded more than 3000 putative specialized metabolites from 22 different classes, which clustered into 386 unique families. Only 4% of these families had an experimentally-characterized reference from the MIBiG database. Non-ribosomal peptide synthetases (NRPS) were the most common class of specialized metabolites and were the only class detected in all genomes. The diversity of terpenes was larger than described before for *Pseudomonas*, 74% of the genomes analyzed encoded at least one of 19 terpene families. We correlated the presence/absence of 105 NRPS from 75 *P. viridiflava* isolates with their effect on plant size after infection, and identified putative NRPSs involved in virulence. Our work is one of the first to describe the specialized metabolite potential of phyllosphere *Pseudomonas*, a genus known for its metabolic diversity. We identified NRPSs that could contribute to the differences in virulence of closely-related *P. viridiflava* isolates and that await for experimental validation.

AUTHOR CONTRIBUTIONS
Authors are listed in the order to be used for publication.

1. Alejandra Duque-Jaramillo, Max Planck Institute for Biology Tübingen, Germany.
   Devised the study, predicted biosynthetic gene clusters and gene cluster families, analyzed all the data, wrote the manuscript.
2. Haim Ashkenazy, Max Planck Institute for Biology Tübingen, Germany.
   Selected the representative *Pseudomonas* genomes
3. Talia L. Karasov[1,2], Max Planck Institute for Biology Tübingen, Germany; School of Biological Sciences, University of Utah, USA.
   Devised the study, performed the axenic infections, provided input for manuscript.
4. Detlef Weigel, Max Planck Institute for Biology Tübingen, Germany.
   Provided input for the manuscript. Corresponding author.

STATUS IN PUBLICATION PROCESS
Advanced manuscript, awaiting for submission to target journal.

INTRODUCTION

Bacteria produce a wide range of specialized metabolites, also known as secondary metabolites, which mediate their interactions with other organisms and the environment. Specialized metabolites are defined as those metabolites not essential for basal bacterial growth or division, but instead play a role in nutrient acquisition, quorum sensing, defense and virulence (1,2). Some specialized metabolites from bacteria also have medical relevance for humans, serving as antibiotic, antifungal, anthelminthic, immunosuppressant and cholesterol-lowering agents (3).

Specialized metabolites also mediate the interaction between bacteria and plants (4). Plant growth and defense against pathogenic bacteria are commonly affected by these metabolites, although the mechanisms are not yet well understood. Examples include lipopeptides produced by *Streptomyces*, *Pseudomonas*, and *Bacillus* genera that can act as biosurfactants and have antimicrobial and cytotoxic activity (5). The biosurfactant activity contributes to increase the availability of water-insoluble substrates and to promote swarming (6). There are also lipopeptides, such as syringomycin and syringopeptin, that are produced by plant-pathogenic *Pseudomonas* and contribute to their virulence; non-pathogenic bacteria's lipopeptides, on the other hand, can induce resistance against microbial pathogens in the host (5). Moreover, several plant-associated bacteria, including *Pseudomonas*, can produce indole-3-acetic acid (auxin), a phytohormone, and use it to manipulate host development as well as defense response and systemic acquired resistance (7–9).

The *Pseudomonas* genus is well-known for its environmental ubiquity and its ability to produce a vast array of specialized metabolites (6,10). It encompasses free-living species as well as non-pathogenic and pathogenic species that are able to infect not only plants but also humans and other animals (11). There are three main *Pseudomonas* lineages: *P. aeruginosa*, *P. fluorescens* and *P. pertucinogena*, each containing several phylogenetic groups (12,13). In particular, the *P. fluorescens* lineage comprises 5 groups, including the *P. syringae* and *P. fluorescens* groups (12,14). Members of *Pseudomonas* are metabolically diverse and can produce an extensive collection of specialized metabolites, such as siderophores, lipopeptides, terpenes, polyketides and non-ribosomal peptides (6,7,15). Some of these specialized metabolites contribute to niche adaptation and most are produced only by specific lineages (6).

Specialized metabolites are produced by biosynthetic gene clusters (BGCs), a group of two or more genes that encode a biosynthetic pathway for the production of a specialized metabolite and its variants (16). The BGC repertoire of the genus *Pseudomonas* has previously been investigated: a total of 24 major BGCs classes were predicted in 37 *Pseudomonas* reference genomes (17), the most abundant classes being non-ribosomal peptide synthetases (NRPSs), ribosomally synthesized and post-translationally modified peptides (RiPPs), redox-cofactors and N-acetylglutaminylglutamine amide (NAGGN)(17). Similarly, 30 BGC classes were

detected in *Pseudomonas* type strains (10). From these studies, we learned that *Pseudomonas* genomes encode on average 10 BGCs, ranging from 1 to 23 (17); and rarely any genome has more than 15 BGCs (7). There is extensive diversity in the metabolite potential at the intra-species level as well: *P. fluorescens* isolates have between 7 and 18 BGCs, 13 on average (17,18).

In this study, we investigated the *P. fluorescens* lineage, with a specific focus on *Pseudomonas viridiflava*. This species is particularly interesting because it is an agricultural pest and a pathogen of *Arabidopsis thaliana*, but the mechanisms of virulence are largely unknown. We have shown that *P. viridiflava* lacks most of the known virulence effectors, but at the same time observed high virulence on *A. thaliana* under laboratory conditions (19), which could be due to the production of specialized metabolites. *Pseudomonas viridiflava* is a globally-distributed opportunistic plant pathogen (20–22). It belongs to the *P. syringae* species complex from the *P. fluorescens* lineage and has been isolated from wild populations in the midwestern USA, Japan and across Europe (19,23,24). In addition to plants, *P. viridiflava* is found in a diverse range of environments, such as leaf litter, rain, snow and river water (25). Several virulence factors have been described in *P. viridiflava*, including pectate lyase enzymes that can degrade the host's cell wall, and the presence of the effector AvrE, which promotes an aqueous apoplast suitable for bacterial growth (26–28). In *A. thaliana* populations of Southwest Germany, *P. viridiflava* is prevalent in the epi- and endophytic leaf compartments (19). Most of these isolates are pathogenic on *A. thaliana* under lab conditions, with inter-strain variation in virulence (19,21,29). The genomes of these *P. viridiflava* isolates contain AvrE and the pectate lyase gene *pel*, but lack other effectors and the genes required for the biosynthesis of the known toxins coronatine, syringomycin, syringopeptide, mangotoxin, phaseolotoxin and tabtoxin are absent (19). Even though virulence factors AvrE and *pel* are conserved, the varying degree of virulence of these isolates suggests that other factors are involved in the process of infection and virulence. While it is not yet clear what factors determine the extent of *P. viridiflava* virulence on *A. thaliana*, specialized metabolites may play a major role. We therefore set out to study the specialized metabolite potential of *P. viridiflava*.

Here, we described the specialized metabolite potential of 284 *Pseudomonas* isolates from Southwestern Germany, with an emphasis on *P. viridiflava*. By integrating the pattern of BGC presence/absence with virulence data on *A. thaliana*, we identified putative BGCs correlated with the stronger impact on plant size. In addition, we identified previously unknown metabolites: an uncharacterized tetrapeptide and a variant of the lipopeptide cichofactin. The specialized metabolites we identified are strong candidates for the genetic basis differences in virulence of closely related *P. viridiflava* isolates.

RESULTS

Representative genomes for a collection of plant-associated *Pseudomonas*

Karasov and colleagues have described a collection of 1524 *Pseudomonas* isolated from wild *A. thaliana* populations in Southwest Germany (19). These isolates were classified into 11 operational taxonomic units (OTUs) based on sequence similarity in the V3-V4 region of the 16S rRNA gene, using a 99% similarity cut-off. OTU5 was the most abundant, and it was classified as *P. viridiflava* (19).

We selected representative genomes from this collection based on their gene content. Starting with 1498 complete genome assemblies (BUSCO single genes > 95%), we reduced our dataset to 284 genomes with distinct gene content according to the presence and absence of groups of orthologous genes. These genomes encompassed all the OTUs identified by Karasov et al. (19), referred to as ATUE here and in other work (30,31) and referred to herein as 'clades'. The use of representative genomes resulted in a more balanced representation of these clades (Table 1): in the original work, 88.91% of the genomes corresponded to ATUE5, which was reduced to 57.39% in the present study. Concurrently, the representation of other clades was increased, particularly of ATUE2 and ATUE3. Four of the clades, ATUE8 to ATUE11 contained only a single representative genome.

Table 1. Distribution of isolates in the Karasov et al. (19) **collection and in the selected representatives genomes**

Clade	All		Representative		Change in %	*Pseudomonas* group/subgroup
	n	%	n	%		
ATUE1	15	0.98	8	2.82	1.83	*P. fluorescens/P. mandelii*
ATUE2	67	4.40	49	17.25	12.86	*P. fluorescens*
ATUE3	37	2.43	28	9.86	7.43	*P. syringae*
ATUE4	27	1.77	19	6.69	4.92	*P. fluorescens*
ATUE5	1355	88.91	163	57.39	-31.52	*P. syringae*
ATUE6	12	0.79	9	3.17	2.38	*P. rhizosphaerae*
ATUE7	4	0.26	4	1.41	1.15	*P. syringae*
ATUE8	3	0.20	1	0.35	0.16	*P. fluorescens*
ATUE9	2	0.13	1	0.35	0.22	*P. syringae*
ATUE10	1	0.07	1	0.35	0.29	*P. fluorescens/P. corrugata*
ATUE11	1	0.07	1	0.35	0.29	*P. syringae*
Total	1524	100.00	284	100.00		

Several *Pseudomonas* species co-exist in the leaves of *Arabidopsis thaliana*

Pseudomonas species can occupy a wide variety of ecological niches, and they can be pathogenic, like many strains from the *P. syringae* complex, or commensals, like *P. fluorescens.* Certain specialized metabolites might be more useful in one niche or host compared to the other, so it is important to identify the *Pseudomonas* in our representative genomes. To identify the phylogenetic relationship of our representative isolates, we generated a maximum likelihood tree based on the *gyrB* (DNA gyrase beta subunit) gene sequence extracted from the genome assemblies, together with that of 105 reference *Pseudomonas* strains (Table S1). This gene provides high resolution at the intraspecies level for phylogenetic analyses (32). As the taxonomy of *Pseudomonas* is complex, we report here the broad-sense *Pseudomonas* groups (11,14) for each ATUE clade; a group aggregates several species and can be further divided into subgroups.

We found ATUE3 and ATUE5 clades closely related to the *P. syringae* group, and ATUE1, ATUE2, ATUE8 clades to the *P. fluorescens* group. No closely-related reference was identified for the ATUE6 and ATUE7 clades. ATUE5 clustered with *P. viridiflava* genomes, in accordance with Karasov et al. (19)(Figure 1). We confirmed and refined the taxonomic classification of each representative genome using GTDB, the genome taxonomy database, which uses 120 universal bacterial genes (33,34). With this approach, ATUE6 genomes belonged to the *P. rhizosphaerae* group, ATUE10 to the *P. corrugata* group and ATUE7, 9 and 11, to the *P. syringae* group (Table 1). The GTBD species classification for each representative genome can be found in table S2.

No clade was monophyletic except for ATUE5, as there was more than one *Pseudomonas* species in all but ATUE5 clades (Table S2). In particular, ATUE2 genomes were assigned to 14 different species, each with 1 to 8 members. ATUE3 isolates belonged to 7 different species, the main one being *P. avellaneda* with 15 representatives. ATUE1 and ATUE6 contained 4 species while ATUE4 contained 3. Interestingly, the only clade with more than one genome and only one species was ATUE5 (Table S2). Only for one isolate, belonging to ATUE1, no match at the species level was obtained with the GTDB.

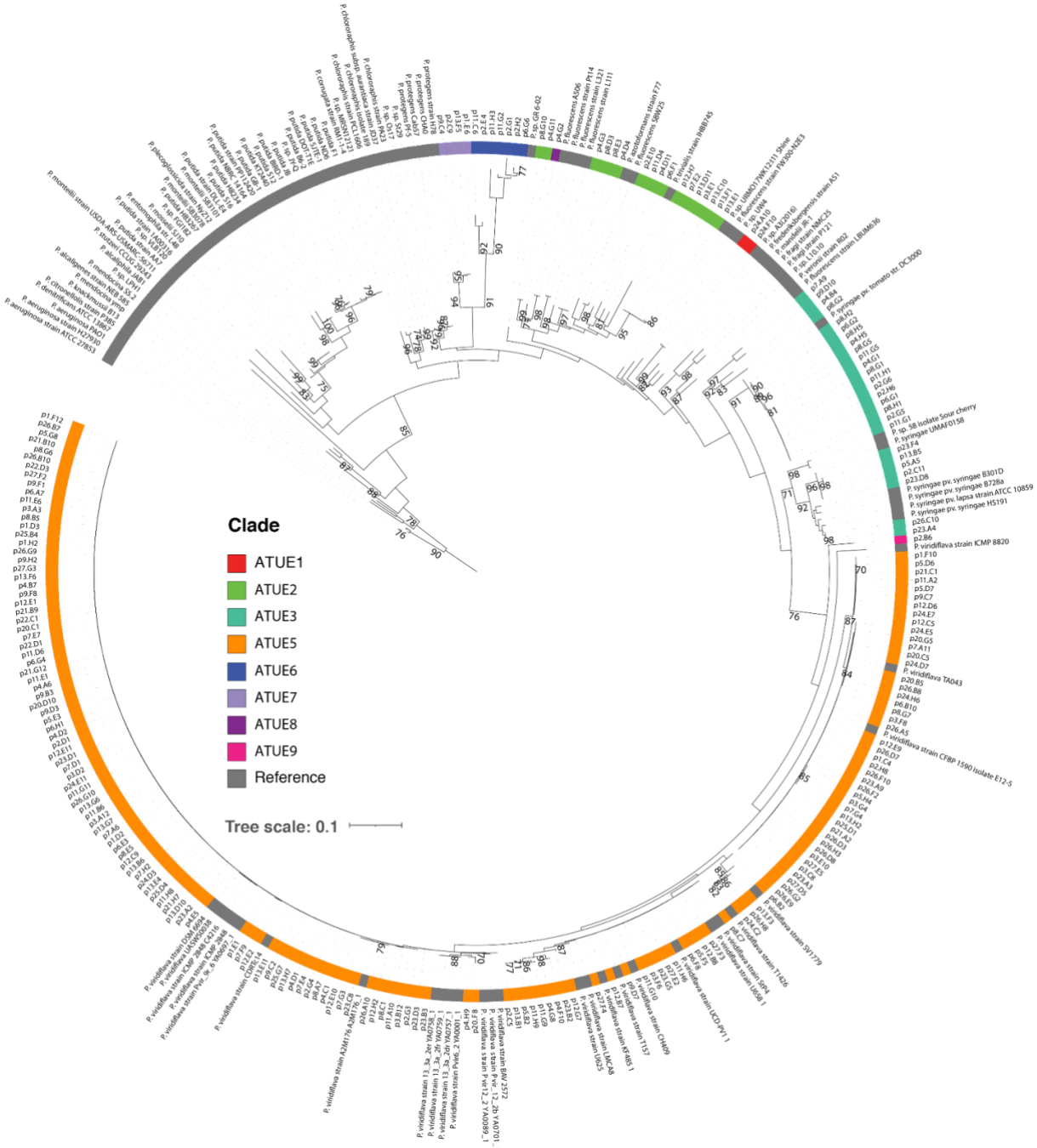

Figure 1 - Phylogenetic relationship between the selected representative genomes and reference *Pseudomonas*. Maximum likelihood tree based on the *gyrB* gene sequence of reference and representative *Pseudomonas* genomes. The tree was rooted on *P. aeruginosa.* Only bootstrap values above 70 are shown. The color circle indicates the clade of our representative genomes, gray indicates the reference genomes obtained from the NCBI. *gyrB* gene sequence extraction was not successful for ATUE4, 10 and 11 and thus are not displayed.

The biosynthetic potential of *A. thaliana*-associated *Pseudomonas* is broad and its products are largely unknown

To characterize the specialized metabolite potential of *A. thaliana*-associated *Pseudomonas*, we used antiSMASH (35) to predict BGCs in the representative genomes (19). It is possible that two or more BGC overlap in one genomic region; when this occurred, we counted each BGC independently.

In total, 3,037 BGCs from 22 classes were predicted. The average genome had 10.69 BGCs (median = 9, min = 3, max = 25; Figure 2A). The distribution of BGCs varied by taxonomic clade: ATUE2 encoded the largest number of BGCs, 13 on average, and ATUE6 the smallest, 5.6 on average; the remaining clades had between 8.3 and 11 BGCs on average, except for the single-genome ATUE8 that had 15 (Figure 2B, Table S3). Three ATUE3 isolates had a higher number of BGCs (>22) compared to the other genomes of this clade. Two of these genomes were classified as *P. syringae* and the other as *P. congelans*. Four other genomes in ATUE3 were classified as *P. congelans*, harboring 9 to 11 BGCs, but none else was identified as *P. syringae*.

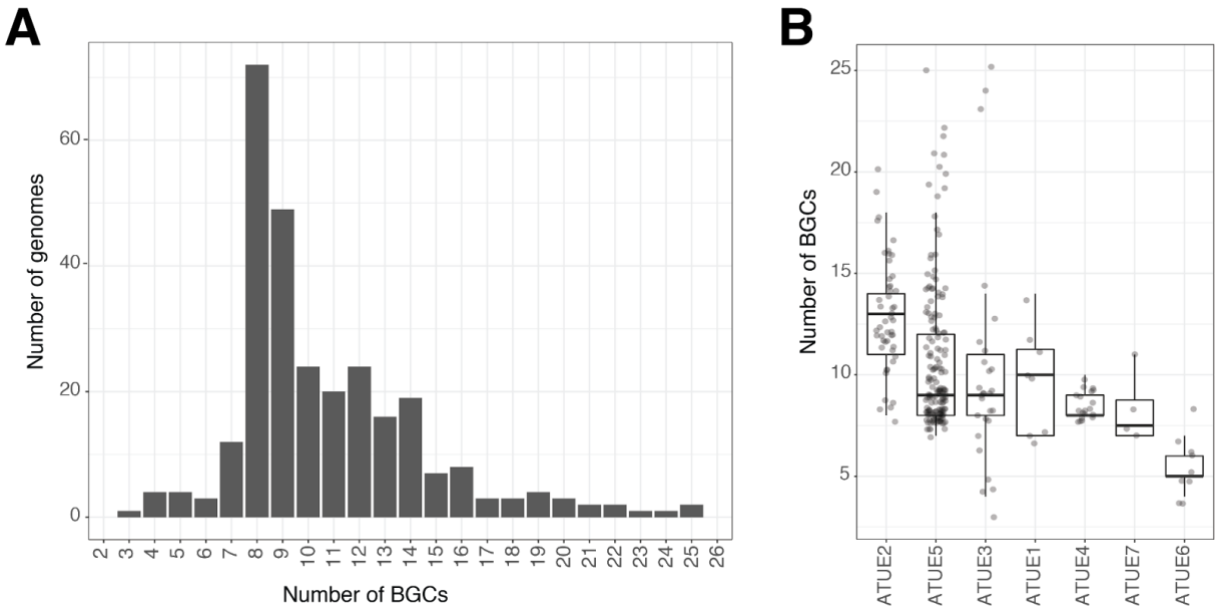

Figure 2 - There are differences in the BGC distribution among clades. (A) Distribution of the number of BGCs across all genomes. **(B)** Number of BGCs by ATUE clade. The boxes represent the first and third quartile of the data and the vertical line, the median. Whiskers below and above the boxes extend to the minimum and maximum value, respectively. Only ATUEs with more than one genome are shown in B. BGCs = biosynthetic gene clusters.

The most abundant BGCs classes were the NRPS and NRPS-like, accounting for 56% of all the predicted BGCs. The dipeptide N-acetylglutaminylglutamine amide (NAGGN), aryl polyenes, terpenes, siderophores and bacteriocins were common as well, with more than 150 BGCs predicted across all genomes (Table 2). On the other hand, three or less BGCs were found for TfuA-related RiPPs, lanthipeptide, lasso peptide, trans-AT PKS and acyl amino acids. There was a significant phylogenetic signal in the total number of BGCs, as well as in the number of bacteriocins, ectoine, NRPS, phenazine, resorcinol and thiopeptide BGCs (Moran's I < 0.05).

Table 2. Number of predicted biosynthetic gene clusters per class

BGC class	n	%	average n per genome (min - max)
NRPS	1379	45.41	4.86 (1 - 18)
NRPS-like	322	10.60	1.13 (0 - 4)
NAGGN	279	9.19	0.98 (0 -1)
Aryl polyene	264	8.69	0.93 (0 -2)
Terpene	223	7.34	0.78 (0 -3)
Siderophore	215	7.08	0.76 (0 -3)
Bacteriocin	170	5.60	0.60 (0 - 3)
Betalactone	73	2.40	0.26 (0 -2)
Type 1 PKS	24	0.79	0.08 (0 -1)
Thiopeptide	23	0.76	0.08 (0 - 2)
Butyrolactone	15	0.49	0.05 (0 - 2)
Hserlactone	9	0.30	0.03 (0 -2)
Ectoine	8	0.26	0.03 (0 - 1)
LAP	6	0.20	0.02 (0 -1)
Phenazine	6	0.20	0.02 (0 -1)
Resorcinol	6	0.20	0.03 (0 -1)
CDPS	5	0.16	0.03 (0 -1)
TfuA-related RiPPs	3	0.10	0.01 (0 -1)
Lanthipeptide	2	0.07	0.007 (0 -1)
Lassopeptide	2	0.07	0.007 (0 -1)
Trans-AT PKS	2	0.07	0.007 (0 -1)
Acyl amino acids	1	0.03	0.003 (0 -1)
Total	3037	100	10.69 (3 - 25

BGC: biosynthetic gene cluster, CDPS: tRNA-dependent cyclodipeptide synthases, hserlactone: homoserine lactone, LAP: Linear azol(in)e-containing peptide, NAGGN: N-acetylglutaminylglutamine amide, NRPS: non-ribosomal peptide synthetases, PKS: polyketide synthases.

All genomes encoded at least one NRPS, with 4.86 on average and 18 maximum per genome; this was the only BGC class present in all genomes (Figure 3, Table S2). NRPS-like BGCs were the next class by prevalence, with an average of 1.13 per genome (min = 0 , max = 4). Many BGC classes had a binary distribution, being either present once or absent in a given genome; this was the case for acyl amino acids, butyrolactone, CDPS, ectoine, lanthipeptide, LAP, lassopeptide, NAGGN, phenazine, resorcinol, type 1 PKS, TfuA-related RiPPs and transAT-PKS. There were maximum 2 aryl polyene, betalactone, homoserine lactone and thiopeptide BGCs, and 3 bacteriocin, terpene and siderophore BGCs predicted per genome.

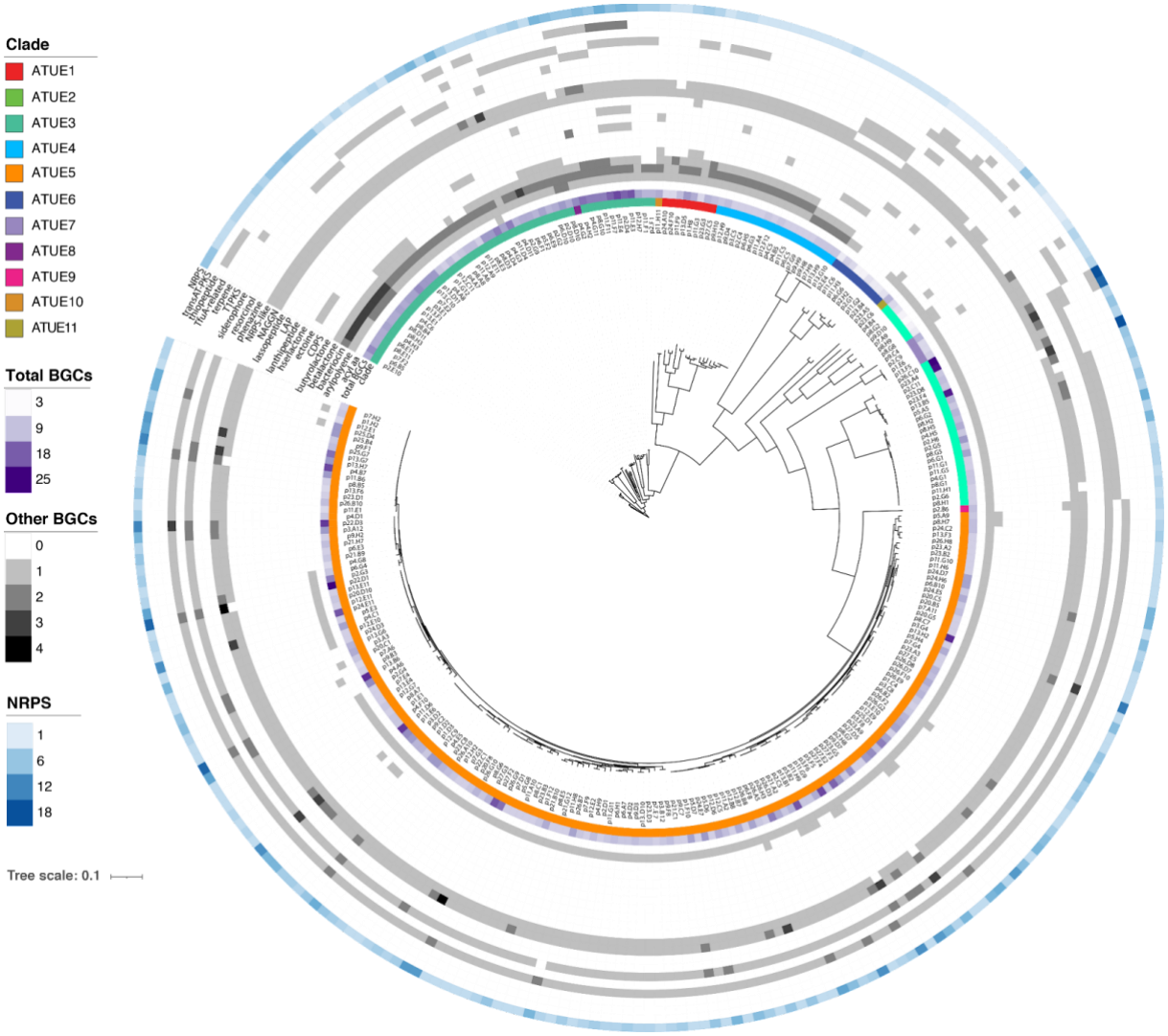

Figure 3 - Distribution of BGCs in plant-associated *Pseudomonas*. Maximum likelihood phylogenetic tree of 284 genomes from (19) annotated with clade (first circle inside), total number of BGCS (second circle, purple), number of BGCs per class (circles 3 to 23, gray scale), and number of NRPS BGCs (last circle, blue). aa: amino acids, BGC: biosynthetic gene cluster, CDPS: tRNA-dependent cyclodipeptide synthases, hserlactone: homoserine lactone, LAP: Linear azol(in)e-containing peptide, NAGGN: N-acetylglutaminylglutamine amide, NRPS: non-ribosomal peptide synthetases, PKS: polyketide synthases.

The annotation of BGCs does not inform us about the uniqueness or conservation of each BGC, which is important for our aim to identify putative BGCs involved in virulence. To address this, we clustered the predicted BGCs into families based on their similarity using BiG-SCAPE (36). BiG-SCAPE generates sequence similarity networks of the BGC and groups them into gene cluster families.

The BGCs were clustered into 386 unique families spanning 7 classes; we did not detect BGCs classified as saccharides. Sixteen families were included in more than one class and were removed for further analyses, except when the product was a PKS-NRPS hybrid. For these families, we kept only the annotation of PKS-NRPS hybrids

and discarded the separate PKS and NRPS ones. In total, 374 gene cluster families were included (Table 3).

Table 3. Number of predicted gene cluster families

Family Class	n	%
NRPS	279	74.60
Others	51	13.64
RiPPs	22	5.88
Terpene	16	4.28
PKS other	1	0.27
PKS-NRP hybrids	4	1.07
PKS type I	1	0.27
Total	374	100.00

NRPS: non-ribosomal peptide synthetases, PKS: polyketide synthases, RiPPs: Ribosomally synthesized and post-translationally modified peptides.

Genomes had on average 9.96 gene cluster families (median = 9, min = 3, max = 24; Figure 4A). This distribution was very similar to that of BGCs. Indeed, we found a strong positive correlation between the number of BGCs and the number of gene cluster families in a genome (Pearson's correlation coefficient = 0.97, p-value < 2.2e-16; Figure 4B). This suggests that there is no redundancy in BGCs, as every BGC in a genome belongs to a different gene cluster family.

We then assessed the conservation of gene cluster families across the *Pseudomonas* isolates, that is, in how many genomes a given family is found. On average, a gene cluster family was present in 7.57 genomes, i.e. 2.66% of the genomes analyzed, ranging from 1 to 129 (median = 2; Figure 4C). Three families were present in at least one-third of the genomes: one was an NRPS (129 isolates) and the remaining two belong to the class 'other' (100 and 92 isolates), encoding a siderophore and an aryl polyene, respectively.

Terpenes were the fifth most common BGC type. However, very few terpenes have been described in *Pseudomonas* before (7,17). We found 19 gene cluster families classified as terpenes (16 when families assigned to two or more classes were removed), distributed in 210 (73.94%) isolates (Figure S1). Most isolates encoded only one terpene gene cluster family, and ATUE10 alone had no gene cluster families of this class. All isolates of clades ATUE5 to ATUE 9 and ATUE11 encoded at least one terpene family; while only 1 ATUE3 (3.6%) and 2 (10.5%) ATUE4 isolates had one. Within a clade, there could be more than one terpene family: ATUE5 had 3, while ATUE1 and ATUE6 had 2 families. The biggest terpene diversity was found in ATUE2,

even when only 50% of the isolates had a terpene: 7 families were detected, 4 of them found in only 1 or 2 genomes.

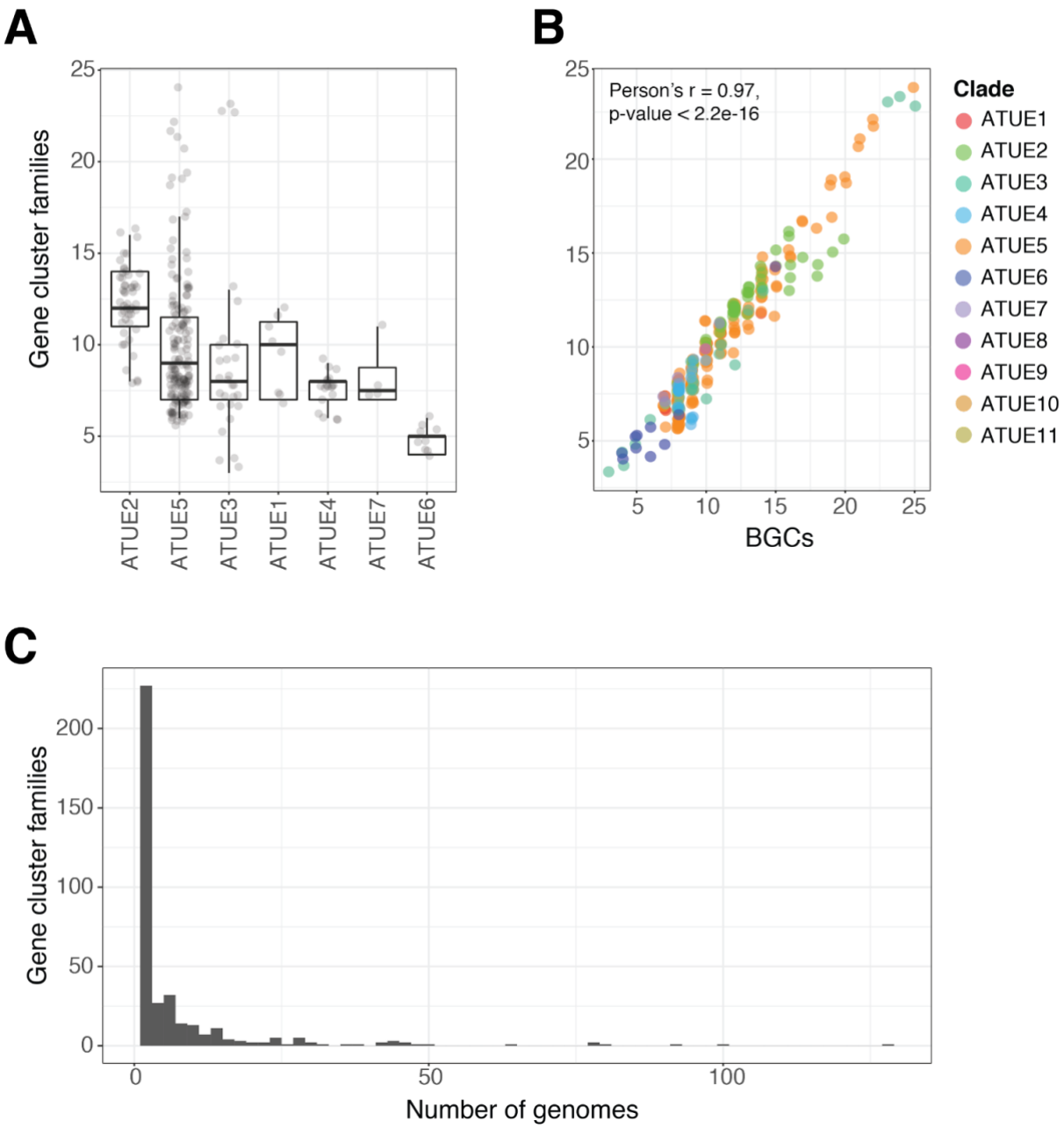

Figure 4 - The distribution of gene cluster families is similar to that of BGCs. (A) Number of gene cluster families per clade. The boxes represent the first and third quartile of the data and the vertical line, the median. Whiskers below and above the boxes extend to the minimum and maximum value, respectively. **(B)** Number of gene cluster families vs. number of BGCs in each genome. There was a strong positive correlation between both measurements (Pearson's correlation coefficient = 0.97, p-value < 2.2e-16). (**C)** Number of genomes in which a gene cluster family is present. BGC = biosynthetic gene cluster.

Of the 386 gene cluster families identified, only 13 (3.9%) included experimentally characterized BGCs according to the MIBiG database. Obafluorin, poaeamide B, massetolide A, viscosin and sessilin/tolaasin were present only in ATUE2 isolates; syringolin A, syringopeptin/syringomycin and syringofactin in ATUE3 isolates, arthrofactin/anikasin in ATUE1, pseudomonine in ATUE8, and poaeamide and cichofactin in ATUE5. The syringofactin gene cluster family was also encoded by isolates from ATUE5 and ATUE11. Finally, the family of the siderophore pyoverdine was found mostly in isolates from ATUE2, and one genome of ATUE1, 3, 8 and 10 each. These results indicate that most of the specialized metabolites potentially produced by *Pseudomonas*, only a few have been characterized and their function is known.

The presence/absence of biosynthetic gene cluster correlates with virulence on *A. thaliana*

The presence or absence of specific gene cluster families could contribute to virulence of a bacterial isolate on its host plant. We tested this hypothesis by reanalyzing an experimental dataset we had previously generated, where *A. thaliana* Ey15-2 plants were infected with 75 *P. viridiflava* ATUE5 isolates, 13 of which were present in the representative genomes described above, and plant size was measured after infection as a proxy for virulence (T. Karasov, personal communication; Table S4). We correlated the presence of gene cluster families in these isolates with plant size after infection.

We identified 105 NRPS, 4 terpenes, 1 RiPP and 30 gene cluster families classified as ‘other’ in these 75 ATUE5 isolates. Focusing on the NRPS, we found 8 families correlated with the size of infected host plants, measured as green pixels seven days post-infection (p-value < 0.05; Table 4). Only one family, FAM_02554, was negatively correlated with plant size, meaning isolates encoding this gene cluster family were more virulent on *A. thaliana* (Figure 5). This family was present in 24 (32%) isolates, and it was somewhat related to siderophore pyoverdine according to the MIBiG database. Most of the families were encoded by only a few isolates; after FAM_02554, FAM_02006 was the second most prevalent family, encoded by 18 genomes, followed by FAM_02516 and FAM_1848, encoded by 10 and 7 genomes respectively. FAM_02006 has a high similarity to rhizomide, while FAM_02516 and FAM_1848 had similarity to cichofactin and the siderophore crochelin A, respectively.

Table 4. NRPS families correlated with plant size after infection

NRPS Family	Number of isolates	Spearman's rho
FAM_02470	2	0.26
FAM_02353	1	0.18
FAM_02468	4	0.07
FAM_02357	1	0.18
FAM_02358	1	0.18
FAM_02554	24	-0.28
FAM_01848	7	0.14
FAM_02516	10	0.17
FAM_02006	18	0.24

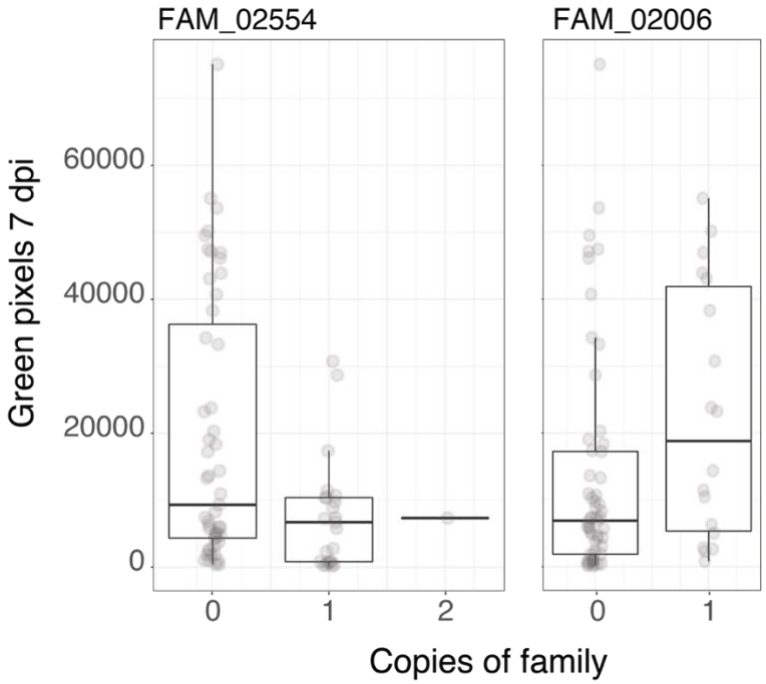

Figure 5 - Gene cluster families presence/absence could impact pathogen virulence. Number of green pixels of *A. thaliana* Ey15-2 plants infected with 75 *P. viridiflava* ATUE5 isolates. Green pixels were measured at 7 dpi, and are shown for two NRPS families. The boxes represent the first and third quartile of the data and the vertical line, the median. Whiskers below and above the boxes extend to the minimum and maximum value, respectively. dpi: days post-infection.

We determined the prevalence of families FAM_02554 and FAM_02006 in the 284 representative *Pseudomonas*, and found they were encoded by 21 to 24 isolates, most of which belonged to ATUE5, with small contributions from ATUE3 and ATUE9. Taking into account that both ATUE3 and ATUE9 isolates belong to the *P. syringae* group, known for its phytopathogenic strains, the specialized metabolites encoded by this gene cluster families could also have a role in their virulence.

Visual inspection of the structure of the BGCs encoded by these 75 *P. viridiflava* ATUE5 isolates revealed two lipopeptide clusters, a tetra and an octapeptide, for which we could not predict the metabolite produced. Further investigation showed that the octapeptide was a variant of the virulence factor cichofactin (described in Helmle et al., in preparation); whereas we were unable to characterize the tetrapeptide.

DISCUSSION

Specialized metabolites play a role in the interaction with other organisms and the environment, increasing the fitness of the producer organism (2). The *Pseudomonas* genus is known for its ability to encode and produce a large number of specialized metabolites (6,37), although the specific biosynthetic capabilities of leaf-associated bacteria has not been characterized.

Here, we describe the specialized metabolite potential of *Pseudomonas* isolated from wild *A. thaliana* plants in Southwest Germany, using 284 genomes representative of a larger collection described previously by Karasov et al. (19). First, we determined the nature of the *Pseudomonas* from our representative genomes. This is important since different *Pseudomonas* groups have different lifestyles and thus, different requirements in terms of specialized metabolites.

We used the housekeeping gene *gyrB* for phylogenetic analysis, since the 16S rRNA gene sequence has resolution to genera only and multi-locus sequence analysis with additional housekeeping genes, such as *rpoD*, *gyrB* and *cts* is required for species classification (12). The taxonomic assignment based on the *gyrB* gene and the GTBD database were consistent, confirming the suitability of this marker gene to achieve good-resolution at the species level (11,26,32). All the representative genomes were classified as *Pseudomonas_E*, which corresponds to the *P. fluorescens* lineage (12). Within this lineage, the main groups represented were the pathogenic *P. syringae* group and the commensal or beneficial clade *P. fluorescens*. Our collection was dominated by *P. viridiflava*, which belongs to the *P. syringae* species complex of the *P. fluorescens* lineage (26). This clade was the only one where all genomes were assigned to the same species.

Pseudomonas isolates encode a wide range of specialized metabolites gene clusters. We found that, on average, a genome encoded 11 BGCs, ranging from 3 to 25. Plant-associated *Pseudomonas* have the potential to produce 22 of the 56 BGC classes predicted by antiSMASH (35). Previous BGCs surveys of the *Pseudomonas* genus have found between 1 and 23 BGCs per genome, spanning 24 BGC classes in total (7,10,17). This is consistent with our results, even when we used only *Pseudomonas* isolated from *A. thaliana* leaves.

Few bacterial strains have the potential to produce more than 10 specialized metabolites (38). *Burkholderia* and *Streptomyces* are among the bacteria genera with

the largest biosynthetic potential: *Streptomyces scabiei* strain 87.22 encodes 32 BGCs (38). We found isolates encoding up to 25 BGCs; the genomes encoding more than 20 BGCs belonged exclusively to the ATUE3 and ATUE5 clades, and represented only 2.8% of the genomes included in this study. This suggests that isolates with a large biosynthetic potential are not the norm in *A. thaliana*-associated *Pseudomonas*, but that large variability is common, also among closely related strains.

It is important to note some caveats of BGCs annotation: first, the number of BGCs and the class they belong to can change depending on the tool (and the version thereof) that is used. Second, when using draft genomes, i.e. not closed, as is the case in this work, the BGCs encoding a single product might be split in two contigs. This can lead to an artificially high number of predicted BGCs, since each fragment of the BGCs in a different contig is identified as an additional gene cluster by antiSMASH (39). Indeed, split of a single BGC on different contigs was reported for some *Pseudomonas* NRPS, especially pyoverdine-related (10). We assessed this possibility in our data by predicting BGCs in 5 *Pseudomonas* genomes assembled in our lab using Nanopore, and found the biggest difference to be three BGCs less compared to the respective draft assemblies (not shown). To overcome this issue, closed genomes generated by either short- or long-read technologies are necessary. Fortunately, the decreasing cost of sequencing technologies enables this. In addition, the quality of the genomes used to annotate BGCs should always be assessed and the physical location of the detected BGCs identified, since BGCs on the contig edges are more likely to be split.

Clustering of BGCs into gene cluster families resulted in a very similar distribution of families compared to BGCs. This suggests that there is no redundancy in the specialized metabolites that can be produced by an isolate, so that each product is encoded by only one BGC in each isolate. Less than 5% of the gene cluster families we found had a closely-related hit in the MIBiG database. This indicates that the diversity of specialized metabolites is still underexplored in *Pseudomonas* despite this genus being known to harbor a large biosynthetic potential.

Most of this diversity is attributed to NRPS and NRPS-like BGCs, which made up 56% of the total BGCs we identified. NRPS are indeed the most common BGCs in other studies of the biosynthetic potential of *Pseudomonas* (7,10,17). NRPS have diverse functions, and can act as phytotoxins, antimicrobials, siderophores and surfactants (15). Interestingly, NRPS are only the fifth most common BGCs class in the entire domain of bacteria (40); the most common, saccharides, were not found in *Pseudomonas* in this study. Our results also indicate terpene diversity in *Pseudomonas* is larger than previous reports: we found 19 terpene gene cluster families, being present in 1 to 69 isolates. Until 2009, no terpenes nor terpene synthases had been described in *Pseudomonas (6)*, while more recent work has identified terpene BGCs in the genus, albeit in low numbers and with a narrow distribution (7,17). Terpenes are the largest and most structurally-diverse class of specialized metabolites (41); they are produced by plants and fungi, and

phytopathogenic fungi produce terpenes that act as toxins (42). The ability of bacteria to synthesize them has been recognized only recently, and thus the role of bacterial terpenes in disease is poorly studied. We found terpenes families were exclusive of one or a few clades, and they could have a role in niche specialization.

We identified NRPS families putatively involved in virulence of *P. viridiflava* ATUE5 on *A. thaliana*. Only one of these families was negatively correlated with plant size and thus, positively correlated with virulence. The closest known BCG to this family was pyoverdine, a high-affinity iron chelator synthesized by fluorescent *Pseudomonas* (6). Iron is an essential element for most, if not all, organisms, but due to its low bioavailability there is strong competition for it, including in host - pathogen interactions. For instance, vertebrates sequester iron upon infection, reducing its availability for the pathogen (43). A role of siderophores in promoting infection by the bacterial phytopathogens *Dickeya dadantii* and *Erwinia amylovora* has been shown (43), and their synthesis is required for full virulence of *P. syringae* pv. *tabaci* 6605 (44). Pyoverdine from *P. fluorescens* has been shown to weaken *A. thaliana* defense in order to promote plant growth under iron-limiting conditions (45). However, results are not always consistent regarding the role of siderophores on bacterial virulence (43). This BGC is a prime candidate for knock-out and knock-in studies to evaluate its role in virulence in this system.

Gene cluster families positively correlated with the effect on plant size were related to rhizomide, cichofactin and corchelin A. Cichofactin is involved in virulence, swarming and biofilm formation in *P. cichorii* (46), so it is interesting that our data suggest its absence is correlated with more virulence. However, the identity of this gene cluster family to cichofactin was around 65%, so they could have different functions. Rhizomide and crochelin are less well-characterized. Crochelin is a siderophore, as is pyoverdine, that uses an unusual γ-amino acid to bind iron (47), while rhizomide was described in *Burkholderia* species (48), and had weak protective activity against cucumber downy mildew when 6 phytopathogens were tested, and inhibited the growth of *Staphylococcus aureus* and *Bacillus subtilis in vitro* (48). Further characterization of these specialized metabolites is required in order to assess their role in virulence; however, many BGCs are expressed only under particular conditions, which cannot always be replicated in the lab, hindering their characterization.

We focused in this report on ATUE5 isolates since its virulence mechanisms are largely unknown despite it being agriculturally relevant and is a natural pathogen of the model plant *A. thaliana*. Including other ATUE clades could provide additional BGCs with potential roles in ATUE5 virulence. In addition, it could enable the identification of metabolites mediating the host protection against ATUE5 by other clades, particularly for ATUE2, that was recently described (31). In addition to determining potential BGCs associated with bacterial virulence, we found two new metabolites produced by *Pseudomonas*, and are in the process of characterizing their chemical and biological activity.

This work describes the vast biosynthetic potential for specialized metabolites in plant-associated *Pseudomonas*, and its relation with virulence, specifically in the context of a natural *A. thaliana* - *P. viridiflava* (ATUE5) pathosystem. To the best of our knowledge, this is the first study to do so. We identified NRPS putatively involved in virulence on *A. thaliana*. However, the product and function of most of the *Pseudomonas* BGCs are still unknown. Subsequent studies should focus on the elucidation of the conditions under which metabolites are produced and on their biological and ecological relevance.

METHODS

Pseudomonas representative genomes

The genomes used in this study were described by Karasov and colleagues and are available online (19). To select representative genomes, we first assessed the completeness of each genome using BUSCO version 3.0.2 (49). The 1498 genomes with a single-copy completeness score equal to or above 95% were clustered using the following procedure:

1. Sort all genome assemblies by their total length, from the longest to the shortest.

2. In each iteration, the longest assembly is set as the ‘representative’ of a newly formed cluster.

3. Assign members to the newly formed cluster based on the similarity of their orthology groups presence-absence profiles. We iterate over all genomes not yet assigned to a cluster, and compute the Jaccard similarity coefficient (50) between the orthology groups presence- absence profile of said genome and the cluster ‘representative’. Orthology groups were obtained from Karasov et al. PanX assignments (19). The Jaccard similarity coefficient was calculated as:

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

Where M_{11} is the number of orthology groups shared between the strains; M_{10} and M_{01} are orthology groups present in the representative and the assessed genome, respectively.

If the Jaccard similarity coefficient is greater or equal to 0.99, the genome is assigned to the cluster and removed from the list.

4. Repeat steps 2 and 3 until all genomes are assigned to a cluster.

Phylogenetics analysis

We obtained 105 reference *Pseudomonas* genomes from the NCBI (Table S1). From these reference genomes, together with the representatives selected above, we extracted a 751 bp fragment of the DNA gyrase beta subunit *gyrB* gene. We used the amplicon command from SeqKit version 2.2.0 (51) with the primers gyrB+271ps and gyrB-1022 (52). The taxonomic classification of the representative genomes was obtained using GTDB-Tk version 2.1.0 (53) and the release 207 of the Genome Taxonomy Database GTDB (33). We used the classify_wf workflow and default parameters.

Annotation of biosynthetic gene clusters

We used antiSMASH 5.1.2 (35) to predict biosynthetic gene clusters (BGCs) on each genome. When more than one BGC was predicted in a region, each BGC was counted independently in its corresponding class. For example, a region with “NRPS, terpene, siderophore” accounted for three BGCs in the analysis. Sequence similarity networks of the BGC were generated using BiG-SCAPE version 1.0.1 (36) with the default parameters. BiG-SCAPE assigns gene cluster families to one or more classes based on their antiSMASH annotation. The possible classes are PKS type I, PKS other, NRPS, RiPPs, saccharides, terpene, PKS/NRPS hybrids and ‘other’. We included entries from the Minimum Information about a Biosynthetic Gene (MIBiG) database (54), providing manually-curated BGC annotations with known functions, when running BiG-SCAPE.

Statistical analysis

Data and statistical analyses were performed using R v4.1.0 (55). Tree annotation and visualization was done using iToL (56).

ACKNOWLEDGEMENTS

We thank Nadine Ziemert and Jacobo de la Cuesta-Zuluaga for their support with the specialized metabolite annotation and analysis. We are grateful to Rebecca Schwab for her comments on the manuscript. This work was supported by the Max Planck Society.

REFERENCES

1. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. MBio. 2020 Jun 16;11(3).
2. O’Brien J, Wright GD. An ecological perspective of microbial secondary metabolism. Curr Opin Biotechnol. 2011 Aug;22(4):552–8.
3. Cragg GM, Newman DJ. Natural products: a continuing source of novel drug leads. Biochim Biophys Acta. 2013 Jun;1830(6):3670–95.
4. Etalo DW, Jeon JS, Raaijmakers JM. Modulation of plant chemistry by beneficial root microbiota. Nat Prod Rep. 2018 May 1;35(5):398–409.

5. Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. Natural functions of lipopeptides from Bacillus and Pseudomonas: more than surfactants and antibiotics. FEMS Microbiol Rev. 2010 Nov;34(6):1037–62.
6. Gross H, Loper JE. Genomics of secondary metabolite production by Pseudomonas spp. Nat Prod Rep. 2009 Nov;26(11):1408–46.
7. Stringlis IA, Zhang H, Pieterse CMJ, Bolton MD, de Jonge R. Microbial small molecules - weapons of plant subversion. Nat Prod Rep. 2018 May 25;35(5):410–33.
8. De Meyer G, Capieau K, Audenaert K, Buchala A, Métraux JP, Höfte M. Nanogram amounts of salicylic acid produced by the rhizobacterium Pseudomonas aeruginosa 7NSK2 activate the systemic acquired resistance pathway in bean. Mol Plant Microbe Interact. 1999 May;12(5):450–8.
9. Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol Rev. 2007 Jul;31(4):425–48.
10. Saati-Santamaría Z, Selem-Mojica N, Peral-Aranega E, Rivas R, García-Fraile P. Unveiling the genomic potential of Pseudomonas type strains for discovering new natural products. Microb Genom. 2022 Feb;8(2).
11. Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E. Phylogenomics and systematics in Pseudomonas. Front Microbiol. 2015 Mar 18;6:214.
12. Lalucat J, Mulet M, Gomila M, García-Valdés E. Genomics in Bacterial Taxonomy: Impact on the Genus Pseudomonas. Genes. 2020 Jan 29;11(2).
13. García-Valdés E, Lalucat J. Pseudomonas: Molecular phylogeny and current taxonomy. In: Pseudomonas: Molecular and Applied Biology. Cham: Springer International Publishing; 2016. p. 1–23.
14. Girard L, Lood C, Höfte M, Vandamme P, Rokni-Zadeh H, van Noort V, et al. The Ever-Expanding Pseudomonas Genus: Description of 43 New Species and Partition of the Pseudomonas putida Group. Microorganisms. 2021 Aug 18;9(8).
15. Götze S, Stallforth P. Structure, properties, and biological functions of nonribosomal lipopeptides from pseudomonads. Nat Prod Rep. 2020 Jan 1;37(1):29–54.
16. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol. 2015 Sep;11(9):625–31.
17. Alam K, Islam MM, Li C, Sultana S, Zhong L, Shen Q, et al. Genome Mining of Pseudomonas Species: Diversity and Evolution of Metabolic and Biosynthetic Potential. Molecules. 2021 Dec 12;26(24).
18. Rieusset L, Rey M, Muller D, Vacheron J, Gerin F, Dubost A, et al. Secondary metabolites from plant-associated Pseudomonas are overproduced in biofilm. Microb Biotechnol. 2020 Sep;13(5):1562–80.
19. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe. 2018 Jul 11;24(1):168–79.e4.
20. Wilkie JP, Dye DW, Watson DRW. Further hosts of Pseudomonas viridiflava. New Zealand Journal of Agricultural Research. 1973 Aug 1;16(3):315–23.
21. Lundberg DS, de Pedro Jové R, Ayutthaya PPN, Karasov TL, Shalev O, Poersch K, et al. Contrasting patterns of microbial dominance in the Arabidopsis thaliana

phyllosphere. bioRxiv. 2021. p. 2021.04.06.438366. doi: 10.1101/2021.04.06.438366v2

22. Goss EM, Kreitman M, Bergelson J. Genetic diversity, recombination and cryptic clades in Pseudomonas viridiflava infecting natural populations of Arabidopsis thaliana. Genetics. 2005 Jan;169(1):21–35.
23. Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J. Pseudomonas viridiflava and P. syringae--natural pathogens of Arabidopsis thaliana. Mol Plant Microbe Interact. 2002 Dec;15(12):1195–203.
24. Goss EM, Bergelson J. Fitness consequences of infection of Arabidopsis thaliana with its natural bacterial pathogen Pseudomonas viridiflava. Oecologia. 2007 May;152(1):71–81.
25. Bartoli C, Berge O, Monteil CL, Guilbaud C, Balestra GM, Varvaro L, et al. The Pseudomonas viridiflava phylogroups in the P. syringae species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. Environ Microbiol. 2014 Jul;16(7):2301–15.
26. Lipps SM, Samac DA. Pseudomonas viridiflava: An internal outsider of the Pseudomonas syringae species complex. Mol Plant Pathol. 2022 Jan;23(1):3–15.
27. Xin XF, Nomura K, Aung K, Velásquez AC, Yao J, Boutrot F, et al. Bacteria establish an aqueous living space in plants crucial for virulence. Nature. 2016 Nov 24;539(7630):524–9.
28. Hu Y, Ding Y, Cai B, Qin X, Wu J, Yuan M, et al. Bacterial effectors manipulate plant abscisic acid signaling for creation of an aqueous apoplast. Cell Host Microbe. 2022 Apr 13;30(4):518–29.e6.
29. Shalev O, Karasov TL, Lundberg DS, Ashkenazy H, Pramoj Na Ayutthaya P, Weigel D. Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant. Nat Ecol Evol. 2022 Apr;6(4):383–96.
30. Karasov TL, Neumann M, Shirsekar G, Monroe G, PATHODOPSIS Team, Weigel D, et al. Drought selection on Arabidopsis populations and their microbiomes. bioRxiv. 2022. doi: 10.1101/2022.04.08.487684v1
31. Shalev O, Ashkenazy H, Neumann M, Weigel D. Commensal Pseudomonas protect Arabidopsis thaliana from a coexisting pathogen via multiple lineage-dependent mechanisms. ISME J. 2022 May;16(5):1235–44.
32. Sarris PF, Trantas EA, Mpalantinaki E, Ververidis F, Goumas DE. Pseudomonas viridiflava, a multi host plant pathogen with significant genetic variation at the molecular level. PLoS One. 2012 Apr 27;7(4):e36090.
33. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018 Nov;36(10):996–1004.
34. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022 Jan 7;50(D1):D785–94.
35. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019 Apr 29.

36. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020 Jan;16(1):60–8.
37. Shahid I, Malik KA, Mehnaz S. A decade of understanding secondary metabolism in Pseudomonas spp. for sustainable agriculture and pharmaceutical applications. Environmental Sustainability. 2018 Mar 1;1(1):3–17.
38. Baldeweg F, Hoffmeister D, Nett M. A genomics perspective on natural product biosynthesis in plant pathogenic bacteria. Nat Prod Rep. 2019 Feb 20;36(2):307–25.
39. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W204–12.
40. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014 Jul 17;158(2):412–21.
41. Helfrich EJN, Lin GM, Voigt CA, Clardy J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. Beilstein J Org Chem. 2019 Nov 29;15:2889–906.
42. Cimmino A, Andolfi A, Evidente A. Phytotoxic Terpenes Produced by Phytopathogenic Fungi and Allelopathic Plants. Nat Prod Commun. 2014 Mar 1;9(3):1934578X1400900330.
43. Franza T, Expert D. Role of iron homeostasis in the virulence of phytopathogenic bacteria: an “à la carte” menu. Mol Plant Pathol. 2013 May;14(4):429–38.
44. Taguchi F, Suzuki T, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y. The siderophore pyoverdine of Pseudomonas syringae pv. tabaci 6605 is an intrinsic virulence factor in host tobacco infection. J Bacteriol. 2010 Jan;192(1):117–26.
45. Trapet P, Avoscan L, Klinguer A, Pateyron S, Citerne S, Chervin C, et al. The Pseudomonas fluorescens Siderophore Pyoverdine Weakens Arabidopsis thaliana Defense in Favor of Growth in Iron-Deficient Conditions. Plant Physiol. 2016 May;171(1):675–93.
46. Pauwelyn E, Huang CJ, Ongena M, Leclère V, Jacques P, Bleyaert P, et al. New linear lipopeptides produced by Pseudomonas cichorii SF1-54 are involved in virulence, swarming motility, and biofilm formation. Mol Plant Microbe Interact. 2013 May;26(5):585–98.
47. Baars O, Zhang X, Gibson MI, Stone AT, Morel FMM, Seyedsayamdost MR. Crochelins: Siderophores with an Unprecedented Iron-Chelating Moiety from the Nitrogen-Fixing Bacterium Azotobacter chroococcum. Angew Chem Int Ed Engl. 2018 Jan 8;57(2):536–41.
48. Wang X, Zhou H, Chen H, Jing X, Zheng W, Li R, et al. Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. Proc Natl Acad Sci U S A. 2018 May 1;115(18):E4255–63.
49. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol. 2019;1962:227–45.
50. Jaccard P. The distribution of the flora in the alpine zone. New Phytol. 1912 Feb;11(2):37–50.

51. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016 Oct 5;11(10):e0163962.
52. Hwang MSH, Morgan RL, Sarkar SF, Wang PW, Guttman DS. Phylogenetic characterization of virulence and resistance phenotypes of Pseudomonas syringae. Appl Environ Microbiol. 2005 Sep;71(9):5182–91.
53. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. bioRxiv. 2022. p. 2022.07.11.499641. doi: 10.1101/2022.07.11.499641v1
54. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 2020 Jan 8;48(D1):D454–8.
55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
56. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.

SUPPLEMENTARY MATERIAL

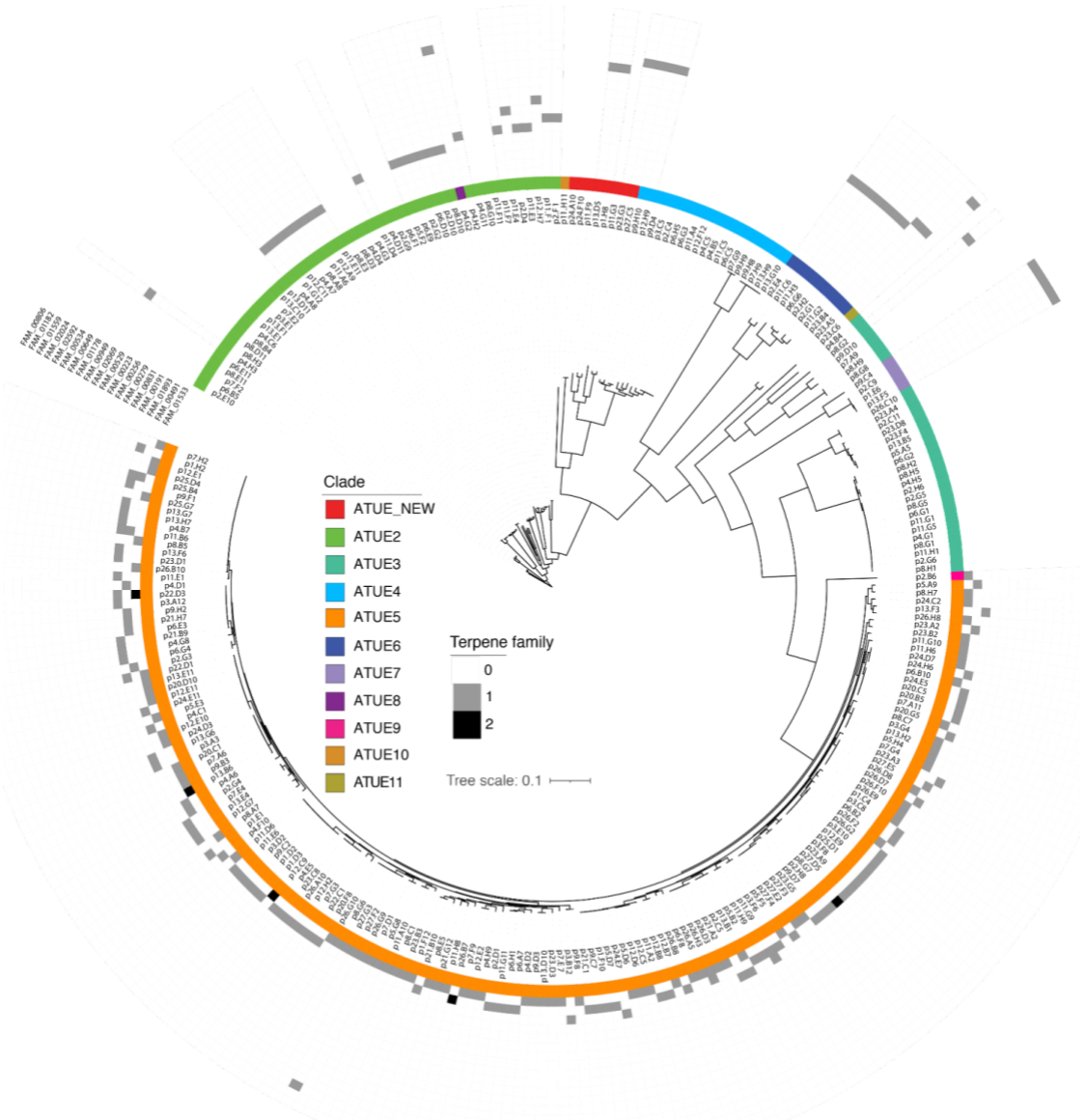

Figure S1 - Terpene BGCs are distributed throughout the phylogeny of plant-associated *Pseudomonas*. Distribution of the 19 terpene gene cluster families identified in 284 genomes. The maximum likelihood phylogenetic tree from (19), was annotated with clade (first circle inside) and the number of copies of each gene cluster family in each isolate (remaining circles, gray scale).

Table S1. Genomes included in this study

Representative genomes from Karasov et al., 2018 (24)		Reference genomes	
NCBI ID	**Name**	**NCBI ID**	**Name**
GCF_900582925.1	p1.C4	GCA_000006765.1	*P. aeruginosa* PAO1
GCF_900583045.1	p1.D2	GCA_001618925.1	*P. aeruginosa* strain ATCC 27853
GCF_900583005.1	p1.D3	GCA_001516325.2	*P. aeruginosa* strain H27930
GCF_900583065.1	p1.E1	GCA_001597285.1	*P. alcaligenes* strain NEB 585
GCF_900583105.1	p1.E6	GCA_001941865.1	*P. alcaliphila* JAB1
GCF_900583195.1	p1.F10	GCA_002007785.1	*P. azotoformans* strain F77
GCF_900583175.1	p1.F12	GCA_001602135.1	*P. chlororaphis* isolate 189
GCF_900583255.1	p1.G12	GCA_000698865.1	*P. chlororaphis* strain PA23
GCF_900583335.1	p1.H2	GCA_000963835.1	*P. chlororaphis* strain PCL1606
GCF_900583455.1	p1.H8	GCA_000761195.1	*P. chlororaphis* subsp. *aurantiaca* strain JD37
GCF_900589175.1	p2.B6	GCA_001586155.1	*P. citronellolis* strain P3B5
GCF_900589295.1	p2.C11	GCA_001708425.1	*P. corrugata* strain RM1-1-4
GCF_900589285.1	p2.C4	GCA_000349845.1	*P. denitrificans* ATCC 13867
GCF_900589325.1	p2.C5	GCA_000026105.1	*P. entomophila* str. L48
GCF_900589395.1	p2.C9	GCA_000262325.2	*P. fluorescens* A506
GCF_900589405.1	p2.D1	GCA_000009225.1	*P. fluorescens* SBW25
GCF_900589385.1	p2.D10	GCA_001307155.1	*P. fluorescens* strain FW300-N2E3
GCF_900589345.1	p2.D4	GCA_001708465.1	*P. fluorescens* strain L111
GCF_900589445.1	p2.E10	GCA_001708445.1	*P. fluorescens* strain L321
GCA_900589535.1	p2.E4	GCA_001612705.1	*P. fluorescens* strain LBUM636
GCF_900589545.1	p2.F1	GCA_001747385.1	*P. fluorescens* strain Pt14
GCA_900589655.1	p2.G1	GCA_002128325.1	*P. fragi* strain NMC25
GCF_900589665.1	p2.G2	GCA_001543265.1	*P. fragi* strain P121
GCF_900589725.1	p2.G3	GCA_001952935.1	*P. frederiksbergensis* strain AS1
GCF_900589675.1	p2.G4	GCA_000689415.1	*P. knackmussii* B13
GCF_900589735.1	p2.G5	GCA_000257545.3	*P. mandelii* JR-1
GCF_900589695.1	p2.G6	GCA_000733715.2	*P. mendocina* S5.2
GCF_900589715.1	p2.G9	GCA_000016565.1	*P. mendocina* ymp
GCA_900589765.1	p2.H2	GCA_000510285.1	*P. monteilii* SB3078
GCF_900589785.1	p2.H6	GCA_000510325.1	*P. monteilii* SB3101
GCF_900589825.1	p2.H8	GCA_001534745.1	*P. monteilii* strain USDA-ARS-USMARC-56711
GCF_900589855.1	p3.A12	GCA_000498975.2	*P. mosselii* SJ10
GCF_900589885.1	p3.A3	GCA_000831585.1	*P. plecoglossicida* strain NyZ12

GCF_900589975.1	p3.B12	GCA_000828695.1	*P. protegens* Cab57 DNA
GCF_900590165.1	p3.C5	GCA_000397205.1	*P. protegens* CHA0
GCF_900590105.1	p3.C8	GCA_000012265.1	*P. protegens* Pf-5
GCF_900590245.1	p3.D2	GCA_002006545.1	*P. protegens* strain H78
GCF_900590325.1	p3.E1	GCA_000226035.3	*P. putida* B6-2
GCF_900590255.1	p3.E10	GCA_000183645.1	*P. putida* BIRD-1
GCF_900590405.1	p3.F6	GCA_000281215.1	*P. putida* DOT-T1E
GCF_900590445.1	p3.F8	GCA_000016865.1	*P. putida* F1
GCF_900590565.1	p3.G4	GCA_000019125.1	*P. putida* GB-1
GCF_900590645.1	p4.A6	GCA_000410575.1	*P. putida* H8234
GCF_900590695.1	p4.A7	GCA_000325725.1	*P. putida* HB3267
GCF_900590725.1	p4.A8	GCA_001767335.1	*P. putida* JB
GCF_900590755.1	p4.B4	GCA_000007565.2	*P. putida* KT2440
GCF_900590795.1	p4.B5	GCA_000412675.1	*P. putida* NBRC 14164 DNA
GCF_900590865.1	p4.B7	GCA_000264665.1	*P. putida* ND6
GCF_900590815.1	p4.C1	GCA_000495455.2	*P. putida* S12
GCF_900590905.1	p4.C5	GCA_000219705.1	*P. putida* S16
GCF_900590935.1	p4.C6	GCA_000271965.2	*P. putida* SJTE-1
GCF_900590915.1	p4.D1	GCA_001515585.2	*P. putida* strain 1A00316
GCF_900590885.1	p4.D11	GCA_002025705.1	*P. putida* strain AA7
GCF_900591005.1	p4.D2	GCA_000691565.1	*P. putida* strain DLL-E4
GCF_900590965.1	p4.D4	GCA_001886975.1	*P. putida* strain PP112420
GCF_900591055.1	p4.E5	GCA_900074915.1	*P.* sp. 58 isolate Sour cherry
GCF_900591085.1	p4.F10	GCA_001661075.1	*P.* sp. A3 2016
GCF_900591215.1	p4.G1	GCA_000511325.1	*P.* sp. FGI182
GCF_900591185.1	p4.G11	GCA_001655615.1	*P.* sp. GR 6-02
GCF_900591195.1	p4.G2	GCA_001655295.1	*P.* sp. JY-Q
GCF_900591205.1	p4.G3	GCA_001294575.1	*P.* sp. L10.10
GCF_900598765.1	p4.G8	GCA_002037565.1	*P.* sp. LPH1
GCF_900598835.1	p4.H2	GCA_000931465.1	*P.* sp. MRSN12121
GCF_900598825.1	p4.H3	GCA_001547895.1	*P.* sp. Os17 DNA
GCF_900598875.1	p4.H5	GCA_001547915.1	*P.* sp. St29 DNA
GCF_900598895.1	p4.H9	GCA_001511755.1	*P.* sp. URMO17WK12 I11 Shine
GCF_900599035.1	p5.A5	GCA_000316175.1	*P.* sp. UW4
GCF_900598995.1	p5.A9	GCA_000494915.1	*P.* sp. VLB120
GCF_900599045.1	p5.B2	GCA_000267545.1	*P. stutzeri* CCUG 29243

GCF_900599255.1	p5.D6	GCA_001482725.1	*P. syringae* pv. *lapsa* strain ATCC 10859
GCF_900599245.1	p5.D7	GCA_000988485.1	*P. syringae* pv. *syringae* B301D
GCF_900582325.1	p5.E3	GCA_000012245.1	*P. syringae* pv. *syringae* B728a
GCF_900599435.1	p5.F2	GCA_000988395.1	*P. syringae* pv. *syringae* HS191
GCF_900599485.1	p5.F5	GCA_000007805.1	*P. syringae* pv. *tomato* str. DC3000
GCF_900599605.1	p5.G8	GCA_001281365.1	*P. syringae* UMAF0158
GCF_900599695.1	p5.H4	GCA_001186335.1	*P. trivialis* strain IHBB745
GCF_900599775.1	p6.A7	GCA_002028325.1	*P. veronii* strain R02
GCF_900599855.1	p6.B10	GCF_016307575.1	*P. viridiflava* strain 13 3a 2dr YA0757 1
GCF_900599865.1	p6.B2	GCF_016307515.1	*P. viridiflava* strain 13 3a 2er YA0758 1
GCF_900599875.1	p6.B5	GCF_016307535.1	*P. viridiflava* strain 13 3a 2fr YA0759 1
GCF_900599915.1	p6.C5	GCF_022936465.1	*P. viridiflava* strain A2M176 A2M176 1
GCF_900600065.1	p6.D10	GCF_009765495.1	*P. viridiflava* strain BAV 2572
GCF_900600165.1	p6.E11	GCF_001716855.1	*P. viridiflava* strain CDRTc14
GCF_900600115.1	p6.E3	GCF_900184295.1	*P. viridiflava* strain CFBP 1590 isolate E12-5
GCF_900600215.1	p6.E9	GCF_900184295.1	*P. viridiflava* strain CFBP 1590 isolate E12-5
GCF_900600265.1	p6.F1	GCF_002406485.1	*P. viridiflava* strain CH409
GCF_900600395.1	p6.F8	GCF_001305955.1	*P. viridiflava* strain DSM 6694
GCF_900600385.1	p6.G1	GCF_001401215.1	*P. viridiflava* strain ICMP 2848
GCF_900600355.1	p6.G2	GCF_001642795.1	*P. viridiflava* strain ICMP 2848 C4216
GCF_900600375.1	p6.G3	GCF_002723575.1	*P. viridiflava* strain ICMP 8820
GCF_900600335.1	p6.G4	GCF_019083885.1	*P. viridiflava* strain KF485 1
GCA_900600405.1	p6.G6	GCF_000834695.1	*P. viridiflava* strain LMCA8
GCF_900600445.1	p6.H1	GCF_016307685.1	*P. viridiflava* strain Pvir 12 2b YA0701 1
GCF_900600465.1	p6.H5	GCF_016307675.1	*P. viridiflava* strain Pvir 9r 6 YA0697 1
GCF_900600525.1	p7.A11	GCF_016307775.1	*P. viridiflava* strain Pvir12 2 YA0089 1
GCF_900600595.1	p7.A6	GCF_016308365.1	*P. viridiflava* strain Pvir6 2 YA0001 1
GCF_900600635.1	p7.A9	GCF_019104005.1	*P. viridiflava* strain StP4
GCF_900600815.1	p7.D1	GCF_019104065.1	*P. viridiflava* strain SV1779
GCF_900601015.1	p7.E2	GCF_019104045.1	*P. viridiflava* strain T1426
GCF_900601065.1	p7.E4	GCF_019104025.1	*P. viridiflava* strain T157
GCF_900601035.1	p7.E7	GCF_018388545.1	*P. viridiflava* strain U625
GCF_900601165.1	p7.F2	GCF_019083835.1	*P. viridiflava* strain U658 1
GCF_900601095.1	p7.F9	GCF_022698325.1	*P. viridiflava* strain UCD-PV1 1

GCF_900601215.1	p7.G3
GCF_900601225.1	p7.G4
GCF_900601305.1	p7.G9
GCF_900601345.1	p7.H2
GCF_900601405.1	p7.H9
GCF_900601425.1	p8.A7
GCF_900601395.1	p8.A8
GCF_900601515.1	p8.B4
GCF_900601545.1	p8.B5
GCF_900601555.1	p8.C1
GCF_900601665.1	p8.C7
GCF_900601635.1	p8.D10
GCF_900601595.1	p8.D11
GCF_900601735.1	p8.D3
GCF_900601695.1	p8.E11
GCF_900601815.1	p8.E3
GCF_900601785.1	p8.E5
GCF_900601895.1	p8.G1
GCF_900601915.1	p8.G10
GCF_900601905.1	p8.G2
GCF_900601995.1	p8.G5
GCF_900601975.1	p8.G6
GCF_900602025.1	p8.G7
GCF_900602065.1	p8.G8
GCF_900602015.1	p8.H1
GCF_900602005.1	p8.H2
GCF_900602055.1	p8.H3
GCF_900602155.1	p8.H5
GCF_900602105.1	p8.H7
GCF_900602115.1	p8.H9
GCF_900602315.1	p9.B3
GCF_900602375.1	p9.C2
GCF_900602385.1	p9.C4
GCF_900602405.1	p9.C7
GCF_900602545.1	p9.D10
GCF_900602555.1	p9.D3
GCF_900602525.1	p9.D4

GCF_000452485.1	*P. viridiflava* TA043
GCF_000307715.1	*P. viridiflava* UASWS0038

GCF_900602575.1	p9.D7
GCF_900602675.1	p9.F1
GCF_900602805.1	p9.F8
GCF_900602905.1	p9.H10
GCF_900602895.1	p9.H2
GCF_900602975.1	p9.H8
GCF_900573885.1	p9.H9
GCF_900576585.1	p11.A10
GCF_900576625.1	p11.A2
GCF_900576645.1	p11.A4
GCF_900576665.1	p11.A6
GCF_900580505.1	p11.B6
GCF_900580565.1	p11.C5
GCF_900580595.1	p11.C6
GCF_900580675.1	p11.D4
GCF_900580685.1	p11.D6
GCF_900580705.1	p11.E1
GCF_900580655.1	p11.E11
GCF_900580765.1	p11.E3
GCF_900580775.1	p11.E4
GCF_900580785.1	p11.E6
GCF_900580855.1	p11.F1
GCF_900580845.1	p11.F11
GCF_900580795.1	p11.F7
GCF_900580865.1	p11.F9
GCF_900580895.1	p11.G1
GCF_900580885.1	p11.G10
GCF_900580905.1	p11.G11
GCA_900580915.1	p11.G2
GCF_900580925.1	p11.G3
GCF_900580955.1	p11.G5
GCF_900580995.1	p11.G9
GCF_900581015.1	p11.H1
GCF_900581005.1	p11.H11
GCF_900581025.1	p11.H3
GCF_900581055.1	p11.H6
GCF_900581095.1	p11.H8

GCF_900581115.1	p11.H9
GCF_900581175.1	p12.A9
GCF_900581235.1	p12.B7
GCF_900581285.1	p12.B8
GCF_900581295.1	p12.C11
GCF_900581375.1	p12.C5
GCF_900581395.1	p12.C9
GCF_900581475.1	p12.D6
GCF_900581435.1	p12.E1
GCF_900581425.1	p12.E10
GCF_900581405.1	p12.E11
GCF_900581455.1	p12.E2
GCF_900581595.1	p12.E9
GCF_900581605.1	p12.F12
GCF_900581665.1	p12.G7
GCF_900581725.1	p12.H2
GCA_900581805.1	p12.H7
GCF_900581815.1	p12.H9
GCF_900581895.1	p13.B1
GCF_900581935.1	p13.B5
GCF_900581965.1	p13.B6
GCF_900582015.1	p13.C10
GCF_900582095.1	p13.D10
GCF_900582105.1	p13.D11
GCF_900582195.1	p13.D5
GCF_900582235.1	p13.E1
GCF_900582245.1	p13.E11
GCF_900582265.1	p13.E4
GCF_900582335.1	p13.F1
GCF_900582345.1	p13.F3
GCF_900582365.1	p13.F5
GCF_900582375.1	p13.F6
GCF_900582425.1	p13.G10
GCF_900582505.1	p13.G6
GCF_900582515.1	p13.G7
GCF_900582535.1	p13.H2
GCF_900582575.1	p13.H7

GCF_900582625.1	p13.H9
GCF_900583695.1	p20.B5
GCA_900583765.1	p20.C1
GCF_900583735.1	p20.C5
GCF_900583725.1	p20.D10
GCF_900584015.1	p20.F8
GCF_900584115.1	p20.G5
GCF_900584365.1	p21.A2
GCF_900584465.1	p21.B10
GCF_900584485.1	p21.B9
GCF_900584565.1	p21.C1
GCF_900584965.1	p21.G12
GCF_900585135.1	p21.H7
GCF_900585355.1	p22.C1
GCF_900585415.1	p22.D1
GCF_900585425.1	p22.D3
GCF_900585675.1	p23.A2
GCF_900585725.1	p23.A3
GCF_900585685.1	p23.A4
GCF_900585705.1	p23.A5
GCF_900585795.1	p23.A9
GCF_900585785.1	p23.B2
GCF_900585775.1	p23.B3
GCF_900585815.1	p23.B4
GCF_900585905.1	p23.C6
SAMEA104472330	p23.C8
GCF_900585955.1	p23.D1
GCF_900586005.1	p23.D3
GCF_900586015.1	p23.D8
GCF_900586155.1	p23.F4
GCF_900586135.1	p23.G3
GCF_900586165.1	p23.G5
GCF_900576715.1	p24.A10
GCF_900586525.1	p24.C2
GCF_900586655.1	p24.D3
GCF_900586685.1	p24.D7
GCF_900586705.1	p24.E11

GCF_900586745.1	p24.E5
GCF_900586775.1	p24.E7
GCF_900586795.1	p24.F10
GCF_900587125.1	p24.H6
GCF_900587325.1	p25.B4
GCF_900587555.1	p25.D1
GCF_900587545.1	p25.D4
GCF_900587935.1	p25.G7
GCF_900588025.1	p26.A10
GCF_900588065.1	p26.A5
GCF_900588125.1	p26.B10
GCF_900588205.1	p26.B7
GCF_900588245.1	p26.B8
GCF_900588235.1	p26.C10
GCF_900588325.1	p26.D3
GCF_900588315.1	p26.D7
GCF_900588355.1	p26.D8
GCF_900588485.1	p26.E9
GCF_900588455.1	p26.F10
GCF_900588545.1	p26.F2
GCF_900588575.1	p26.G10
GCF_900588565.1	p26.G2
GCF_900588685.1	p26.G9
GCF_900588695.1	p26.H3
GCF_900588765.1	p26.H8
GCF_900588845.1	p27.C5
GCF_900588875.1	p27.D5
GCF_900588925.1	p27.E2
GCF_900588945.1	p27.E5
GCF_900588965.1	p27.F2
GCF_900588955.1	p27.F3
GCF_900588975.1	p27.F4
GCF_900589005.1	p27.G3

Table S2. GTDB taxonomical classification and predicted BGCs of representatives genomes

Isolate	Clade	Species	fastANI reference	Classification method*	Total BGCs	Acyl amino acids	Arylpolyene	Bacteriocin	Betalactone	Butyrolactone	CDPS	Ectoine	Hserlactone	Lanthipeptide	LAP	Lassopeptide	NAGGN	NRPS	NRPS-like	Phenazine	Resorcinol	Siderophore	Type 1 PKS	Terpene	TfuA-related	Thiopeptide	Trans AT-PKS
p1.H8	ATUE1	P_E sp900582195	GCF_900582195.1	a	12	0	1	2	1	1	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0
p11.F9	ATUE1	P_E sp900580865	GCF_900580865.1	a	10	0	1	2	1	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
p11.G3	ATUE1	P_E	N/A	a	7	0	1	1	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p13.D5	ATUE1	P_E sp900582195	GCF_900582195.1	a	14	0	1	2	1	1	0	0	0	0	0	1	1	5	1	0	0	0	0	1	0	0	0
p23.G3	ATUE1	P_E sp001297015	GCF_001297015.1	a	11	0	1	2	2	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p24.A10	ATUE1	P_E gregormendelii	GCF_017114825.1	a	7	0	1	1	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p24.F10	ATUE1	P_E gregormendelii	GCF_017114825.1	a	7	0	1	1	2	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0
p27.C5	ATUE1	P_E sp001297015	GCF_001297015.1	a	10	0	1	2	1	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p1.C4	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p1.D2	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	1	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p1.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p1.E1	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p1.F10	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	6	2	0	0	2	0	1	0	0	0

p1.F12	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p1.H2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p11.A10	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.A2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.B6	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	0	0	0	0	0	0	0	0	0	0	0	0	8	2	0	0	2	0	2	0	0	0
p11.D6	ATUE5	P_E viridiflava	GCF_001642795.1	a	21	0	1	0	0	0	0	0	0	0	0	0	0	1	12	3	0	0	3	0	1	0	0	0
p11.E1	ATUE5	P_E viridiflava	GCF_001642795.1	a	7	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.E6	ATUE5	P_E viridiflava	GCF_001642795.1	a	15	0	1	0	1	0	0	0	0	0	0	0	0	1	8	2	0	0	1	0	1	0	0	0
p11.G10	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.G11	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p11.G9	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	0	1	5	2	0	0	1	0	2	0	0	0
p11.H6	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.H8	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p11.H9	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p12.B7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p12.B8	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	0	1	6	2	0	0	1	0	1	0	0	0
p12.C5	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p12.C9	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	0	1	6	2	0	0	2	0	2	0	0	0
p12.D6	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p12.E1	ATUE5	P_E viridiflava	GCF_001642795.1	a	7	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p12.E10	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p12.E11	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p12.E2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0

p12.E9	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p12.G7	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p12.H2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p13.B1	ATUE5	P_E viridiflava	GCF_001642795.1	a	19	0	1	1	0	0	0	0	0	0	0	0	1	12	1	0	0	1	0	2	0	0	0
p13.B6	ATUE5	P_E viridiflava	GCF_001642795.1	a	22	0	0	0	0	0	0	0	0	0	0	0	1	16	1	0	0	2	0	2	0	0	0
p13.D10	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p13.E11	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p13.E4	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p13.F3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p13.F6	ATUE5	P_E viridiflava	GCF_001642795.1	a	7	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p13.G6	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	0	0	0	0	0	0	0	0	0	0	1	6	1	0	0	2	0	1	0	0	0
p13.G7	ATUE5	P_E viridiflava	GCF_001642795.1	a	17	0	0	0	0	0	0	0	0	0	0	0	1	11	3	0	0	1	0	1	0	0	0
p13.H2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p13.H7	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	0	0	0	0	0	0	0	0	0	0	1	10	1	0	0	1	0	1	0	0	0
p2.C5	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	1	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p2.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p2.G3	ATUE5	P_E viridiflava	GCF_001642795.1	a	16	0	1	0	0	0	0	0	0	0	0	0	1	10	2	0	0	1	0	1	0	0	0
p2.G4	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p2.H8	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p20.B5	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p20.C1	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p20.C5	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p20.D10	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0

p20.F8	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p20.G5	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p21.A2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p21.B10	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	6	2	0	0	1	0	1	0	0	0
p21.B9	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	0	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p21.C1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p21.G12	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p21.H7	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	0	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p22.C1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p22.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	25	0	1	0	0	0	0	0	0	0	0	0	0	16	4	0	0	2	0	2	0	0	0
p22.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	21	0	0	0	0	0	0	0	0	0	0	0	1	13	2	0	0	2	0	3	0	0	0
p23.A2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p23.A3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p23.A9	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	1	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p23.B2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p23.B3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p23.C8	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p23.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p23.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p23.G5	ATUE5	P_E viridiflava	GCF_001642795.1	a	15	0	1	0	0	0	0	0	0	0	0	0	0	8	2	0	0	3	0	1	0	0	0
p24.C2	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p24.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p24.D7	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	0	1	0	0	0

p24.E11	ATUE5	P_E viridiflava	GCF_001642795.1	a	19	0	1	0	0	0	0	0	0	0	0	0	1	12	3	0	0	1	0	1	0	0	0
p24.E5	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p24.E7	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	2	0	1	0	0	0
p24.H6	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	0	1	0	0	0
p25.B4	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	0	0	0	0	0	0	0	0	0	0	1	9	1	0	0	1	0	2	0	0	0
p25.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	7	2	0	0	1	0	1	0	0	0
p25.D4	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	9	1	0	0	1	0	1	0	0	0
p25.G7	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	0	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p26.A10	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p26.A5	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p26.B10	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	0	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p26.B7	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	0	0	2	0	0	0
p26.B8	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	6	3	0	0	1	0	1	0	0	0
p26.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p26.D7	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p26.D8	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p26.E9	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p26.F10	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p26.F2	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	5	2	0	0	1	0	1	0	0	0
p26.G10	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	2	0	1	0	0	0
p26.G2	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p26.G9	ATUE5	P_E viridiflava	GCF_001642795.1	a	17	0	1	0	0	0	0	0	0	0	0	0	1	11	1	0	0	1	0	2	0	0	0
p26.H3	ATUE5	P_E viridiflava	GCF_001642795.1	a	16	0	1	0	0	0	0	0	0	0	0	0	1	10	1	0	0	1	0	2	0	0	0

p26.H8	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p27.D5	ATUE5	P_E viridiflava	GCF_001642795.1	a	15	0	1	1	0	0	0	0	0	0	0	0	1	9	1	0	0	1	0	1	0	0	0
p27.E2	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	2	0	1	0	0	0
p27.E5	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	1	0	0	0
p27.F2	ATUE5	P_E viridiflava	GCF_001642795.1	a	19	0	1	0	0	0	0	0	0	0	0	0	1	11	4	0	0	1	0	1	0	0	0
p27.F3	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p27.F4	ATUE5	P_E viridiflava	GCF_001642795.1	a	18	0	1	0	0	0	0	0	0	0	0	0	1	11	2	0	0	2	0	1	0	0	0
p27.G3	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	5	2	0	0	1	0	1	0	0	0
p3.A12	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	0	0	0	0	0	0	0	0	0	0	1	8	2	0	0	2	0	1	0	0	0
p3.A3	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	0	0	0	0	0	0	0	0	0	0	1	7	1	0	0	2	0	1	0	0	0
p3.B12	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p3.C8	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p3.D2	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	8	2	0	0	1	0	1	0	0	0
p3.E10	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	0	1	0	0	0
p3.F6	ATUE5	P_E viridiflava	GCF_001642795.1	a	20	0	1	1	0	0	0	0	0	0	0	0	0	12	3	0	0	2	0	1	0	0	0
p3.F8	ATUE5	P_E viridiflava	GCF_001642795.1	a	16	0	1	0	0	0	0	0	0	0	0	0	1	11	1	0	0	1	0	1	0	0	0
p3.G4	ATUE5	P_E viridiflava	GCF_001642795.1	a	22	0	1	0	0	0	0	0	0	0	0	0	1	14	2	0	0	3	0	1	0	0	0
p4.A6	ATUE5	P_E viridiflava	GCF_001642795.1	a	15	0	0	0	0	0	0	0	0	0	0	0	1	10	1	0	0	2	0	1	0	0	0
p4.B7	ATUE5	P_E viridiflava	GCF_001642795.1	a	20	0	0	0	0	0	0	0	0	0	0	0	1	13	3	0	0	2	0	1	0	0	0
p4.C1	ATUE5	P_E viridiflava	GCF_001642795.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	0	1	0	0	0
p4.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	0	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p4.D2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p4.E5	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	9	1	0	0	1	0	1	0	0	0

p4.F10	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	3	2	0	0	1	0	1	0	0	0
p4.G8	ATUE5	P_E viridiflava	GCF_001642795.1	a	7	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p4.H9	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p5.A9	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.B2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.D6	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.D7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.E3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.F5	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.G8	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p5.H4	ATUE5	P_E viridiflava	GCF_001642795.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	7	1	0	0	1	0	1	0	0	0
p6.A7	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p6.B10	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	3	2	0	0	1	0	1	0	0	0
p6.B2	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p6.E3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p6.F8	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p6.G4	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p6.H1	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p7.A11	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.A6	ATUE5	P_E viridiflava	GCF_001642795.1	a	7	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.D1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.E4	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.E7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0

p7.F9	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p7.G3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.G4	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p7.H2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p8.A7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p8.B5	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p8.C1	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p8.C7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p8.E5	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p8.G6	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p8.G7	ATUE5	P_E viridiflava	GCF_001642795.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	0	2	0	0	0
p8.H7	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	1	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p9.B3	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	1	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p9.C2	ATUE5	P_E viridiflava	GCF_001642795.1	a	11	0	1	0	1	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p9.C7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p9.D3	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p9.D7	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	0	1	0	0	0
p9.F1	ATUE5	P_E viridiflava	GCF_001642795.1	a	9	0	0	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p9.F8	ATUE5	P_E viridiflava	GCF_001642795.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	0	0	0
p9.H2	ATUE5	P_E viridiflava	GCF_001642795.1	a	8	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	1	0	0	0
p11.H11	ATUE10	P_E sp900581005	GCF_900581005.1	a	12	0	1	2	1	1	0	0	0	1	0	0	1	4	1	0	0	0	0	0	0	0	0
p23.C6	ATUE11	P_E sp900585905	GCF_900585905.1	a	11	0	1	0	0	0	0	0	0	0	1	0	1	6	0	0	0	0	0	1	0	1	0

p1.G12	ATUE2	P_E sp002979555	GCF_002979555.1	a	8	0	1	2	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0
p11.A6	ATUE2	P_E sp002979555	GCF_002979555.1	a	10	0	1	2	1	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p11.D4	ATUE2	P_E sp900580675	GCF_900580675.1	a	16	0	1	1	1	0	0	0	0	0	0	0	1	8	2	0	0	0	0	1	0	1	0
p11.E11	ATUE2	P_E sp002979555	GCF_002979555.1	a	8	0	1	2	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0
p11.E3	ATUE2	P_E orientalis_A	GCF_002934065.1	b	19	0	1	2	1	0	0	0	1	0	1	0	1	8	1	0	0	0	0	1	0	2	0
p11.E4	ATUE2	P_E orientalis_A	GCF_002934065.1	b	20	0	1	2	1	0	0	0	1	0	1	0	1	9	1	0	0	0	0	1	0	2	0
p11.F1	ATUE2	P_E poae	GCF_001439785.1	a	11	0	1	2	0	0	0	0	0	0	0	0	1	5	1	0	0	0	0	1	0	0	0
p11.F11	ATUE2	P_E orientalis_A	GCF_002934065.1	b	16	0	1	2	2	0	0	0	0	0	0	0	1	7	1	0	0	0	1	0	0	1	0
p11.F7	ATUE2	P_E orientalis_A	GCF_002934065.1	b	18	0	1	2	2	0	0	0	1	0	1	0	1	6	1	0	0	0	0	1	0	2	0
p12.A9	ATUE2	P_E sp002979555	GCF_002979555.1	a	10	0	1	2	0	0	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0
p12.C11	ATUE2	P_E sp002979555	GCF_002979555.1	a	10	0	1	2	0	0	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0
p12.H7	ATUE2	P_E trivialis	GCF_001439805.1	a	13	0	1	1	1	0	0	0	0	0	0	0	1	7	1	0	0	0	0	1	0	0	0
p13.C10	ATUE2	P_E sp900583165	GCF_900583165.1	a	14	0	1	2	1	1	0	0	0	0	0	0	1	6	1	0	0	0	1	0	0	0	0
p13.D11	ATUE2	P_E sp900583165	GCF_900583165.1	a	14	0	1	2	1	1	0	0	0	0	0	0	1	6	1	0	0	0	1	0	0	0	0
p13.E1	ATUE2	P_E sp900583165	GCF_900583165.1	a	13	0	1	2	1	1	0	0	0	0	0	0	1	6	1	0	0	0	0	0	0	0	0
p13.F1	ATUE2	P_E sp900583165	GCF_900583165.1	a	13	0	1	2	1	1	0	0	0	0	0	0	1	6	1	0	0	0	0	0	0	0	0
p2.D10	ATUE2	P_E sivasensis	GCF_013778505.1	a	14	0	2	2	1	1	0	0	0	0	0	0	1	5	1	0	0	0	0	1	0	0	0

p2.D4	ATUE2	P_E orientalis_A	GCF_002934065.1	b	18	0	1	2	1	0	0	0	1	0	1	0	1	7	1	0	0	0	0	1	0	2	0
p2.E10	ATUE2	P_E marginalis_B	GCF_001645105.1	a	12	0	1	2	1	0	0	0	0	0	0	0	1	6	1	0	0	0	0	0	0	0	0
p2.F1	ATUE2	P_E poae	GCF_001439785.1	a	11	0	1	2	0	0	0	0	0	0	0	0	1	5	1	0	0	0	0	1	0	0	0
p2.G2	ATUE2	P_E sivasensis	GCF_013778505.1	a	11	0	1	2	1	1	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p2.G9	ATUE2	P_E sp002843605	GCF_002843605.1	a	12	0	1	2	1	0	0	0	0	0	0	0	1	3	3	0	0	0	0	0	0	1	0
p3.E1	ATUE2	P_E sp900583165	GCF_900583165.1	a	15	0	1	2	1	1	0	0	0	0	0	0	1	7	1	0	0	0	1	0	0	0	0
p4.A7	ATUE2	P_E sp002979555	GCF_002979555.1	a	8	0	1	2	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0
p4.A8	ATUE2	P_E sp002979555	GCF_002979555.1	a	9	0	1	2	0	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p4.C6	ATUE2	P_E canadensis	GCF_000503215.1	a	12	0	2	3	1	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
p4.D11	ATUE2	P_E sp003097075	GCF_003097075.1	a	11	0	1	2	1	0	0	0	0	0	0	0	1	5	1	0	0	0	0	0	0	0	0
p4.D4	ATUE2	P_E salomonii	GCF_900107155.1	a	16	0	1	2	1	0	0	1	0	0	0	0	1	8	1	0	0	1	0	0	0	0	0
p4.G11	ATUE2	P_E orientalis_A	GCF_002934065.1	b	16	0	1	2	2	0	0	0	0	0	0	0	1	6	2	0	0	0	1	0	0	1	0
p4.G3	ATUE2	P_E sp900591205	GCF_900591205.1	a	13	1	1	2	1	0	0	0	0	0	0	0	1	4	1	0	0	1	0	0	0	1	0
p4.H2	ATUE2	P_E orientalis_A	GCF_002934065.1	b	17	0	1	2	1	0	0	0	2	0	0	0	1	7	1	0	0	0	0	1	0	1	0
p4.H3	ATUE2	P_E canadensis	GCF_000503215.1	a	13	0	2	3	1	0	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0
p5.F2	ATUE2	P_E sivasensis	GCF_013778505.1	a	11	0	1	2	1	1	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p6.B5	ATUE2	P_E lurida	GCF_002563895.1	a	15	0	1	3	1	0	0	1	0	0	0	0	1	6	1	0	0	0	0	0	0	1	0
p6.D10	ATUE2	P_E sivasensis	GCF_013778505.1	a	12	0	1	2	1	1	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0

p6.E11	ATUE2	P_E lurida	GCF_002563895.1	a	12	0	1	3	1	0	0	1	0	0	0	0	1	3	1	0	0	0	0	0	0	1	0
p6.E9	ATUE2	P_E sivasensis	GCF_013778505.1	a	13	0	1	2	1	1	0	0	0	0	1	0	1	4	1	0	0	0	0	1	0	0	0
p6.F1	ATUE2	P_E sp002843605	GCF_002843605.1	a	12	0	1	3	1	0	0	0	0	0	0	0	1	3	1	0	0	1	0	0	0	1	0
p7.E2	ATUE2	P_E sp900583165	GCF_900583165.1	a	14	0	1	2	1	1	0	0	0	0	0	0	1	6	1	0	0	0	1	0	0	0	0
p7.F2	ATUE2	P_E lurida	GCF_002563895.1	a	12	0	1	3	1	0	0	1	0	0	0	0	1	3	1	0	0	0	0	0	0	1	0
p8.A8	ATUE2	P_E sp002979555	GCF_002979555.1	a	9	0	1	2	0	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p8.B4	ATUE2	P_E canadensis	GCF_000503215.1	a	13	0	2	3	1	0	0	0	0	0	0	0	1	5	1	0	0	0	0	0	0	0	0
p8.D10	ATUE2	P_E sivasensis	GCF_013778505.1	a	12	0	2	2	1	1	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p8.D11	ATUE2	P_E canadensis	GCF_000503215.1	a	12	0	2	3	1	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
p8.D3	ATUE2	P_E salomonii	GCF_900107155.1	a	14	0	1	2	1	0	0	1	0	0	0	0	1	6	1	0	0	1	0	0	0	0	0
p8.E11	ATUE2	P_E lurida	GCF_002563895.1	a	12	0	1	3	1	0	0	1	0	0	0	0	1	3	1	0	0	0	0	0	0	1	0
p8.E3	ATUE2	P_E salomonii	GCF_900107155.1	a	14	0	1	2	1	0	0	1	0	0	0	0	1	6	1	0	0	1	0	0	0	0	0
p8.G10	ATUE2	P_E orientalis_A	GCF_002934065.1	b	16	0	1	2	2	0	0	0	0	0	0	0	1	6	2	0	0	0	1	0	0	1	0
p8.H3	ATUE2	P_E canadensis	GCF_000503215.1	a	12	0	2	3	1	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
p2.G5	ATUE3	P_E avellanae	GCF_000444135.1	a	14	0	1	0	0	0	0	0	0	0	0	0	1	9	1	0	0	1	1	0	0	0	0
p6.G1	ATUE3	P_E avellanae	GCF_000444135.1	a	13	0	1	0	0	0	0	0	0	0	0	0	1	8	1	0	0	1	1	0	0	0	0
p4.H5	ATUE3	P_E avellanae	GCF_000444135.1	a	12	0	1	0	0	0	0	0	0	0	0	0	1	6	2	0	0	1	1	0	0	0	0
p8.H2	ATUE3	P_E avellanae	GCF_000444135.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	1	0	0	0	0
p8.H5	ATUE3	P_E avellanae	GCF_000444135.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	4	2	0	0	1	1	0	0	0	0
p2.H6	ATUE3	P_E avellanae	GCF_000444135.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	1	0	0	0	0

p4.G1	ATUE3	P_E avellanae	GCF_000444135.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	1	0	0	0	0
p6.G2	ATUE3	P_E avellanae	GCF_000444135.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	5	0	0	0	1	1	0	0	0	0
p8.G1	ATUE3	P_E avellanae	GCF_000444135.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	1	0	0	0	0
p8.G5	ATUE3	P_E avellanae	GCF_000444135.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	1	0	0	0	0
p11.G1	ATUE3	P_E avellanae	GCF_000444135.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	1	0	0	0	0
p11.G5	ATUE3	P_E avellanae	GCF_000444135.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	1	0	0	0	0
p11.H1	ATUE3	P_E avellanae	GCF_000444135.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	1	0	0	0	0
p2.G6	ATUE3	P_E avellanae	GCF_000444135.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	4	0	0	0	1	1	0	0	0	0
p8.H1	ATUE3	P_E avellanae	GCF_000444135.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	1	1	0	0	0	0
p13.B5	ATUE3	P_E congelans	GCF_900103225.1	a	23	0	1	0	0	0	0	0	0	0	0	0	0	18	3	0	0	1	0	0	0	0	0
p5.A5	ATUE3	P_E congelans	GCF_900103225.1	a	11	0	1	0	0	0	0	0	0	0	0	0	1	6	2	0	0	1	0	0	0	0	0
p23.D8	ATUE3	P_E congelans	GCF_900103225.1	a	10	0	1	0	0	0	0	0	0	0	0	0	1	6	1	0	0	1	0	0	0	0	0
p2.C11	ATUE3	P_E congelans	GCF_900103225.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	2	0	0	1	0	0	0	0	0
p23.F4	ATUE3	P_E congelans	GCF_900103225.1	a	9	0	1	0	0	0	0	0	0	0	0	0	1	4	2	0	0	1	0	0	0	0	0
p9.D10	ATUE3	P_E sp002699985	GCF_002699985.1	a	4	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
p7.A9	ATUE3	P_E sp002699985	GCF_002699985.1	a	3	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
p4.B4	ATUE3	P_E sp900590755	GCF_900590755.1	a	7	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
p8.G2	ATUE3	P_E sp900601905	GCF_900601905.1	a	6	0	1	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0
p8.G8	ATUE3	P_E sp900602065	GCF_900602065.1	a	5	0	1	1	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0
p8.H9	ATUE3	P_E sp900602065	GCF_900602065.1	a	4	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0

p26.C10	ATUE3	P_E syringae	GCF_000507185.2	a	25	0	1	0	0	0	0	0	0	0	0	0	1	18	2	0	0	1	1	0	0	0	1
p23.A4	ATUE3	P_E syringae	GCF_000507185.2	a	24	0	1	0	0	0	0	0	0	0	0	0	1	16	3	0	0	1	1	0	0	0	1
p11.A4	ATUE4	P_E atacamensis	GCF_004801935.1	b	9	0	1	2	1	0	0	1	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p11.C5	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p12.F12	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p12.H9	ATUE4	P_E atacamensis	GCF_004801935.1	b	10	0	1	2	1	0	0	0	0	1	0	0	1	2	1	0	0	0	0	1	0	0	0
p13.G10	ATUE4	P_E sp900582625	GCF_900582625.1	a	9	0	2	1	0	0	1	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0
p13.H9	ATUE4	P_E sp900582625	GCF_900582625.1	a	9	0	2	1	0	0	1	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0
p2.C4	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p3.C5	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p4.B5	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p4.C5	ATUE4	P_E atacamensis	GCF_004801935.1	b	9	0	1	2	1	0	0	0	0	0	0	0	1	3	1	0	0	0	0	0	0	0	0
p6.C5	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p6.G3	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p6.H5	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p7.G9	ATUE4	P_E sp900573885	GCF_900573885.1	a	8	0	2	1	0	0	0	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0

p7.H9	ATUE4	P_E sp900582625	GCF_900582625.1	a	9	0	2	1	0	0	1	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0
p9.D4	ATUE4	P_E atacamensis	GCF_004801935.1	b	8	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0
p9.H10	ATUE4	P_E atacamensis	GCF_004801935.1	b	9	0	1	2	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0	0
p9.H8	ATUE4	P_E sp900582625	GCF_900582625.1	a	9	0	2	1	0	0	1	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0
p9.H9	ATUE4	P_E sp900573885	GCF_900573885.1	a	8	0	2	1	0	0	0	0	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0
p11.C6	ATUE6	P_E coleopterorum	GCF_900105555.1	a	8	0	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1	1	0	0
p11.G2	ATUE6	P_E sp900596015	GCA_900596015.1	a	5	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0
p11.H3	ATUE6	P_E baltica	GCF_014235765.1	a	7	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1	1	0	0
p2.E4	ATUE6	P_E coleopterorum	GCF_900105555.1	a	5	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0
p2.G1	ATUE6	P_E sp900596015	GCA_900596015.1	a	6	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	1	1	0	0
p2.H2	ATUE6	P_E sp900596015	GCA_900596015.1	a	4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0
p23.A5	ATUE6	P_E sp900585815	GCF_013201545.1	a	6	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0
p23.B4	ATUE6	P_E sp900585815	GCF_013201545.1	a	5	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0
p6.G6	ATUE6	P_E sp900596015	GCA_900596015.1	a	4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0
p1.E6	ATUE7	P_E sp900589395	GCF_900589395.1	a	7	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p13.F5	ATUE7	P_E sp900589395	GCF_900589395.1	a	8	0	1	0	0	0	0	0	0	0	0	0	1	4	1	0	0	0	0	1	0	0	0

p2.C9	ATUE7	P_E sp900589395	GCF_900589395.1	a	7	0	1	0	0	0	0	0	0	0	0	0	1	3	1	0	0	0	0	1	0	0	0
p9.C4	ATUE7	P_E ovata	GCF_003131185.1	a	11	0	1	0	0	0	1	0	0	0	0	0	1	4	2	0	0	0	0	1	0	1	0
p4.G2	ATUE8	P_E synxantha_A	GCF_000263715.2	a	15	0	1	2	1	0	0	0	0	0	0	0	1	8	0	0	0	1	0	1	0	0	0
p2.B6	ATUE9	P_E asturiensis	GCF_900143095.1	a	10	0	1	0	0	0	0	0	0	0	0	1	1	4	1	0	0	1	0	1	0	0	0

* a: taxonomic classification defined by topology and ANI; b: ANI

ANI: average nucleotide identity, CDPS: tRNA-dependent cyclodipeptide synthases, hserlactone: homoserine lactone, LAP: Linear azol(in)e-containing peptide, NA: not applicable, NAGGN: N-acetylglutaminylglutamine amide, NRPS: non-ribosomal peptide synthetases, PKS: polyketide synthases, P_E: *Pseudomonas*_E

Table S3. Number of predicted biosynthetic gene clusters per ATUE

ATUE	n	Mean	Median	Min	Max	IQR
ATUE1	8	9,8	10,0	7	14	4.25
ATUE2	49	13,0	13,0	8	20	3.00
ATUE3	28	10,2	9,0	3	25	3.00
ATUE4	19	8,5	8,0	8	10	1.00
ATUE5	163	10,7	9,0	7	25	4.00
ATUE6	9	5,6	5,0	4	8	1.00
ATUE7	4	8,3	7,5	7	11	1.75
ATUE8	1	15,0	15,0	15	15	0.00
ATUE9	1	10,0	10,0	10	10	0.00
ATUE10	1	12,0	12,0	12	12	0.00
ATUE11	1	11,0	11,0	11	11	0.00

Table S4. Green pixels measured 7 days post-infection with 75 different *P. viridiflava* ATUE5 isolates

Isolate	Mean	Standard error
p1.A2	55074.0	3089.3
p1.A3	17387.5	3908.1
p1.D11	53636.6	2077.5
p1.D2	50165.9	1719.4
p1.E3	49532.0	2651.3
p1.F8	47485.1	3170.2
p1.G2	47152.3	2446.5
p1.G5	46942.3	2848.9
p1.H1	46152.4	2363.5
p13.E11	43911.8	1701.5
p3.A3	4328.6	955.8
p4.A9	3726.0	1224.7
p4.B7	3119.3	415.6
p4.D2	2790.3	790.2
p4.E5	2652.0	639.8
p4.F10	2431.5	635.5
p4.F5	2313.4	693.6
p4.G8	10933.9	1143.9
p6.A10	1851.0	579.1
p6.A9	1839.8	474.5
p6.B9	1039.1	180.0
p6.D1	1013.3	202.7
p6.D6	965.6	191.9
p6.D8	406.9	86.9
p6.E1	841.0	101.8
p7.G11	799.1	150.8
p7.G6	714.6	180.8
p8.B2	444.8	103.9
p8.B3	421.0	65.0
p8.B9	7384.1	1796.4
p8.C7	351.8	53.0
p8.D5	246.1	16.6
p8.E5	191.3	12.4
p13.E3	43081.6	2440.6
p13.F2	40724.3	2185.0
p13.G2	38298.4	2596.5
p13.G4	34247.3	2141.7
p13.G8	33270.8	3211.2

p13.H5	7265.4	1403.6
p20.B12	28676.1	2391.4
p20.B3	23772.4	2298.4
p20.B8	23240.1	2877.7
p21.E3	20275.8	2488.7
p21.E5	19045.8	2225.7
p21.F4	18360.0	2579.1
p21.H4	853.5	124.4
p22.A8	17239.9	1927.7
p22.B5	14361.0	2902.9
p22.C1	13649.5	2142.8
p22.C3	13294.9	2059.3
p22.D2	11449.5	1834.4
p22.D4	9780.0	1911.7
p22.D7	10697.9	1689.1
p22.E7	10467.8	1911.5
p22.F1	10316.3	2446.0
p22.H4	30742.0	2921.0
p23.A2	9325.9	1639.6
p23.A7	8867.6	2009.5
p23.B1	8281.5	1412.4
p23.B3	7430.3	1866.0
p23.B6	4911.8	1524.7
p23.C2	7349.3	1637.6
p23.C4	1895.1	442.2
p23.C9	6891.6	1222.3
p23.D5	6731.3	1332.2
p24.G4	6306.4	1157.9
p24.H2	6005.6	991.1
p25.A12	5953.3	1138.6
p25.A4	5811.0	1594.1
p25.B2	5739.8	1244.4
p25.C11	5309.3	1205.0
p25.C2	5029.1	1684.2
p25.D2	75093.7	2610.5
p25.E3	4728.8	1066.7
p25.F3	4390.1	1210.6

Discussion

In this thesis, I provide an integrative characterization of the interactions between *Arabidopsis thaliana* and its natural pathogen *Pseudomonas viridiflava*. Despite *P. viridiflava* being a prevalent opportunistic pathogen on *A. thaliana* populations, our knowledge about this pathosystem is limited. Understanding how host and pathogen genetic backgrounds interact upon infection remains largely unexplored, even when a better comprehension of host-pathogen dynamics may enable predictions of disease outcome and the development of new treatment or prevention strategies against phytopathogens.

1. Closely-related pathogens elicit different immune responses

In the first chapter, I focused on the host response to infection using a comprehensive approach that integrated genetics/genomics, transcriptomics and metabolomics. Using multiple experimental approaches, I showed that the JA/ET pathway is involved in defense against *P. viridiflava* in an axenic system. This expands the results by Jabok *et al.* (122) on Midwestern USA *P. viridiflava* to the predominant *Pseudomonas* clade in European *A. thaliana* populations, *P. viridiflava* ATUE5 (118,120). I expanded these results by showing that the ET branch of the JA/ET pathway is the one upregulated upon infection, and suggest that this is mediated by an increase in JA levels, transduced via ERF1 and PDF1.2. Indeed, infection with *P. viridiflava* reduces plant growth, and even more so in the JA-insensitive *coi1-16* mutants (97).

Not only did I describe the host response to *P. viridiflava*, but I also included the model pathogen DC3000 for comparison. I found that compared to the model pathogen, *P. viridiflava* elicited a larger remodeling on the overall host transcriptome and metabolome. *P. syringae* encodes the toxin coronatine, which manipulates the host defense response to promote pathogen growth. Coronatine upregulates the JA signaling pathway, leading to a downregulation of the SA pathway that mediates resistance against DC3000 (32). I confirmed that this upregulation of the JA pathway was mediated by MYC2, as reported before (32), and showed that *P. viridiflava* and *P. syringae* use different mechanisms to induce the JA defense signaling pathway.

2. Host-pathogen interactions have two equally important protagonists

I used a diverse collection of *A. thaliana* and *P. viridiflava* genotypes, which allowed me to demonstrate that genotype-by-genotype interactions modify the outcome of infection. Accordingly, it has been shown that genetic variation in the pathogen and the host influence the pathogen transcriptome strategy upon infection and the host defense system (107,108). A differential interaction between the bacterial genotypes with the host genotypes could explain the persistence of closely-related *P. viridiflava* lineages on *A. thaliana* populations described by Karasov and colleagues (118).

Screening several host and pathogen genotypes enabled me to select what I found interesting for further study. Hence, instead of investigating the host response to *P. viridiflava* in a single *A. thaliana* genotype, as is usually done, I used susceptible and resistant *A. thaliana* genotypes and compared their response to the same pathogen. I found that bacterial growth is slower in the resistant host and that it appeared to be primed in the defense response: the resistant genotype displayed higher expression of defense-related genes at the moment of infection compared to the susceptible host. This resulted in a faster transcriptome response to *P. viridiflava* infection and the establishment of resistance. Furthermore, my results suggest that the susceptible host response is qualitatively similar to that of the resistant host, but it occurs slower and is not enough to stop plant death. A similar pattern was reported for DC3000 infection, with susceptible plants and immunity mutants showing an almost identical transcriptome response to resistant plants but with several hours delay (29).

Although not described here, I have generated data not only in *P. viridiflava* ATUE5::p13.G4, the most virulent of the screened isolates, but also on isolate p13.C1, the least virulent. These results indicate that two closely-related isolates impact the host transcriptome in a different way, and hint at the strategy used by *P. viridiflava* ATUE5::p13.G4 to cause disease: plants infected with this isolate had increased detoxification response, which was not the case for plants infected with p13.C1. These results contribute to the identification of *P. viridiflava* virulence strategies and, combined with knowledge on the host biochemistry and metabolism, could be linked to susceptibility or resistance as well.

3. Plant-associated *Pseudomonas* encode diverse specialized metabolites gene clusters

In the second chapter, I shifted my focus to the pathogen. In particular, I described the biosynthetic potential of plant-associated *Pseudomonas* in terms of the repertoire of specialized metabolites, and identified gene clusters putatively involved in virulence. The *Pseudomonas* genus is well-known for its environmental ubiquity and its ability to produce an ample array of specialized metabolites (73,128). However, only one study focusing on plant-associated bacteria has been carried out, specifically in *P. fluorescens* from the rhizosphere (86).

The large collection of 284 *Pseudomonas* genomes that I analyzed proved to have a vast biosynthetic potential, encoding specialized metabolites from more than 20 different classes. I found differences in the distribution of the biosynthetic gene cluster among clades, which could be related to the different environmental niches, neighbors and/or host of each of them (71,129). I highlighted the diversity of terpenes since this is both the largest and the most structurally diverse class of specialized metabolites (130), and it is sparsely found in *Pseudomonas* (85,128). I found 19 terpene gene cluster families, present in 73% of the isolates throughout the phylogeny. This is a larger number of terpenes than reported before in the genus (85,128). In 4 of the 10

clades where terpenes were detected, 50% or less of the isolates encoded one. This underscores the strength of my approach, where I looked into more than a single isolate per species to uncover the complete potential of a species or genus. Such an approach is not the norm in most of the published literature, where only one or a few representatives for each species are analyzed. An interesting next step would be to perform rarefaction analyses to determine when the complete biosynthetic potential of a clade is discovered.

4. Strengths and opportunities for improvement

4.1. A natural pathosystem

The *A. thaliana* - *P. viridiflava* is a natural pathosystem. Contrary to other *P. syringae* strains frequently used for phytopathology studies, such as *P. syringae* pv. *tomato* DC3000 and *P. syringae* pv. *syringae* B728a, *P. viridiflava* infects wild *A. thaliana* populations, and has been isolated from plants in the US, Europe and Japan (88,92,97,120,131). Thus, it is well-suited to ask ecologically-relevant questions about adaptation and co-evolution with a natural host.

Based on the ratio of susceptible and resistant individuals from an F_2 population, I posit that resistance to *P. viridiflava* is due to a recessive mechanism. Recessive mutations in genes that encode negative regulators of defense signaling pathways are common and can result in loss of susceptibility (132,133); this would be consistent with a primed defense status of the resistant host Sij-4. If this hypothesis turns out to be true, this would be the first report of recessive resistance to a bacterial pathogen in *A. thaliana*.

4.2. Genetic variation and high throughput

The work I presented here highlights the importance of host and pathogen genetic variation, and prompts for their inclusion in future studies. Only then can we understand the spectrum of defense and virulence mechanisms, which can be specific to certain genotypes (63). Currently, plant-pathogen interaction studies that include genetic variation usually do so for the host only (63,64). In contrast, I included both host and pathogen diversity, and found that resistance and virulence in the *A. thaliana* - *P. viridiflava* pathosystem are continuous, quantitative traits. Including host genetic diversity instead of a single reference genotype allowed me to select susceptible and resistant hosts to characterize the defense response of *A. thaliana*. To the best of my knowledge, my work is the first to use transcriptomics and metabolomics to study the natural *A. thaliana* - *P. viridiflava* pathosystem; few studies in phytopathology combine both approaches. The data me and my collaborators generated should be made publicly available so that other researchers can benefit from it.

I also took advantage of genetic variation when annotating specialized metabolites in *Pseudomonas*, which can act as virulence factors. By correlating the presence/absence of nonribosomal peptide synthetase (NRPS) families in 75 *P.*

viridiflava ATUE5 isolates with plant size after infection with each isolate, I identified one family whose presence increases virulence, and several others with the opposite effect. Although the role of these families in virulence has to be experimentally confirmed, this computational approach leveraged genetic variation in the pathogen to narrow candidates' biosynthetic gene clusters, and can be easily expanded to include other *Pseudomonas* clades or species.

Yet another strength of my work was the use of methods that allowed me to carry out high throughput screens. Employing luminescence as a proxy for bacterial load and the number of green pixels as a proxy for plant size enabled me to perform high-throughput experiments, such as screening 252 host x pathogen genotype combinations or phenotyping hundreds of F_2 individuals for genetic mapping. This would have been painstaking and time-consuming with the more traditional colony counting or weighting the plants. Another technical advantage of my work is the drip-inoculation of the plants without surfactants, which resembles natural infection more than the commonly used syringe infiltration. Although results can be more variable with drip-inoculation than with other infection methods, especially in the absence of surfactants, I believe this is worth it in order to be able to recapitulate all the factors involved in the infection process, including the first obstacle leaf pathogens have to overcome: colonizing the leaf and penetrating into the apoplast. Early in my project, I worried about the inter-treatment variation upon infection. I got access to the data of a colleague who performed similar experiments to me, but using flood inoculation and Silwet, which is expected to increase reproducibility; I found that the variation in my data was similar to theirs, and since them embrace the complexity of the infection process as a feature instead of a bug.

4.3. Pitfalls and shortcomings

My work, however, is not without limitations. First, the infection outcome was measured at a single time point except for transcriptomics, and the results could be different at other time points, reflecting different stages of the infection process. For example, the classification of host genotypes as resistant or susceptible could change when other time points are studied. To rule out that this was the case for the resistant host genotype I used throughout the first chapter, I followed the infected plants for two weeks without observing susceptibility.

I used an axenic system, meaning the only microorganism present was the isolate used for infection. This means that the plant lacked their natural microbiome, which is known to affect susceptibility and resistance to disease (134). In addition, competing microorganisms on the leaf surface, including other pathogens, could also modify *P. viridiflava* virulence (97,98). Nevertheless, I believe this set up is useful to gain insights into the interactions between host and pathogen, particularly when not much is known as was the case for this work. I focused on *A. thaliana* hormone defense pathways, but it is not unlikely that other host pathways or processes play a role in infection. Fortunately, the datasets generated here can be further analyzed and re-analyzed to

gain more insights into the interaction between *A. thaliana* and *P. viridiflava* and to compare with observations from future studies.

The use of an understudied pathosystem like the one I used is both a blessing and a curse. On the one hand, most of the findings are novel, interesting and expand our knowledge; on the other, there is little literature to compare with. My results did not always correspond to what I expected based on the research done with *P. syringae* pv. *tomato* DC3000 and/or with soil-grown plants, and I could not determine whether this was due to the object of study, the researcher, or the interaction between them. All these shortcomings notwithstanding, the consistency of the results I obtained through different approaches gives me confidence in the results I have described.

5. Outlook

Taken together, my work provides a baseline for future studies in the *A. thaliana* - *P. viridiflava* pathosystem. Just like with any phenomenon under study, there is much more to explore in this pathosystem, including questions arising from this thesis.

For instance, further studies are needed to identify the genetic basis of resistance in this pathosystem. My results suggest resistance is a recessive trait, but I was not able to map the gene(s) involved. I used an F_2 population derived from a highly susceptible and a highly resistant parent, but other parental genotypes can be selected based on the axenic screen I presented. In addition, the conservation of the resistance mechanisms I propose, namely a basal upregulation of the defense response, needs to be addressed. This is important since it is known that *A. thaliana* genotypes have different mechanisms of resistance against the same pathogen (63).

If indeed the resistance I observed to *P. viridiflava* is due to higher basal levels of JA, this host genotype should be resistant to other pathogens for which resistance is mediated by this pathway, namely necrotrophs (18), such as the fungal pathogen *Botrytis cinerea*. Conversely, this host genotype should be more susceptible to biotrophic pathogens, provided that increased JA results in decreased SA, the main mediator of resistance against biotrophs. These phenotypes are easy to measure and could help elucidate the ecological advantage of increased JA levels, especially in the face of the defense-growth trade-off (135).

With the knowledge I have generated about the infection process of *P. viridiflava*, it would be ideal to generate metabolomics data from earlier stages of infection, which can be temporarily correlated with the transcriptomic changes observed already at 16 hpi. Furthermore, increased basal levels of the JA hormone in the resistant host would lend support to the mechanisms suggested here. What I hope is most evident from my thesis is that both the host and pathogen genotypes affect the infection outcome, and their interaction can result in two closely-related genotypes having different

phenotypes. It is then important to include this genetic variation in future studies so that the complete spectrum of resistance and virulence can be discovered.

Although it is now known that *P. viridiflava* is common throughout Europe (120), the genomes used in this thesis, and in fact most of the genomes publicly available for *P. viridiflava*, correspond to isolates from Southwest Germany. It will be interesting to generate complete genomes from other geographic locations and annotate their virulence factors, including specialized metabolites. High-quality genomes, ideally closed, are required for accurate annotation of these specialized metabolites.

The prediction of biosynthetic gene clusters can point to metabolites implicated in virulence but, to do so efficiently, a better chemical and biological characterization of specialized metabolites is required. Future studies should focus not only on the computational prediction of biosynthetic gene clusters, but should also aim to structurally and/or functionally characterize the product from these gene clusters. This is imperative to really understand the mechanisms *P. viridiflava* and other pathogens use to cause disease. During the execution of my project, I established a collaboration with Prof. Dr. Harald Groß, an expert in *Pseudomonas* specialized metabolites. His group is currently working on the characterization of the two new lipopeptides from *P. viridiflava* identified here.

My work provides biosynthetic gene clusters potentially involved in *P. viridiflava* pathogenicity on *A. thaliana*. These candidates require experimental validation, such as knock-out and knock-in studies, to confirm their role in virulence. In addition, data about the pathogenicity of a larger collection of *Pseudomonas* isolates, including not only ATUE5 but other clades belonging to the *P. syringae* group, can contribute to the understanding of the virulence mechanisms of *P. viridiflava*. It is still a mystery why some *P. viridiflava* isolates are more virulent than others, and comparative genomic analyses might help elucidate this. Comparisons need to be made not only within *P. viridiflava*/ATUE5, but also between this clade and others described. *P. viridiflava* is a generalist pathogen, and it lacks most of the effectors and toxins that contribute to virulence on other *Pseudomonas.* I believe the combination of genomics and phenotypic data, as I did here, will provide clues about this mystery, as it has done for others (136). The ultimate test will be to assess how the findings in an axenic system translate to the complex environment plants and bacteria encounter outside the lab.

All in all, the work I presented here is at the forefront of research in plant-microbe interaction. It does not only recognize the importance of host and pathogen genotypes, and their interaction; it also integrates several methodologies to produce a comprehensive picture of the *A. thaliana* - *P. viridiflava* pathosystem, where both protagonists are studied. Just as I proposed at the beginning of this thesis, I have shed light on the resistance and virulence mechanisms in this pathosystem, and provided the baseline for future studies.

References

1. Laibachs F. Arabidopsis thaliana (L.) HEYNH. als Objekt fur genetische und entwicklungsphysiologische Untersuchungen. Bot Arch. 1943;44:439–55.
2. Koornneef M, Meinke D. The development of Arabidopsis as a model plant. Plant J. 2010 Mar;61(6):909–21.
3. Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, et al. 50 years of Arabidopsis research: highlights and future directions. New Phytol. 2016 Feb;209(3):921–44.
4. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, et al. The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol. 2005 Jul;3(7):e196.
5. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000 Dec 14;408(6814):796–815.
6. Laibach F. Über sommer-und winterannuelle Rassen von Arabidopsis thaliana (L.) Heynh. Ein Beitrag zur Ätiologie der Blütenbildung. Beitr Biol Pflanzen. 1951;28:173–210.
7. Kugler I. Untersuchungen über das Keimverhalten einiger Rassan van Arabidopsis thaliana (L.) Heynh. beitrag zur Ätiologie der Blütenbildung. Beitr Biol Pflanzen. 1951;28:173–210.
8. Weigel D. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol. 2012 Jan;158(1):2–22.
9. Alonso-Blanco C, Koornneef M. Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends Plant Sci. 2000 Jan;5(1):22–9.
10. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016 Jul 14;166(2):481–91.
11. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet. 2011 Aug 28;43(10):956–63.
12. Chae E, Bomblies K, Kim ST, Karelina D, Zaidem M, Ossowski S, et al. Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell. 2014 Dec 4;159(6):1341–51.
13. Alonso JM, Ecker JR. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nat Rev Genet. 2006 Jul;7(7):524–36.
14. Ngou BPM, Ding P, Jones JDG. Thirty years of resistance: Zig-zag through the plant immune system. Plant Cell. 2022 Apr 26;34(5):1447–78.
15. Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, et al. A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. Cell. 2019 Aug 22;178(5):1260–72.e14.
16. Jones JDG, Dangl JL. The plant immune system. Nature. 2006 Nov 16;444(7117):323–9.
17. Bi G, Su M, Li N, Liang Y, Dang S, Xu J, et al. The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. Cell. 2021 Jun 24;184(13):3528–41.e12.
18. Glazebrook J. Contrasting mechanisms of defense against biotrophic and

necrotrophic pathogens. Annu Rev Phytopathol. 2005;43:205–27.

19. Reymond P, Weber H, Damond M, Farmer EE. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell. 2000 May;12(5):707–20.
20. Lefevere H, Bauters L, Gheysen G. Salicylic Acid Biosynthesis in Plants. Front Plant Sci. 2020 Apr 17;11:1261.
21. Ruan J, Zhou Y, Zhou M, Yan J, Khurshid M, Weng W, et al. Jasmonic Acid Signaling Pathway in Plants. Int J Mol Sci. 2019 May 20;20(10).
22. Wasternack C, Song S. Jasmonates: biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. J Exp Bot. 2017 Mar 1;68(6):1303–21.
23. Li N, Han X, Feng D, Yuan D, Huang LJ. Signaling Crosstalk between Salicylic Acid and Ethylene/Jasmonate in Plant Defense: Do We Understand What They Are Whispering? Int J Mol Sci. 2019 Feb 4;20(3).
24. Binder BM. Ethylene signaling in plants. J Biol Chem. 2020 May 29;295(22):7710–25.
25. Checker VG, Kushwaha HR, Kumari P, Yadav S. Role of Phytohormones in Plant Defense: Signaling and Cross Talk. In: Singh A, Singh IK, editors. Molecular Aspects of Plant-Pathogen Interaction. Singapore: Springer Singapore; 2018. p. 159–84.
26. Mur LAJ, Kenton P, Atzorn R, Miersch O, Wasternack C. The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. Plant Physiol. 2006 Jan;140(1):249–62.
27. Betsuyaku S, Katou S, Takebayashi Y, Sakakibara H, Nomura N, Fukuda H. Salicylic Acid and Jasmonic Acid Pathways are Activated in Spatially Different Domains Around the Infection Site During Effector-Triggered Immunity in Arabidopsis thaliana. Plant Cell Physiol. 2018 Jan 1;59(1):8–16.
28. Zhang N, Zhou S, Yang D, Fan Z. Revealing Shared and Distinct Genes Responding to JA and SA Signaling in Arabidopsis by Meta-Analysis. Front Plant Sci. 2020 Jun 26;11:908.
29. Mine A, Seyfferth C, Kracher B, Berens ML, Becker D, Tsuda K. The Defense Phytohormone Signaling Network Enables Rapid, High-Amplitude Transcriptional Reprogramming during Effector-Triggered Immunity. Plant Cell. 2018 Jun;30(6):1199–219.
30. Caarls L, Pieterse CMJ, Van Wees SCM. How salicylic acid takes transcriptional control over jasmonic acid signaling. Front Plant Sci. 2015 Mar 25;6:170.
31. Hou S, Tsuda K. Salicylic acid and jasmonic acid crosstalk in plant immunity. Essays Biochem. 2022 Jun 14.
32. Zheng XY, Spivey NW, Zeng W, Liu PP, Fu ZQ, Klessig DF, et al. Coronatine promotes Pseudomonas syringae virulence in plants by activating a signaling cascade that inhibits salicylic acid accumulation. Cell Host Microbe. 2012 Jun 14;11(6):587–96.
33. Ton J, Flors V, Mauch-Mani B. The multifaceted role of ABA in disease resistance. Trends Plant Sci. 2009 Jun;14(6):310–7.
34. Berens ML, Wolinska KW, Spaepen S, Ziegler J, Nobori T, Nair A, et al. Balancing

trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. Proc Natl Acad Sci U S A. 2019 Feb 5;116(6):2364–73.

35. Denancé N, Sánchez-Vallet A, Goffner D, Molina A. Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. Front Plant Sci. 2013 May 24;4:155.
36. Koskella B. The phyllosphere. Curr Biol. 2020 Oct 5;30(19):R1143–6.
37. Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012 Dec;10(12):828–40.
38. Xu N, Zhao Q, Zhang Z, Zhang Q, Wang Y, Qin G, et al. Phyllosphere Microorganisms: Sources, Drivers, and Their Interactions with Plant Hosts. J Agric Food Chem. 2022 Apr 27;70(16):4860–70.
39. Passarelli-Araujo H, Franco GR, Venancio TM. Network analysis of ten thousand genomes shed light on Pseudomonas diversity and classification. Microbiol Res. 2022 Jan;254:126919.
40. Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E. Phylogenomics and systematics in Pseudomonas. Front Microbiol. 2015 Mar 18;6:214.
41. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005 Feb 15;102(7):2567–72.
42. Baltrus DA, McCann HC, Guttman DS. Evolution, genomics and epidemiology of Pseudomonas syringae: Challenges in Bacterial Molecular Plant Pathology. Mol Plant Pathol. 2017 Jan;18(1):152–68.
43. Xin XF, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. Nat Rev Microbiol. 2018 May;16(5):316–28.
44. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol. 2012 Aug;13(6):614–29.
45. Morris CE, Lamichhane JR, Nikolić I, Stanković S, Moury B. The overlapping continuum of host range among strains in the Pseudomonas syringae complex. Phytopathology Research. 2019 Jan 16;1(1):1–16.
46. Morris CE, Monteil CL, Berge O. The life history of Pseudomonas syringae: linking agriculture to earth system processes. Annu Rev Phytopathol. 2013 May 6;51:85–104.
47. Dillon MM, Thakur S, Almeida RND, Wang PW, Weir BS, Guttman DS. Recombination of ecologically and evolutionarily significant loci maintains genetic cohesion in the Pseudomonas syringae species complex. Genome Biol. 2019 Jan 3;20(1):3.
48. Berge O, Monteil CL, Bartoli C, Chandeysson C, Guilbaud C, Sands DC, et al. A user's guide to a data base of the diversity of Pseudomonas syringae and its application to classifying strains in this phylogenetic complex. PLoS One. 2014 Sep 3;9(9):e105547.
49. Araki H, Tian D, Goss EM, Jakob K, Halldorsdottir SS, Kreitman M, et al. Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. Proc Natl Acad Sci U S A. 2006 Apr 11;103(15):5887–92.

50. Bartoli C, Berge O, Monteil CL, Guilbaud C, Balestra GM, Varvaro L, et al. The Pseudomonas viridiflava phylogroups in the P. syringae species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. Environ Microbiol. 2014 Jul;16(7):2301–15.
51. O'Brien HE, Thakur S, Guttman DS. Evolution of plant pathogenesis in Pseudomonas syringae: a genomics perspective. Annu Rev Phytopathol. 2011;49:269–89.
52. Xin XF, Nomura K, Aung K, Velásquez AC, Yao J, Boutrot F, et al. Bacteria establish an aqueous living space in plants crucial for virulence. Nature. 2016 Nov 24;539(7630):524–9.
53. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. Nature. 2001 Jun 14;411(6839):826–33.
54. Deslandes L, Rivas S. Catch me if you can: bacterial effectors and plant targets. Trends Plant Sci. 2012 Nov;17(11):644–55.
55. Dillon MM, Almeida RND, Laflamme B, Martel A, Weir BS, Desveaux D, et al. Molecular Evolution of Pseudomonas syringae Type III Secreted Effector Proteins. Front Plant Sci. 2019 Apr 5;10:418.
56. Bender Carol L., Alarcón-Chaidez Francisco, Gross Dennis C. Pseudomonas syringae Phytotoxins: Mode of Action, Regulation, and Biosynthesis by Peptide and Polyketide Synthetases. Microbiol Mol Biol Rev. 1999 Jun 1;63(2):266–92.
57. Stringlis IA, Zhang H, Pieterse CMJ, Bolton MD, de Jonge R. Microbial small molecules - weapons of plant subversion. Nat Prod Rep. 2018 May 25;35(5):410–33.
58. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, et al. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 Pseudomonas syringae isolates. PLoS Pathog. 2011 Jul;7(7):e1002132.
59. O'Neill EM, Mucyn TS, Patteson JB, Finkel OM, Chung EH, Baccile JA, et al. Phevamine A, a small molecule that suppresses plant immune responses. Proc Natl Acad Sci U S A. 2018 Oct 9;115(41):E9514–22.
60. Schellenberg B, Ramel C, Dudler R. Pseudomonas syringae virulence factor syringolin A counteracts stomatal immunity by proteasome inhibition. Mol Plant Microbe Interact. 2010 Oct;23(10):1287–93.
61. Arrebola E, Cazorla FM, Perez-García A, de Vicente A. Chemical and metabolic aspects of antimetabolite toxins produced by Pseudomonas syringae pathovars. Toxins . 2011 Sep;3(9):1089–110.
62. Xin XF, He SY. Pseudomonas syringae pv. tomato DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants. Annu Rev Phytopathol. 2013 May 31;51:473–98.
63. Velásquez AC, Oney M, Huot B, Xu S, He SY. Diverse mechanisms of resistance to Pseudomonas syringae in a thousand natural accessions of Arabidopsis thaliana. New Phytol. 2017 Jun;214(4):1673–87.
64. Bartoli C, Roux F. Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. Front Plant Sci. 2017 May 23;8:763.
65. Cunnac S, Chakravarthy S, Kvitko BH, Russell AB, Martin GB, Collmer A. Genetic disassembly and combinatorial reassembly identify a minimal functional

repertoire of type III effectors in Pseudomonas syringae. Proc Natl Acad Sci U S A. 2011 Feb 15;108(7):2975–80.

66. Hu Y, Ding Y, Cai B, Qin X, Wu J, Yuan M, et al. Bacterial effectors manipulate plant abscisic acid signaling for creation of an aqueous apoplast. Cell Host Microbe. 2022 Apr 13;30(4):518–29.e6.
67. Roussin-Léveillée C, Lajeunesse G, St-Amand M, Veerapen VP, Silva-Martins G, Nomura K, et al. Evolutionarily conserved bacterial effectors hijack abscisic acid signaling to induce an aqueous environment in the apoplast. Cell Host Microbe. 2022 Apr 13;30(4):489–501.e4.
68. Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A, et al. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc Natl Acad Sci U S A. 2005 Aug 2;102(31):11064–9.
69. Hutchison ML, Tester MA, Gross DC. Role of biosurfactant and ion channel-forming activities of syringomycin in transmembrane ion flux: a model for the mechanism of action in the plant-pathogen interaction. Mol Plant Microbe Interact. 1995 Jul;8(4):610–20.
70. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, et al. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. Proc Natl Acad Sci U S A. 2003 Sep 2;100(18):10181–6.
71. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. MBio. 2020 Jun 16;11(3).
72. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol. 2015 Sep;11(9):625–31.
73. Gross H, Loper JE. Genomics of secondary metabolite production by Pseudomonas spp. Nat Prod Rep. 2009 Nov;26(11):1408–46.
74. Etalo DW, Jeon JS, Raaijmakers JM. Modulation of plant chemistry by beneficial root microbiota. Nat Prod Rep. 2018 May 1;35(5):398–409.
75. Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. Natural functions of lipopeptides from Bacillus and Pseudomonas: more than surfactants and antibiotics. FEMS Microbiol Rev. 2010 Nov;34(6):1037–62.
76. De Meyer G, Capieau K, Audenaert K, Buchala A, Métraux JP, Höfte M. Nanogram amounts of salicylic acid produced by the rhizobacterium Pseudomonas aeruginosa 7NSK2 activate the systemic acquired resistance pathway in bean. Mol Plant Microbe Interact. 1999 May;12(5):450–8.
77. Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol Rev. 2007 Jul;31(4):425–48.
78. Vinatzer BA, Yan S. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol Plant Pathol. 2008 Jan;9(1):105–18.
79. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W339–46.

80. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021 Jul 2;49(W1):W29–35.
81. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014 Jul 17;158(2):412–21.
82. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020 Jan;16(1):60–8.
83. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 2020 Jan 8;48(D1):D454–8.
84. Götze S, Stallforth P. Structure, properties, and biological functions of nonribosomal lipopeptides from pseudomonads. Nat Prod Rep. 2020 Jan 1;37(1):29–54.
85. Alam K, Islam MM, Li C, Sultana S, Zhong L, Shen Q, et al. Genome Mining of Pseudomonas Species: Diversity and Evolution of Metabolic and Biosynthetic Potential. Molecules. 2021 Dec 12;26(24).
86. Rieusset L, Rey M, Muller D, Vacheron J, Gerin F, Dubost A, et al. Secondary metabolites from plant-associated Pseudomonas are overproduced in biofilm. Microb Biotechnol. 2020 Sep;13(5):1562–80.
87. Baldeweg F, Hoffmeister D, Nett M. A genomics perspective on natural product biosynthesis in plant pathogenic bacteria. Nat Prod Rep. 2019 Feb 20;36(2):307–25.
88. Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J. Pseudomonas viridiflava and P. syringae--natural pathogens of Arabidopsis thaliana. Mol Plant Microbe Interact. 2002 Dec;15(12):1195–203.
89. Ishiga Y, Ishiga T, Uppalapati SR, Mysore KS. Arabidopsis seedling flood-inoculation technique: a rapid and reliable assay for studying plant-bacterial interactions. Plant Methods. 2011 Oct 6;7:32.
90. Katagiri F, Thilmony R, He SY. The Arabidopsis thaliana-pseudomonas syringae interaction. Arabidopsis Book. 2002 Mar 27;1:e0039.
91. Karasov TL, Neumann M, Duque-Jaramillo A, Kersten S, Bezrukov I, Schröppel B, et al. The relationship between microbial population size and disease in the Arabidopsis thaliana phyllosphere. bioRxiv. 2020. p. 828814. doi: 10.1101/828814v2
92. Goss EM, Bergelson J. Fitness consequences of infection of Arabidopsis thaliana with its natural bacterial pathogen Pseudomonas viridiflava. Oecologia. 2007 May;152(1):71–81.
93. Goss EM, Bergelson J. Variation in resistance and virulence in the interaction between Arabidopsis thaliana and a bacterial pathogen. Evolution. 2006 Aug;60(8):1562–73.
94. Pagán I, García-Arenal F. Tolerance of Plants to Pathogens: A Unifying View. Annu Rev Phytopathol. 2020 Aug 25;58:77–96.
95. Kover PX, Schaal BA. Genetic variation for disease resistance and tolerance among Arabidopsis thaliana accessions. Proc Natl Acad Sci U S A. 2002 Aug

20;99(17):11270–4.

96. Shalev O, Ashkenazy H, Neumann M, Weigel D. Commensal Pseudomonas protect Arabidopsis thaliana from a coexisting pathogen via multiple lineage-dependent mechanisms. ISME J. 2022 May;16(5):1235–44.
97. Lundberg DS, de Pedro Jové R, Ayutthaya PPN, Karasov TL, Shalev O, Poersch K, et al. Contrasting patterns of microbial dominance in the Arabidopsis thaliana phyllosphere. bioRxiv. 2021. p. 2021.04.06.438366. doi: 10.1101/2021.04.06.438366v2
98. Shalev O, Karasov TL, Lundberg DS, Ashkenazy H, Pramoj Na Ayutthaya P, Weigel D. Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant. Nat Ecol Evol. 2022 Apr;6(4):383–96.
99. Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, et al. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods. 2005 Jun;2(6):443–8.
100. Choi KH, Schweizer HP. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. Nat Protoc. 2006;1(1):153–61.
101. Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. Elife. 2013 Oct 29;2:e01123.
102. Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H, et al. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. Proc Natl Acad Sci U S A. 2018 Jun 12;115(24):E5440–9.
103. Lees JA, Ferwerda B, Kremer PHC, Wheeler NE, Serón MV, Croucher NJ, et al. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. Nat Commun. 2019 May 15;10(1):2176.
104. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nat Rev Microbiol. 2012 Sep;10(9):618–30.
105. Lewis LA, Polanski K, de Torres-Zabala M, Jayaraman S, Bowden L, Moore J, et al. Transcriptional Dynamics Driving MAMP-Triggered Immunity and Pathogen Effector-Mediated Immunosuppression in Arabidopsis Leaves Following Infection with Pseudomonas syringae pv tomato DC3000. Plant Cell. 2015 Nov;27(11):3038–64.
106. Nobori T, Velásquez AC, Wu J, Kvitko BH, Kremer JM, Wang Y, et al. Transcriptome landscape of a bacterial pathogen under plant immunity. Proc Natl Acad Sci U S A. 2018 Mar 27;115(13):E3055–64.
107. Zhang W, Corwin JA, Copeland D, Feusier J, Eshbaugh R, Chen F, et al. Plastic Transcriptomes Stabilize Immunity to Pathogen Diversity: The Jasmonic Acid and Salicylic Acid Networks within the Arabidopsis/Botrytis Pathosystem. Plant Cell. 2017 Nov;29(11):2727–52.
108. Zhang W, Corwin JA, Copeland DH, Feusier J, Eshbaugh R, Cook DE, et al. Plant-necrotroph co-transcriptome networks illuminate a metabolic battlefield. Elife. 2019 May 13;8.
109. Castro-Moretti FR, Gentzel IN, Mackey D, Alonso AP. Metabolomics as an Emerging Tool for the Study of Plant-Pathogen Interactions. Metabolites. 2020 Jan 29;10(2).
110. Sotelo-Silveira M, Chauvin AL, Marsch-Martínez N, Winkler R, de Folter S.

Metabolic fingerprinting of Arabidopsis thaliana accessions. Front Plant Sci. 2015 May 27;6:365.

111. Ward JL, Forcat S, Beckmann M, Bennett M, Miller SJ, Baker JM, et al. The metabolic transition during disease following infection of Arabidopsis thaliana by Pseudomonas syringae pv. tomato. Plant J. 2010 Aug;63(3):443–57.
112. Seybold H, Demetrowitsch TJ, Hassani MA, Szymczak S, Reim E, Haueisen J, et al. A fungal pathogen induces systemic susceptibility and systemic shifts in wheat metabolome and microbiome composition. Nat Commun. 2020 Apr 20;11(1):1910.
113. Parker D, Beckmann M, Zubair H, Enot DP, Caracuel-Rios Z, Overy DP, et al. Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by Magnaporthe grisea. Plant J. 2009 Sep;59(5):723–37.
114. Tugizimana F, Djami-Tchatchou AT, Steenkamp PA, Piater LA, Dubery IA. Metabolomic Analysis of Defense-Related Reprogramming in Sorghum bicolor in Response to Colletotrichum sublineolum Infection Reveals a Functional Metabolic Web of Phenylpropanoid and Flavonoid Pathways. Front Plant Sci. 2018;9:1840.
115. Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, et al. Comparative metabolomics and transcriptomics of plant response to Tomato yellow leaf curl virus infection in resistant and susceptible tomato cultivars. Metabolomics. 2015 Feb 1;11(1):81–97.
116. Nobori T, Wang Y, Wu J, Stolze SC, Tsuda Y, Finkemeier I, et al. Multidimensional gene regulatory landscape of a bacterial pathogen in plants. Nat Plants. 2020 Jul;6(7):883–96.
117. Mauch-Mani B, Slusarenko AJ. Arabidopsis as a model host for studying plant-pathogen interactions. Trends Microbiol. 1993 Oct;1(7):265–70.
118. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe. 2018 Jul 11;24(1):168–79.e4.
119. Lipps SM, Samac DA. Pseudomonas viridiflava: An internal outsider of the Pseudomonas syringae species complex. Mol Plant Pathol. 2022 Jan;23(1):3–15.
120. Karasov TL, Neumann M, Shirsekar G, Monroe G, PATHODOPSIS Team, Weigel D, et al. Drought selection on Arabidopsis populations and their microbiomes. bioRxiv. 2022. doi: 10.1101/2022.04.08.487684v1
121. Xin XF, Nomura K, Ding X, Chen X, Wang K, Aung K, et al. Pseudomonas syringae Effector Avirulence Protein E Localizes to the Host Plasma Membrane and Down-Regulates the Expression of the NONRACE-SPECIFIC DISEASE RESISTANCE1/HARPIN-INDUCED1-LIKE13 Gene Required for Antibacterial Immunity in Arabidopsis. Plant Physiol. 2015 Sep;169(1):793–802.
122. Jakob K, Kniskern JM, Bergelson J. The Role of Pectate Lyase and the Jasmonic Acid Defense Response in Pseudomonas viridiflava Virulence. Mol Plant Microbe Interact. 2007 Feb 1;20(2):146–58.
123. Araki H, Innan H, Kreitman M, Bergelson J. Molecular evolution of pathogenicity-island genes in Pseudomonas viridiflava. Genetics. 2007 Oct;177(2):1031–41.

124. Gonzalez ME, Marco F, Minguet EG, Carrasco-Sorli P, Blázquez MA, Carbonell J, et al. Perturbation of spermine synthase gene expression and transcript profiling provide new insights on the role of the tetraamine spermine in Arabidopsis defense against Pseudomonas viridiflava. Plant Physiol. 2011 Aug;156(4):2266–77.
125. Marina M, Sirera FV, Rambla JL, Gonzalez ME, Blázquez MA, Carbonell J, et al. Thermospermine catabolism increases Arabidopsis thaliana resistance to Pseudomonas viridiflava. J Exp Bot. 2013 Mar;64(5):1393–402.
126. Rossi FR, Marina M, Pieckenstain FL. Role of Arginine decarboxylase (ADC) in Arabidopsis thaliana defence against the pathogenic bacterium Pseudomonas viridiflava. Plant Biol . 2015 Jul;17(4):831–9.
127. Burdon JJ, Laine AL. Genetics of host plant resistance and pathogen infectivity and aggressiveness. In: Evolutionary Dynamics of Plant-Pathogen Interactions. Cambridge University Press; 2019.
128. Saati-Santamaría Z, Selem-Mojica N, Peral-Aranega E, Rivas R, García-Fraile P. Unveiling the genomic potential of Pseudomonas type strains for discovering new natural products. Microb Genom. 2022 Feb;8(2).
129. O'Brien J, Wright GD. An ecological perspective of microbial secondary metabolism. Curr Opin Biotechnol. 2011 Aug;22(4):552–8.
130. Helfrich EJN, Lin GM, Voigt CA, Clardy J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. Beilstein J Org Chem. 2019 Nov 29;15:2889–906.
131. Goss EM, Kreitman M, Bergelson J. Genetic diversity, recombination and cryptic clades in Pseudomonas viridiflava infecting natural populations of Arabidopsis thaliana. Genetics. 2005 Jan;169(1):21–35.
132. Hashimoto M, Neriya Y, Yamaji Y, Namba S. Recessive Resistance to Plant Viruses: Potential Resistance Genes Beyond Translation Initiation Factors. Front Microbiol. 2016 Oct 26;7:1695.
133. Iyer-Pascuzzi AS, McCouch SR. Recessive resistance genes and the Oryza sativa-Xanthomonas oryzae pv. oryzae pathosystem. Mol Plant Microbe Interact. 2007 Jul;20(7):731–9.
134. Vogel CM, Potthoff DB, Schäfer M, Barandun N, Vorholt JA. Protective role of the Arabidopsis leaf microbiota against a bacterial pathogen. Nat Microbiol. 2021 Dec;6(12):1537–48.
135. Huot B, Yao J, Montgomery BL, He SY. Growth-defense tradeoffs in plants: a balancing act to optimize fitness. Mol Plant. 2014 Aug;7(8):1267–87.
136. Pacheco-Moreno A, Stefanato FL, Ford JJ, Trippel C, Uszkoreit S, Ferrafiat L, et al. Pan-genome analysis identifies intersecting roles for Pseudomonas specialized metabolites in potato pathogen inhibition. Elife. 2021 Dec 31;10.

Acknowledgments

There are several people without whom this thesis would have not been possible, it would have taken longer and/or it would be less good (assuming it is good, that is). I want to thank them for their support and contributions during the past years.

To Detlef Weigel, for taking me as a PhD student even when my experience with plants was non-existent, for providing a project that encompassed everything I wanted to learn during my PhD, for being supportive of what I have wanted to do, for making the lab a good place to work, and for giving me a great daily supervisor. I appreciate having you as my boss.

To Talia Karasov, for teaching me how to work with plants and infections, for being patient and enthusiastic, for meeting with me regularly, even from a different continent, and for taking screenshots of my results, which made me feel good about them. I am forever grateful for your tireless support. I want to be as kind, dedicated and knowledgeable as you when I grow up.

To my TAC members, Georg Felix, Felicity Jones and Nadine Ziemert: thank you for your input during our meetings, for your enthusiasm about my project and for giving me the confidence I needed to start writing this dissertation. Thanks to Georg Felix, who served as my University supervisor and reviewer, and to Nadine Ziemert, whose expertise enabled the specialized metabolite chapter.

I got help from many in the lab. My project covered several methods, most of which were new to me, and this forced me to ask for help while giving me the opportunity to interact with my colleagues. Ilja Bezrukov helped me with the picture segmentation. Wei Yuan, Thanvi Srikant, Max Collenberg and Bridgit Waithaka-Vasilljevic helped me with RNA-seq. Cristina Barragan, Gautam Shirsekar, Ulrich Lutz and Ilja Bezrukov helped me with the genetic mapping that did not make it into this thesis. Manuela Neumann was my to-go person when I was finding my way around the lab and the methods. Gautam Shirsekar stops by my desk to say hola and give me suggestions and feedback. Moreover, thank you to all in the lab (and the offices) for being willing to answer my questions and share your knowledge, and specially for being nice.

One of the reasons I was able to complete this work was the people around me in the lab, who provided me with support, motivation, gossip, social gatherings and food: Adrián, Thanvi, Sergio, Cristina, Shanshan, Andrea, Miriam, Natalie, Martina. Leo and Andrea, even when in another lab. Very special mention to Rebecca Schwab for the constant supply of sweets, snacks and cakes, for answering my many questions about a variety of topics, for sharing her vision of science and life, for practicing mentoring and for making plans with me that allowed me to write this thesis, even when I thought I was not ready or there was not enough time. The lab would not be the same without you. I thank all those who knew I was writing and did not talk to me or did not let me

take care of the BBQ; although a bit weird, I appreciated it as it made me feel supported.

I am thankful for the friends I have kept through the years, who listen to me and give me support and comfort, to whom I can say I am struggling without fear, and who share their struggles with me so that I feel less alone: Angélica, Poust, Elisa, Wendy, Trang, Marina, Daniel. I hope I can give other as much love and comfort as you have to me.

To my family, who for the past four years have asked how my plants and infections were going, who feed me delicious food I miss when I visit, who have not doubted that I could do this and who give me so much love even from afar. I miss you constantly and would love to have you around.

To COVID, for forcing virtual defenses and giving me the opportunity to share mine with family and friends abroad.

There is one more person without whom this thesis would *literally* not exist: Jacobo. The reason why I am in Tübingen, I did a PhD in plants and I wake up every day with love. Your support through every step that led here is invaluable, and even more so is knowing that you will support whatever comes after. I am in awe of who you are, who you have become, the love you show and the amount of bad jokes you can produce. Thank you for teaching me the command line and statistics, for sharing your R knowledge, for reading every report, abstract and section of this thesis, for making breakfast every morning and dinner whenever I was working late. Lo valoro!

Support and help are a constant in these acknowledgments. I hope I can provide these to those around me, just as you all have done to me.