THE UNIVERSITY OF CHICAGO

# ECOLOGICAL ADAPTATION IN THE UBIQUITOUS PLANT-PATHOGEN PSEUDOMONAS SYRINGAE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY
REBECCA SUE SATTERWHITE

CHICAGO, ILLINOIS
MARCH 2024

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

Thank you to Hannah Whitehurst for expert technical assistance with NGS and for being a true friend! Thank you to Hanna Maerkle for sharing her genius and undying patience. Thank you to Sara, Amir, and Mina for much-needed moral support. Thank you to my wonderful friends from my early years at UChicago who helped me stay inspired. Thank you to my family for unending support, especially my aunts Nancy and Marcy and my cousins Skye and Bethany.

# CHAPTER 1

# INTRODUCTION

Microbes are variable organisms in a variable natural world. Microbial taxa harbor some of the most interesting ecological adaptations, from dispensable chromosomes in filamentous fungi (Galazka and Freitag 2014) to a bacterial toxin that mimics a fungal exudate in order to highjack the plant immune response (Weiler et al. 2014). The extensive phenotypic and genetic variation characteristic of microbes has made them difficult to classify into species (Cohan 2001; Cohan 2002). In fact, traditional attempts at microbial pathogen classification have tended to use host phenotype rather than characteristics of the pathogen itself (Lelliott et al. 1966).

The question of how to appropriately define microbial species has inspired a wealth of literature but no clear solution (Lan and Reeves 2000; Cohan 2001; Fraser et al. 2007; Shapiro et al. 2016). Ecotype theory is one useful measure for understanding ecological adaptation of evolving lineages (Cohan 2002). An ecotype consists of the populations occupying a particular niche, wherein genetic divergence is limited by natural selection which repeatedly purges variation (Cohan 2002). Predictions for a stable ecotype include 1) that the group should form a monophyletic clade, 2) that the clade should resemble a star, such that each member of the ecotype is approximately equally related to each other, and 3) that relatedness within the ecotype is higher than between the ecotype and other groups (Cohan 2002; Ward et al. 2008). This concept has been invoked to describe many microbial species, including communities of extremophiles living in natural thermal geysers (Keller et al. 2023) as well as the opportunistic plant pathogen *P. syringae* (Baltrus et al. 2016; Dillon et al. 2019B).

The relationship between pathogens and their plant hosts is often likened to a molecular 'arms race', where microbes have arsenals of virulence factors that hosts experience counter-selection to recognize (Bergelson et al. 2001; Stahl and Bishop 2000; Dawkins and Krebs 1979). Virulence factors can alter a host interaction indirectly by exuding toxins or other

secondary metabolites (Bender et al. 1999), or directly by injecting effector proteins into the cytoplasm through the Type III Secretion System (Collmer et al. 2000). Effectors are a well-studied class of virulence factors that have been considered double-edged swords (Russell et al. 2015) because they represent both the possibility of subverting a host immune response and of being recognized by the host and triggering an immune response (Block and Alfano 2011). Effectors are of particular interest to any study of ecological adaptation because a strain's unique effector load is thought to be a key determinate of its host range (Dillon et al. 2019).

Another class of gene that has received a lot of attention are singletons, or strain-specific genes. Once thought to be sequencing artifacts, it now seems that many singleton genes are of ecological relevance (Wilson et al. 2005; Fakhar et al. 2023). Theory does not predict the maintenance of large numbers of neutral alleles (Wolf et al. 2016) at population sizes as large as those typical of bacteria (Bobay and Ochman 2018), thus the observation of rare genetic variants in high numbers has led to the speculation that such variants are maintained by selection in various environments. Plants, for example, likely represent ecologically distinct environments for their pathogens; that is, the single example we have of the same Tn-library being grown in multiple hosts found that distinct genes were required for growth on different hosts (Hellman et al. 2020). However, outside of this example, the extent to which there exists a host-specific accessory genome remains to be determined.

## Study system

*P. syringae* is ubiquitous on our planet, with a global distribution and the potential to impose significant yearly costs related to patterns of epidemic growth on crop cultivars (Vanneste 2016; Cameron and Sarojini 2013). As a species, this Gammaproteobacteria emerged long before land plants, approximately 1.75 Gya, and likely harbors significant adaptations to aquatic and icy environments (Morris et al. 2013). While most *P. syringae* isolates are

from diseased crop plants, an increasing number of genomes from non-crop sources have been sequenced (Monteil et al. 2013). Traditional classification of the species was predicated on plant pathogenicity phenotypes, including toxin production and plant response. Prior to 1970, strains were grouped using the LOPAT tests: L, levan production; O, oxidase production; P, pectinolitic activity; A, arginine dihydrolase production; and T, tobacco hypersensibility (Lelliott et al. 1966). When the LOPAT group 1 strains were collapsed into the *P. syringae* species, the problem of what to do with the several remaining strains that defied LOPAT classification led to the proposal that they be classified as pathovars of their particular host-of-isolation (Young et al. 1978). Today, *P. syringae* is classified into more than 60 pathovars of various crop cultivars (Dillon et al. 2019B). However, pathovars can grow in hosts that span multiple species (Morris et al. 2019), and attempts to incorporate genomic information have often found an incongruence between host association and phylogeny (Sarkar and Guttman 2004; Baltrus et al. 2011; Baltrus et al. 2016; Dylan et al. 2019A). Thus, classification of this species is still debated (Baltrus 2016; Gomila et al. 2017).

*P. syringae* is genetically variable, and contains a large number of genes important for ecological interactions with other organisms (Dillon et al. 2019A). For example, *P. syringae* has a molecular superpower; it produces a protein that nucleates ice at temperatures slightly below freezing (Lindow 1983). The resultant ice damages plant cells, giving *P. syringae* a chance to infect. In addition to ice nucleation, each strain contains an arsenal of virulence factors that contribute to the infection success of this opportunistic pathogen. For example, fluorescent Pseudomonads secrete coronatine, which is a jasmonic acid mimic that induces stomatal opening and disrupts the salicylic acid immune response (Melotto et al. 2006). Another common approach is for the pathogen to target host signaling during infection; for example, the secreted effector HopBF1 is a protein kinase that targets the plant Hsp90 chaperone, disabling it from activating immune receptors (Lopez et al. 2019). This variety

of virulence strategies is not surprising considering that the pathogen is generally thought to be winning the evolutionary arms race; pathogens have shorter generation times and larger population sizes than macroorganisms (Kaltz and Shykoff 1998). However, one important example found maladaptation in strains of *P. syringae* isolated from the natural host Arabidopsis thaliana, in that isolates tended to be recognized by the plant and induce a host immune response (Kniskern et al. 2011). These results suggest that *A. thaliana* is winning the arms race against *P. syringae*; thus, it is likely that whether the pathogen wins the arms race is specific to the organisms being compared.

With this dissertation, I aimed to learn about ecological adaptation in *P. syringae* using three approaches. The first approach was entirely computational; I leveraged the existing genomic data for three sets of *P. syringae* isolates from either *A. thaliana*, crop, or non-host environmental sources to answer questions about their relatedness and the overlap of traits. I found that the core genomes of the three groups represent distinctly evolving lineages, indicating the suitability of this selection of ecological groups. Consistent with the findings of others, I found that host is a poor predictor of pathogen relatedness, indicating that the genes related to host-specificity are likely found in the variable (not core) genome. I also found that the *A. thaliana* group is likely to represent a stable ecotype. Though I found that the environmental and crop isolates are more closely related to each other than either are to the *A. thaliana* strains, I did detect some degree of specialization within the environmental strains as well. The second approach I took focused on a subset of five *P. syringae* host-pathogen pairs, with isolates from crops and *A. thaliana*. For these strains, I compared effector repertoires and assayed growth in three *in vitro* and five *in planta* environments. I used effector load to make predictions about the performance of strains *in vitro* and *in planta*; I found that pathogens are demonstrably able to overcome any predicted costs associated with high effector load in these experiments. The *in vitro* experiments revealed a strain-wide preference for one media, KB, and an environment-wide advantage for one strain, NP29.

However, the best grower for the *in planta* experiments was a different strain, DC3000. I also found patterns consistent with local adaptation in two of five cases, in that growth on the host-of-isolation was significantly higher than growth in all other hosts. In a third approach I attempted to use TnSeq to detect a host-specific accessory genome, if one exists, but was unable to generate the sequences required to statically test hypotheses. I summarize attempts and lessons learned herein.

Together, these studies reflect the dynamic and somewhat unpredictable evolutionary landscape of *P. syringae*. I found that effector repertoire is not always a good predictor of growth, and that host association is not a good predictor of pathogen phylogeny. It seems a good deal is learned from considering the relationship of *P. syringae* with *A. thaliana*; for example, the strain NP29 exhibits growth patterns consistent with local adaptation to *A. thaliana*, but also elicits host recognition immune response (Kniskern et al. 2011). Considering that this strain, like many isolated from *A. thaliana*, carries a minimal effector repertoire, it seems quite likely that the plant has the advantage in the arms race. However, outside of *A. thaliana*, host is only minimally (if at all) associated with pathogen phylogeny, a trend which appears to be universal in core genome trees of this species. The finding of three out of five infections not exhibiting patterns of local adaptation reflects that patterns of local adaptation should not be assumed, as is implied by the use of a pathovar system. For this reason, I believe it makes sense to use a species delineator that is fully sequence based (Baltrus 2016). Ultimately it will take a more detailed understanding of the evolutionary history and interactions between hosts and their pathogens to be able to predict the outcome of infection.

# CHAPTER 2

# COMPARATIVE GENETICS OF PSEUDOMONAS SYRINGAE FROM CROPS, THE ENVIRONMENT, AND THE NATURAL HOST ARABIDOPSIS THALIANA

## 2.1 Abstract

The plant-pathogen *P. syringae* has a wide host range that includes a large number of plant hosts (Berge et al. 2014). It has been found on multiple occasions that host association is not a good predictor of pathogen phylogeny (Baltrus et al. 2011; Baltrus et al. 2016; Dillon et al. 2019B); what, then, makes for an appropriate ecological grouping? I answer this question using a balanced but non-random subset of strains from each of three sources: crop hosts, the natural host *A. thaliana*, and non-host environmental reservoirs. I find evidence that these three groups are appropriate to compare, in that they show evidence of evolving distinctly according to the core genome phylogeny. As predicted, I never find that host association is a good predictor of pathogen phylogeny, which indicates that host-specific adaptations are more likely to be present in the variable than core genomes. Another central finding based on several points of evidence is that the *A. thaliana*-derived isolates are likely to represent a stable ecotype. Finally, I describe the overlapping vs. group-specific functions, effectors, virulence factors, and antibiotic resistance genes. I find that effectors delineate by group: there were few effectors in the *A. thaliana*-derived group, but a large number shared by the crop and environmental groups. This was not the case for general virulence factors and antibiotic resistance genes, which tended to be shared by all groups. These results inform our understanding of ecological adaptation and gene content evolution in *P. syringae*.

## 2.2 Introduction

*P. syringae* is an opportunist: a highly adaptable (Spiers et al. 2000) phytopathogen that maintains a large accessory genome which is still increasing as new isolates are sequenced (Dillon et al. 2019B). The species has a wide host range that includes many of the crop plants that we eat, as well as natural, weedy hosts like the model plant *A. thaliana*. Attempts to classify stains based on host-association have generally found that one is not a good predictor of the other. In one example, Sarkar and Guttman (2004) found less than 20% of genetic variation in the core phylogeny was explained by host association. More recently, Dillon et al. published the largest pangenome analysis to date of *P. syringae* (2019A), which demonstrated little congruence of host association with pathogen phylogeny using the largest sample possible. Thus, it may be that the host is only one of a large number of ecological forces driving pathogen evolution, rather than a central driver.

One attempt to categorize genetically similar microbial species with shared ecology is ecotype theory (Cohan 2002; Ward et al. 2008). An aim of ecotype theory is to allow for systematic avoidance of grouping ecologically distinct populations into the same species (Cohan 2006). An ecotype consists of the genotypes adapted to a particular niche (Cohan 2002). Ecotypes should delineate based on true ecological differences, thus the source of strains informs what comprises an appropriate grouping. For example, *P. syringae* has a long evolutionary history of survival in a plant-free world (Morris et al. 2013) but at times acts as an opportunistic phytopathogen. It has been found that some environmental isolates are closely related, and in fact in the same phylogroup as those isolated from crop hosts (Monteil et al. 2013). Thus, it appears unlikely that strains from environmental sources are ecologically distinct from plant pathogens. I do expect some differences; as they must survive in a stochastic natural environment, I predict that environmental strains should be specifically enriched for functions related to stress from harsh conditions. Additionally, environmental strains have not necessarily had to contend with a host immune response, which

includes two phases of defensive action that exerts a strong selective pressure *in planta* (Jones and Dangl 2006). Survival in a host environment is thus likely to require virulence factors, defined as any molecular product that is useful in establishing or promoting a host infection or microbe-microbe interaction (Barrett et al. 2009). These include structural proteins located in the cell membrane that are responsible for host recognition and secreted secondary metabolites, like toxins. For example, the toxin Coronatine is a mimic of a plant stress hormone, methyl-jasmonate, and has been shown to modulate virulence in some isolates of *P. syringae* (Bender et al. 1999). Since virulence factors can also drive microbe-microbe interactions, either directly (reviewed in Trunk et al. 2019; Snelders et al.2020) or indirectly by altering the local environment (Barrett et al. 2009), they are likely not limited to host-associated strains.

Secreted effectors in particular are a well-studied type of virulence factor, the compliment of which can determine host range of a particular pathogen (Dillon et al. 2019A). A recent study gathered sequences from all known Type III *P. syringae* effectors (Dillon et al. 2019A), identifying 14,613 sequences spanning 70 gene families and 523 alleles (the PsyTec Compendium), and asked whether each allele was recognized by *A. thaliana* ecotype Col-0, a standard genotype that is considered representative of the species. They found that a majority of effector alleles (more than 70%) were recognized by *A. thaliana*, suggesting the presence of a selective force within the *A. thaliana* environment on *P. syringae* for the loss of effectors. In fact, it is known that *A. thaliana*-derived strains of *P. syringae* contain few effectors relative to those derived from crop hosts (Karasov et al. 2017). Still, a few secreted effectors are conserved species-wide as part of the core genome of *P. syringae* (Dillon et al. 2019B), indicating the historical importance of effectors for survival of this species.

Herein I explore a balanced, non-random subset of *P. syringae* strains isolated from (i) environmental sources, (ii) the natural host *A. thaliana*, and (iii) crop hosts, all from the same geographic region, and explore the characteristics of the core- and pan-genomes, ev-

idence of selection in select virulence factors, and the functions that are enriched amongst isolates from different environments. Given previous findings of 1) close relatedness between crop and environmentally-derived isolates (Monteil et al. 2013), and 2) lower singleton load for *A. thaliana*-derived than crop-derived isolates (Karasov et a. 2017), I expect the crop and environmental groups to be more similar to each other than either group is to the *A. thaliana*-derived isolates, and that the crop and *A. thaliana*-derived isolates represent distinct ecological groupings, such that there should be marked differences in gene content between the two plant-associated groups. Since virulence factors can be useful for microbe-microbe interactions, which are likely to be important for all *P. syringae* isolates, I do not predict the same differences in presence of virulence factors. Similarly, I expect antibiotic resistance genes should be useful for all life histories and do not expect a signature of ecological differentiation in antibiotic resistance genes.

## 2.3 Methods

### *2.3.1 Genome selection and quality control*

The vast majority of strains with published genomes from environmental sources were isolated in France, so I chose to limit this analysis to strains isolated from France to avoid biases introduced by disparate localities. In order to cover a breadth of strain sources, I chose three groups of isolates: a crop set comprised of 24 isolates from various crop cultivars, an Arabidopsis-derived set (Bergelson and Roux; unpublished data) of 25 isolates from the natural, weedy host, and an environmental set (Monteil et al. 2013) of 25 isolates derived from non-host sources (Table 2.1). These are herein referred to as 'crop', 'At', and 'env', respectively. It is important to note that these groups do not represent true random samples; the crop and env groups were chosen based on covering a range of sources while spanning multiple phylogroups, whereas the At group was isolated from a single host species and only

represents phylogroup seven. To my knowledge this is the first pangenome analysis based on comparing *P. syringae* genomes from these groups.

Genomes used in this study were initially in various stages of completion; the crop and At-derived genomes were generally already high-quality assemblies, but many of the environmental genomes were available only as Sequencing Read Archive files that needed to be assembled. I downloaded the SRAs from NCBI; if there were multiple sequencing runs available, I concatenated those reads prior to assembly. I checked initial quality of the reads using multiqc (Ewels et al. 2016); all samples had adapter contamination, so I used bbduk (Bushnell 2014) to remove adapters and confirmed success of the trimming with multiqc. I performed de novo assembly with SPAdes (Bankevich et al. 2012), with standard settings and the flag "–phredoffset -33". I performed final quality control with QUAST (Gurevich et al. 2013) to confirm that, for strains from At and crops, the quality of these assemblies was comparable to the previously assembled genomes. All genomes had greater than 98% completion and less than 5% redundancy, and were reannotated as part of my pipeline for consistency. Genome sizes ranged from 5.7 Mbp to 6.4 Mbp, with the number of genes ranging from 5053 to 6230, consistent with *P. syringae* genomes used in published studies (Monteil et al. 2016; Karasov et al. 2017).

### *2.3.2 Pangenome analysis with anvi'o*

To process FASTA files for each of the 54 assemblies and perform a pangenome analysis, I used anvi'o v.7.1. (Eren et al. 2015). For all analyses, I filtered out contigs with less than 1000 bp. My pipeline was, in brief: (1) simplify FASTA definition lines, (2) generate anvi'o contig databases for each genome, (3) identify open reading frames using Prodigal (Hyatt et al. 2010) v2.60X, (4) annotate genes with functions using the NCBI's Clusters of Orthologous Groups (COG) (Tatusov et al. 1997), (5) identify single copy core genes using HMMER v3.2.1X (Eddy 2008) and an included collection of HMM profiles spanning

Table 2.1: Strains used in this study

| Source | Group | Strain | Length | GC | Genes | Gene clusters | Singletons |
|---|---|---|---|---|---|---|---|
| At | At | BACCB1 | 5956328 | 0.5934181 | 5256 | 5066 | 33 |
| At | At | BELLA1 | 6131582 | 0.5917228 | 5440 | 5201 | 0 |
| At | At | BULAA1 | 6095915 | 0.5933493 | 5412 | 5184 | 71 |
| At | At | CARLA2 | 5958805 | 0.5934361 | 5280 | 5074 | 57 |
| At | At | CASTIA1 | 6109650 | 0.5918044 | 5443 | 5214 | 55 |
| At | At | CERNB2 | 5933176 | 0.5936613 | 5280 | 5069 | 36 |
| At | At | JACOC2 | 6085767 | 0.5931229 | 5410 | 5198 | 122 |
| At | At | LABAA1 | 5963389 | 0.5933328 | 5297 | 5069 | 39 |
| At | At | LABAC2 | 6205186 | 0.5911724 | 5527 | 5297 | 108 |
| At | At | LABASB1 | 5983879 | 0.5923683 | 5291 | 5087 | 80 |
| At | At | LACRC1 | 6224453 | 0.5921640 | 5502 | 5207 | 50 |
| At | At | LANTC1 | 5979876 | 0.5932215 | 5315 | 5106 | 67 |
| At | At | LANTC2 | 6052562 | 0.5925558 | 5387 | 5174 | 59 |
| At | At | LAUZA1 | 6054542 | 0.5926389 | 5407 | 5182 | 76 |
| At | At | LUZEE6 | 6135905 | 0.5925885 | 5433 | 5218 | 86 |
| At | At | MAZAA1 | 5917839 | 0.5926844 | 5275 | 5052 | 46 |
| At | At | MERVA2 | 6106471 | 0.5901238 | 5453 | 5231 | 110 |
| At | At | PREIA1 | 6054887 | 0.5921972 | 5363 | 5170 | 53 |
| At | At | RAYRB2 | 6013490 | 0.5931921 | 5307 | 5096 | 43 |
| At | At | RAYRB3 | 6157987 | 0.5921425 | 5541 | 5267 | 95 |
| At | At | REALA1 | 6130861 | 0.5917210 | 5442 | 5202 | 1 |
| At | At | SAUBA1 | 6199415 | 0.5908778 | 5506 | 5238 | 59 |
| At | At | SAUBA2 | 6042794 | 0.5925673 | 5394 | 5204 | 69 |
| At | At | SIMOA1 | 5975575 | 0.5925974 | 5287 | 5088 | 60 |
| At | At | VILLEMA1 | 6008606 | 0.5927132 | 5316 | 5095 | 55 |
| apricot | crop | 41a | 5983849 | 0.5911207 | 5140 | 4965 | 79 |
| beet | crop | ICMP11935 | 5984734 | 0.5874234 | 6230 | 5311 | 249 |
| cantaloupe | crop | CC440 | 5933410 | 0.5903929 | 5086 | 4911 | 104 |
| cantaloupe | crop | CC457 | 5979293 | 0.5899748 | 5125 | 4912 | 11 |
| cantaloupe | crop | CC94 | 6149858 | 0.5925856 | 5696 | 5136 | 121 |
| celery | crop | ICMP11947 | 5826748 | 0.5936028 | 5075 | 4865 | 87 |
| kidney bean | crop | CFBP13578 | 6147017 | 0.5917576 | 5510 | 5270 | 124 |
| kidney bean | crop | CFBP8160 | 6099296 | 0.5875884 | 5436 | 5108 | 106 |
| kidney bean | crop | CFBP8779 | 6140205 | 0.6210577 | 5566 | 5314 | 167 |
| kidney bean | crop | CFPB8161 | 6245039 | 0.5872357 | 5552 | 5227 | 21 |
| kiwifruit | crop | CFBP3846 | 6229862 | 0.5864898 | 6027 | 5220 | 136 |
| kiwifruit | crop | CFBP8180 | 6261671 | 0.5868877 | 5592 | 5252 | 40 |
| kiwifruit | crop | ICMP13102 | 5954747 | 0.5932086 | 5239 | 4972 | 68 |
| leek | crop | ICMP8960 | 6065267 | 0.5762752 | 5719 | 5275 | 95 |
| leek | crop | ICMP8961 | 5827909 | 0.5776667 | 5522 | 5083 | 47 |
| lemon | crop | ICMP4917 | 5928691 | 0.5918071 | 5366 | 5067 | 126 |
| oleander | crop | ICMP16944 | 5801705 | 0.5814913 | 5517 | 5109 | 87 |
| polyanthus | crop | ICMP18417 | 6358112 | 0.5860426 | 5801 | 5337 | 214 |
| pumpkin | crop | P12832 | 5909338 | 0.5906652 | 5094 | 4918 | 18 |
| pumpkin | crop | P77 | 6133764 | 0.5914094 | 5345 | 5122 | 19 |
| pumpkin | crop | P89 | 6222478 | 0.5908248 | 5474 | 5238 | 57 |
| radish | crop | CFBP13571 | 6026070 | 0.6223839 | 5421 | 5186 | 96 |
| radish | crop | CFBP13572 | 6020391 | 0.5915216 | 5374 | 5162 | 63 |
| soybean | crop | ICMP5027 | 5926573 | 0.5792360 | 5873 | 5324 | 372 |
| epilithic biofilm | env | **SZB0006** | 5676844 | 0.5887377 | 5252 | 4911 | 59 |
| epilithic biofilm | env | **SZB0065** | 5741751 | 0.5881180 | 5444 | 4986 | 56 |
| irrigation water | env | **CMW0021** | 5772452 | 0.5880607 | 5540 | 5039 | 55 |
| irrigation water | env | GAW0231 | 5840762 | 0.5878986 | 5377 | 5091 | 64 |
| lake water | env | CC1543 | 5716914 | 0.5920264 | 5205 | 4925 | 87 |
| lake water | env | CC1544 | 5858806 | 0.5908691 | 5281 | 5052 | 122 |
| leaf litter | env | **CCV0502** | 5921468 | 0.5864232 | 5398 | 5129 | 73 |
| rain | env | **CST0002** | 5900870 | 0.5853555 | 5605 | 5152 | 47 |
| rain | env | **CST0009** | 5853848 | 0.5873849 | 5342 | 5099 | 68 |
| rain | env | **CST0076** | 5807499 | 0.5865578 | 5310 | 5037 | 39 |
| rain | env | **CST0094** | 5862602 | 0.5860812 | 5345 | 5105 | 36 |
| rain | env | LYR0002 | 6051922 | 0.5805665 | 5767 | 5232 | 122 |
| river water | env | **CMW0036** | 5700152 | 0.5894355 | 5137 | 4918 | 60 |
| river water | env | UB246 | 6088488 | 0.5711407 | 5585 | 5387 | 193 |
| river water | env | UB303 | 6141482 | 0.5918656 | 5318 | 5043 | 78 |
| snow | env | CC1557 | 5758024 | 0.5855686 | 5053 | 4792 | 218 |
| snow | env | CC1559 | 5842720 | 0.5886823 | 5289 | 5017 | 115 |
| snowpack | env | **CCV0450** | 5885620 | 0.5876735 | 5319 | 5095 | 97 |
| snowpack | env | **CCV0611** | 5733670 | 0.5885114 | 5163 | 4949 | 52 |
| snowpack | env | CSZ0137 | 6298766 | 0.5700948 | 5617 | 5416 | 157 |
| snowpack | env | **CSZ0279** | 5754966 | 0.5885847 | 5217 | 4986 | 46 |
| snowpack | env | CSZ0324 | 6389627 | 0.5704835 | 5722 | 5505 | 159 |
| snowpack | env | **CSZ0720** | 5746602 | 0.5891179 | 5197 | 4972 | 34 |
| stream water | env | CCE0118 | 6223941 | 0.5705909 | 5606 | 5386 | 217 |
| stream water | env | **CSZ0259** | 5767871 | 0.5886453 | 5517 | 5024 | 92 |

bacteria and archaea, (6) build an anvi'o genomes storage database, and finally (7) build the pangenome. Additional anvi'o scripts were used for particular analyses: I estimated completion of the genomes using the script anvi-estimate-genome-completeness. I obtained statistics on contig quality using the script anvi-display-contigs-stats. I generated a summary file of the pangenome in order to extract gene frequencies using the script anvi-summarize. I employed fastANI (Jain et al. 2018) to calculate Average Nucleotide Identity (ANI) for each group of genomes using the script anvi-compute-genome-similarity. ANI is a calculation of the average nucleotide identity of all orthologous genes common to any two genomes (Jain et al. 2018). I estimated functional enrichment for each group based on COG annotations using the script anvi-compute-functional-enrichment-in-pan. This script works by finding the COG functional annotations that appear most frequently in a gene cluster, then fitting a logistic regression to the occurrence of each function using group as the explanatory variable (Shaiber et al. 2020). The equality of proportions across group affiliation is tested using a Rao score test, which provides an enrichment score (test statistic) and p-value, which is then converted to a q-value to account for multiple testing (Shaiber et al. 2020). Finally, I extracted sequences for accessory, single copy core, and singleton genomes using the script anvi-get-sequences-for-gene-clusters. I chose to use a hard cut-off when defining core (found in all) vs. accessory (found in some) genomes in order to make the comparisons as straightforward as possible, since core genes are more likely to represent vertically inherited genes than the accessory genome (Lan and Reeves 2000).

### *2.3.3 ANI of* A. thaliana*-derived isolates from France and the Midwest USA*

Midwest strains used for this comparison were sequenced by Karasov et al. (2017). These genomes are *P. syringae* isolated from *A. thaliana* growing in agricultural fields in Indiana and Michigan between 2000 and 2014 and are publicly available. I prepared the genome FASTA files using the anvi'o pipeline described above, and used fastANI to calculate ANI

with the script anvi-compute-genome-similarity. I parsed the results with custom R scripts.

### *2.3.4 Additional genomic analyses*

Computational analyses were done in R v.4.2.2 (R Core Team 2022). Identification of virulence factors and antibiotic resistance genes were done using ABRicate version 0.8.13 (Seemann 2023) with databases VFDB (Chen et al. 2016) and MEGAres (Bonnin et al. 2023) for virulence factor and antibiotic resistance gene detection, respectively. Identification of Type III secreted effectors required rebuilding the PsyTec database (LaFlamme et al. 2020), which is published as a supplementary file to the original paper. With Blast+ for the command line (Camacho et al. 2009), I used the commands makeblastdb to rebuild the published sequence database and blastn to blast my genomes against it. I parsed the output with custom R scripts, and limited my analyses to matches with at least a 99% identity over 90% of the query sequence. To generate phylogenies, I first used trimAl (Capella-Gutierrez et al. 2009) to remove nucleotide positions that were gap characters in more than half of the sequences. I then used IQ-TREE (Nguyen et al. 2015) with the general matrix model 'WAG' to infer a maximum likelihood tree with 1,000 bootstraps. Trees were annotated and formatted in iTOL (Letunic et al. 2021). For selection analyses, DNA sequences were extracted from the pangenome using the anvi'o script anvi-summarize and then aligned with muscle v. 3.8.1551 (Edgar 2004). Alignments were deduplicated using the bbmap script dedupe.sh (Bushnell 2014). GC content was calculated using SeqKit (Shen et al. 2016), and Tajima's D and was calculated using the R package pegas (Paradis 2010).

## 2.4 Results

### *2.4.1 Relatedness of the experimental system*

In order to explore the evolutionary relationships in my experimental system, I generated a phylogeny using the 1,471 single copy core genes found in all isolates (Fig. 2.1). I found that there was general agreement between my tree and previous representations of the *P. syringae* phylogeny (e.g., Dillon et al. 2019B); canonical phylogroups 1-4, 7, 10, and 13 are represented (Fig. 2.1). One cluster of two crop isolates (CFBP8779 and CFBP13571) form a monophyletic clade that likely corresponds to an early-branching, secondary phylogroup (8, 9, or 11), but I was unable to identify which as the representative genomes are currently unavailable on NCBI. Another cluster of four env isolates (CCE0118, CSZ0137, XCS0324, and UB246) represent phylogroup 13, and fifteen env isolates clustered together with crop isolate CFBP3846 to form a monophyletic group representing phylogroup 1A. The remaining genomes from the env group are mixed with the crop group and distributed between phylogroups 2 and 3c.

There were two central results from Fig. 2.1. One is that the At group was the most completely delineated, with all isolates clustered together in a clade representing phylogroup 7 that also contains two crop isolates (CFBP13572 and CFBP13578). In fact, the At group appears as a 'star' clade with a single ancestral node, such that members of this group are approximately equally related to each other (Cohan 2002). These results agree with my predictions that the At group would have the greatest within-group similarity, whereas the env and crop clades would share a greater degree of similarity to each other. But it is important to note that even the most congruous cluster contained two isolates from different crops (Fig. 2.1). Phylogroup 1A is the second most congruous grouping: 56% (14/25) of the env isolates cluster together while mostly excluding crop isolates. This indicates that while some crop and env isolates are close relatives, consistent with predictions based on previous

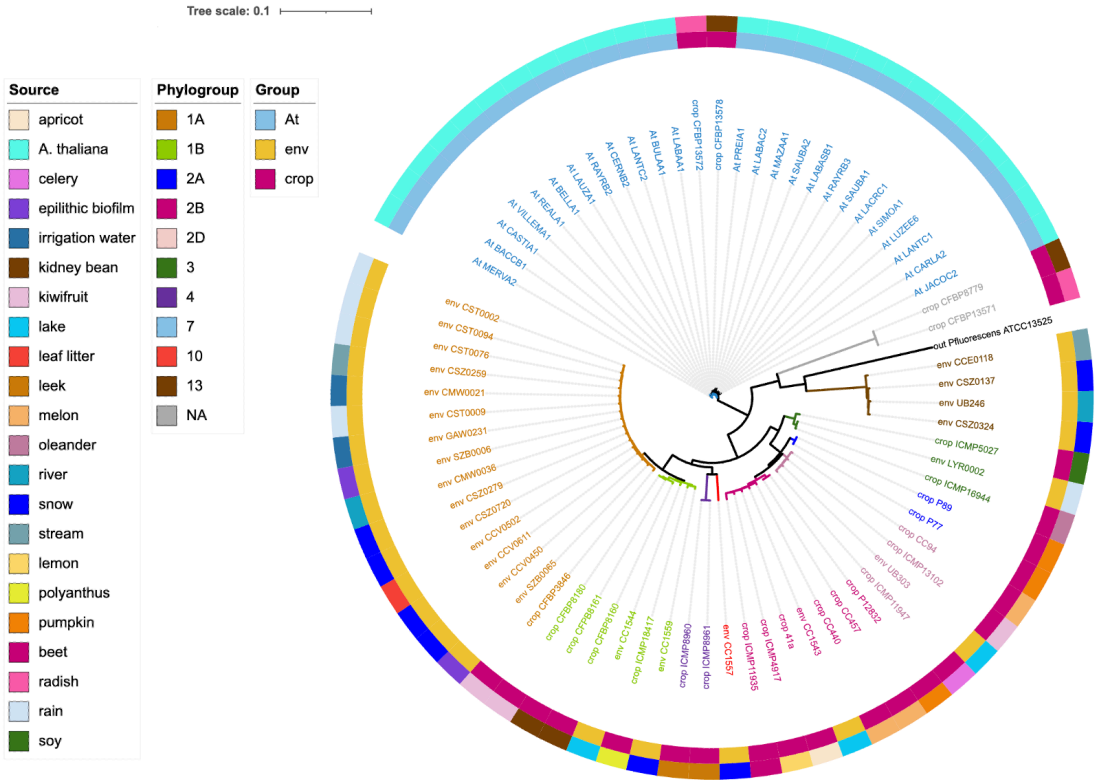



Figure 2.1: Single copy gene tree. Maximum likelihood phylogeny built with 10,000 bootstraps. Colors of the nodes and strain names correspond to phylogroup. Rings show group and source as indicated in the legend.

findings (Monteil et al. 2013), there is another group of environmental isolates more closely related to each other than to crop isolates. Thus, though the env and crop groups overlap in some cases, there is still some signal of within-group separation for the env isolates, which is not something I had predicted.

The other central result from Fig. 2.1. is that, though it is tempting to speculate that strain relatedness should resolve in a pattern consistent with host association, this is generally not what I see. Instead, each phylogroup tends to contain strains from more than one source (Fig. 2.1). For example, isolates from rain are found in phylogroups 1A and 3, and isolates from pumpkin are found in phylogroups 2A and 2B. One exception is phylogroup 4 which is represented only by two isolates from leek (Fig. 2.1). This finding generally agrees with the results of others (Dillon et al. 2019B; Baltrus et al. 2011), and suggests that genes that are responsible for host-specific adaptations are not in the core genome, but instead are members of the variable genome.

### *2.4.2 Comparisons of ANI*

To estimate sequence similarity directly, I compared ANI within and between groups (Table 2.2). Consistent with all At isolates being from a single clade (Fig. 2.1), the At group had the highest within-group ANI (0.97). In contrast, the crop and env groups both had within-group ANI close to 0.90 (Table 2.2). This is relatively low for strains of the same species, as exemplified by a 2004 work with *Escherichia coli* that found ANI $> 94\%$ was a species-level cut-off determined by both DNA-DNA reassociation and by their own empirical comparisons of 9 genomes (Konstantinidis and Tiedje 2004). The At group ANI was so much higher than either the env or crop groups' that I reasoned it may be more similar to other, allopatric At-derived groups. Therefore, I re-calculated ANI of the original 25 French *A. thaliana*-derived isolates when combined with another group of 16 *A. thaliana*-derived isolates from the Midwest United States (Karasov et al. 2017). I first checked the within group ANI of

Table 2.2: ANI within and between groups

| Comparison | ANI |
|---|---|
| At | 0.973 |
| crop | 0.885 |
| env | 0.911 |
| At & crop | 0.849 |
| Crop & env | 0.872 |
| At & env | 0.837 |
| Michigan At | 0.955 |
| At & Michigan At | 0.966 |

the Midwest isolates, which was 0.955, similar to the French At group. I next found that the two groups together had an average ANI of 0.966, such that the French *A. thaliana*-derived group had a more similar nucleotide complement to isolates from an allopatric *A. thaliana* population than to either group of crop- or environmentally-derived strains used in this study.

### *2.4.3 Characteristics of the core and variable genomes*

The pangenome of 74 genomes consisted of 400,532 genes divided into 20,454 gene clusters. Of these, 2,743 (13.4%) gene clusters were core, shared by all isolates, leaving 17,711 (86.6%) variable gene clusters and 6,383 (31.2%) singletons, found in only a single isolate. I found that, on average, nearly 50% of a strain's gene clusters were variable (At: 2417/ 5160 = 46.8%; env: 2347/5090 = 46.1%; and crop: 2394/5137 = 46.6%). This is similar to a previous estimate obtained by 19 pathotype genomes from various crop hosts (43%: Baltrus et al. 2011). The similarity of this percentage across groups reveals that the number of variable genes was consistent despite differences in habitat.

However, there were significant differences in the numbers of singletons per strain from a given group. Though all groups displayed an enrichment of rare and singleton genes (Fig. 2A), the crop and env groups had significantly more singletons per strain than the At group (Fig. 2B; Welch two sample t-tests, $p < 0.0002$ in both cases). This is consistent with the

findings of Karasov et al. (2017), who detected a significant enrichment of singletons in strains derived from crop hosts compared to strains derived from *A. thaliana* isolated in the midwestern United States. There was no difference in the number of singletons per strain for the crop and env groups (Welch two sample t-test, $p > 0.05$), revealing a similarity in their genome compositions.

The core genomes of strains derived from At (N = 4438 core genes) was larger than that from env strains (N = 3255 core genes), which was larger than the core genome from strains derived from crops (N = 2947 core genes). These estimates are in agreeance with the patterns of clustering seen in Fig. 2.1 and of ANI in Table 2.2, such that the At-derived strains are the most similar within a group and thus had the greatest amount of overlap in gene content. Additionally, I sought to explore the genes essential for survival by comparing the identity of the core genomes for each group to the core genome of the whole pangenome. I found that 40-60% of each group's core genome was conserved in the core of the entire pangenome (39% for the At group, 54% for the env group, and 59% for the crop group). This is also reflective of greater relative divergence of the At group in that less than half of the gene clusters in the At core are present in the whole pangenome core. In contrast, for the env and crop groups, a majority of within-group core genes are conserved in the whole pangenome's core genome.

### *2.4.4 Selection in the pangenome*

Elevated AT (relative to GC) nucleotide content, which comes from a mutational bias toward AT-rich sequences (Hershberg and Petrov 2010), is believed to be a signature of reduced purifying selection. In fact, the GC content of the core (59.9%) is significantly greater than the GC content of the singletons (50.7%) as well as the entire accessory genome considered together (57.6%; Welch two-sample t-tests each with $p <$ 2.2e-16), consistent with the expectation of greater purifying selection on the core. These findings are consistent with the

results of Bohlin et al. (2017) who compared the GC content of core and accessory genomes for an array of organisms, including *P. aeruginosa*, and found that the accessory genomes had significantly lower GC content than the core genes. Further evidence of purifying selection acting on the core genome comes from my observation that GC content in the core genomes of the three groups are approximately the same (61% for the At and env group and 60% for the crop group).

To further examine the action of selection, I calculated Tajima's D for a subset of genes in the core and accessory genomes (Tajima 1989). Tajima's D is a test of neutral evolution that compares the number of pairwise differences and segregating sites; Tajima's D equals zero in cases of fully neutral evolution, whereas values significantly above zero are consistent with balancing selection, and values below zero are consistent with a selective sweep. Note, however that demographic processes can also impact Tajima's D, with high values consistent with population contraction and low values consistent with population expansion. For this exploration, I chose a set of 20 genes that are well-represented in all three groups of isolates, five with housekeeping functions and 15 with virulence functions (Table 2.3). Half of the 20 genes revealed evidence of non-neutral evolution (with significant p-values after Bonferroni correction for multiple comparisons); in all of these cases, Tajima's D was less than zero (Table 2.3) and thus consistent with either recent selective sweeps having removed diversity from the population or a population expansion. Two of the genes with significant Tajima's D were housekeeping genes GapA and GyrB, plus three core and five variable virulence factors (Table 2.3).

For the same set of 20 genes I consider whether the At, core, and env groups represent statistically significant ecological groups using unweighted Unifrac distance. Unifrac is a phylogenetic method by which to measure differences in communities as the fraction of the branch length of the tree that leads to one environment or the other (Lozupone and Knight 2005). I performed this analysis for the SCG core genome first, and found that the

Table 2.3: Tajima's D for a subset of 20 genes

| | FUNCTION | D | P | P.ADJUST | ANNOTATION |
|---|---|---|---|---|---|
| **VARIABLE GENES** | | | | | |
| **ALKA** | virulence | −1.75 | 0.0580 | 1.00 | 3-methyladenine DNA glycosylase |
| **FEPA** | virulence | −3.35 | 0 | 0 | Outer membrane receptor for ferrienterochelin |
| **HIPA** | virulence | −2.76 | 0.0000370 | 0.000740 | Permease of the DMT superfamily |
| **OMPV** | virulence | −3.50 | 0 | 0 | Outer membrane protein for murein synthesis |
| **RAYT** | virulence | −3.40 | 0 | 0 | REP element-mobilizing transposase RayT |
| **TRA5** | virulence | −3.17 | 0 | 0 | Transposase InsO and inactivated derivatives |
| **Y0BV** | virulence | −0.603 | 0.585 | 1.00 | Site-specific recombinase |
| **CORE GENES** | | | | | |
| **CTS/GLTA** | housekeeping | −0.357 | 0.767 | 1.00 | Citrate synthase |
| **GAPA** | housekeeping | −3.26 | 0 | 0 | Glyceraldehyde-3-phosphate dehydrogenase |
| **GYRB** | housekeeping | −3.35 | 0 | 0 | DNA gyrase/topoisomerase IV, subunit B |
| **PGI** | housekeeping | 0.726 | 0.472 | 1.00 | Glucose-6-phosphate isomerase |
| **RPOD** | housekeeping | −2.10 | 0.0120 | 0.240 | DNA-directed RNA polymerase, sigma subunit |
| **FLGI** | virulence | 0.681 | 0.500 | 1.00 | Flagellar basal body P-ring protein |
| **FLHA** | virulence | −3.55 | 0 | 0 | Flagellar biosynthesis protein |
| **FLIL** | virulence | 0.672 | 0.501 | 1.00 | Type IV pilus assembly protein, ATPase |
| **HCP** | virulence | −3.16 | 0 | 0 | Type VI protein secretion system component |
| **MBTH** | virulence | −0.698 | 0.519 | 1.00 | Regulates adenylation domains of NRPSs |
| **PILM** | virulence | −1.03 | 0.313 | 1.00 | Type IV pilus assembly protein |
| **PILO** | virulence | 0.948 | 0.358 | 1.00 | Flagellar motor switch protein |
| **PILT** | virulence | −3.38 | 0 | 0 | Type IV pilus assembly protein |

three groups represent distinct communities according to Unifrac (Table 2.4). This finding provides confidence that I have chosen appropriate ecological groupings for my strains. In fact, the three groups represented significantly different communities for the majority of genes compared (Table 2.4).

Two representative virulence factors that had significantly negative Tajima's D value (that is, they exhibit evidence of evolving under purifying selection) were Hcp and Tra5. For Hcp, the results of the Unifrac tests indicated that the At, crop, and env groups represented distinct communities. This was not the case for Tra5, for which the results of the Unifrac test indicated that the three groups were not evolving distinctly. Hcp is found in all isolates, Tra5 is found in only 56 of the 74 genomes, which is another indication of the differential effects of horizontal gene transfer and/or selection on these two genes. The individual gene trees reflect different evolutionary histories than the core gene tree; for example, the At group is not a monophyletic group in either single gene tree (Fig. 2.2). Additionally, no tree displays a perfect delineation between environmental and crop clusters; instead I see consistent overlap between a portion of crop and environmental strains. For example, in the Hcp tree, there is a group of environmentally-derived isolates that clusters with the At phylogroup seven isolates (CC1559, CSZ0279, and CCV0502), and in the Tra5 tree, there are five At group isolates (LABASB1, MAZAA1, BELLA1, REALA1, LUZEE6) that cluster with the majority of crop and env group genomes (Fig. 2.2). Thus, in these examples of the evolutionary histories of genes in this system I find imperfect phylogenetic delineation of the env and crop groups. Additionally, I found evidence of lateral transfer and selection among the three groups, such that in both the Hcp and Tra5 trees, at least some At-derived genes cluster with the crop and env groups as opposed to within a larger At group (Fig 2.2).

Additionally, for both trees I again find that host association does not resolve in patterns congruent with strain relatedness; rather, host association appears random with respect to placement on either the Hcp or Tra5 trees (Fig. 2.2). This is similar to what I found for the

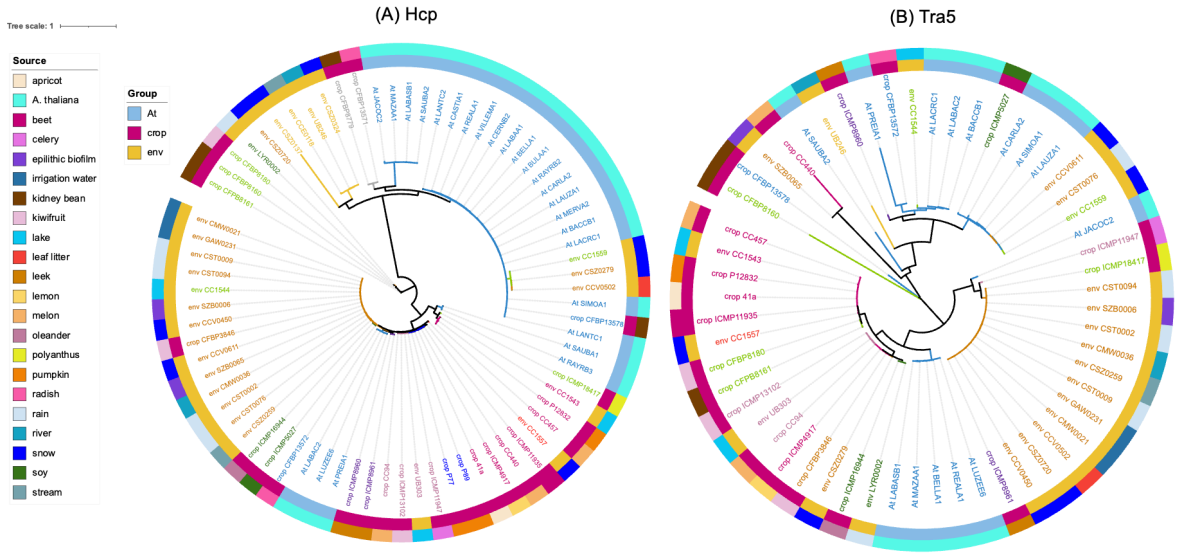



Figure 2.2: Maximum likelihood phylogenies for Hcp and Tra5.

single copy core gene tree (Fig. 2.1), indicating a true failing of host association to predict genetic relatedness. Thus, in these examples of the evolutionary histories of genes in this system I find some phylogenetic delineation of the env and crop groups, despite predicting that they would overlap. Additionally, I found evidence of lateral transfer and selection among the three groups, such that in both the Hcp and Tra5 trees, at least some At-derived genes cluster with the crop and env groups as opposed to within a larger At group (Fig 2.2). I also again find that host association is an overall poor predictor of genetic relatedness, as it does not resolve in patterns similar to the genetic delineations found in any of the three trees examined.

### *2.4.5 Functional categories significantly enriched per each group*

Having considered the relatedness of the At, crop, and env groups, I next asked what genes are significantly enriched across groups. Those enriched in all groups should be related to general housekeeping, cell viability, and ecological interactions, whereas those enriched in specific groups are more likely to reflect adaptation to source environments. For example, multidrug resistance genes are likely important for most *P. syringae* life histories, but functions related to pathogenicity are more likely to play a role in host environments. My phylogenetic analysis indicated that the env and crop groups are not ecologically distinct, thus I expect that they share more gene content than either does with the At group. This means that at least some genes useful for host interactions will be enriched in the env group, though those isolates have not necessarily been selected *in planta*. The At group is the relative outlier in that it has fewer singletons than the other groups (Fig. 2.3.B), so I expect that it is unlikely to contain a wealth of effectors. Below I examine the COG functional categories of genes significantly enriched across groups, and then I focus on the presence of known virulence factors, secreted effectors, and antibiotic resistance genes across groups. I focus on the categories or genes shared by all and specific to the individual groups, rather than pairwise comparisons between groups, in the hopes of identifying ecological adaptations to the particular source environments.

In addition to genes in the unknown or predicted functional categories, the categories enriched in all groups were G: carbohydrate transport and metabolism, I: lipid transport and metabolism, Q: secondary metabolite biosynthesis, T: Signal transduction mechanisms, and V: defense mechanisms. Several of these genes were components of the Type II (PulJ and PulK) and Type IV (PilP, PilY, and PilZ) secretion systems, and one, CheC, is known to regulate the cell motion via a flagellar switch (Park et al. 2004). Aside from those shared across groups, there were two categories of genes specific to the crop group; F: nucleotide transport and metabolism and X: Mobilome: prophages, transposons (Fig. 2.4). Category

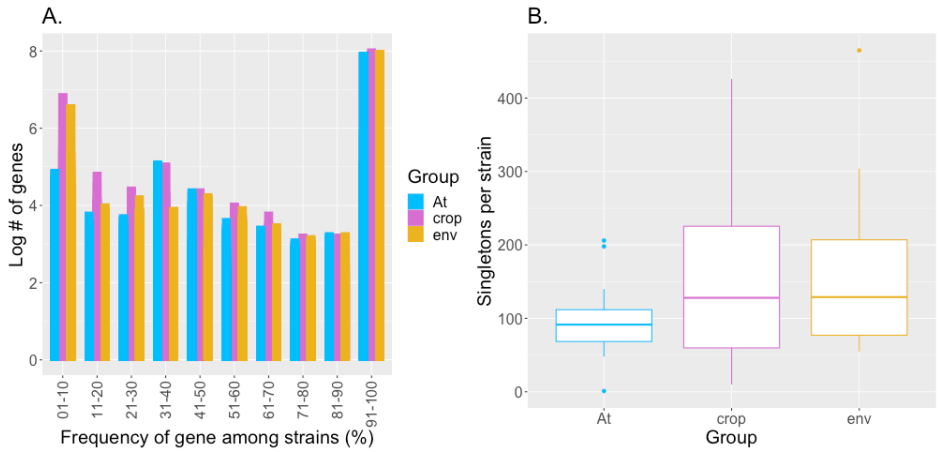


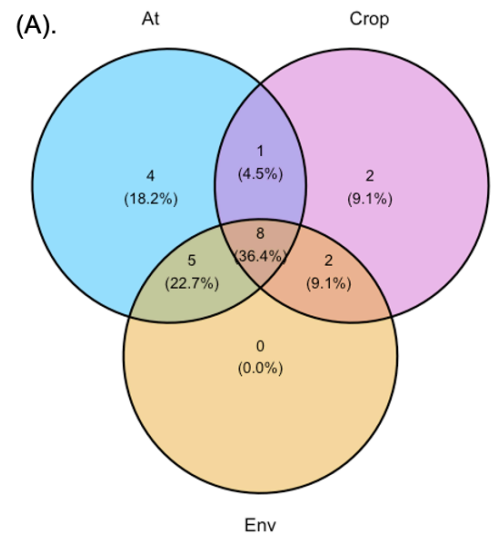

Figure 2.3: A. Frequency of genes in the pangenome. B. Median singletons per genome.

F was represented by a single gene, GDA1, which is a nucleoside diphosphatase conserved across plant-pathogenic microbes (Studholme et al. 2005). Category X was represented by two phage-related genes: AlpA, which has been shown to regulate prophage excision and pathogenicity in *P. putida* (Petitjean et al. 2017), and a hypothetical phage repressor protein. Finally, there were four categories of genes enriched only in the At group, including A: RNA processing and modification, J: Translation, M: Cell membrane biogenesis, and O: Posttranslational modification (Fig. 2.4). Two of these are related to Type I (aprE and a second serine protease) and two to Type II (PulL and PulM) secretion structure and function. Additionally, the At group contained CapA, a gene that guards against cold stress (Craig et al. 2021), and yidD, a component of a multi-gene system (Oxa1/YidC/SpoIIIJ) that inserts proteins into microbial and host cells (Yi and Dalbey 2009). The fact that these genes are present only in the *A. thaliana*-derived isolates indicates the specificity of their adaptive benefit; I expect that they are either directly beneficial in *A. thaliana* infections,

such as perhaps YidD, or benefit interactions with the *A. thaliana* microbiome, like those involved in Type II secretion.

### *2.4.6 Presence of virulence factors*

Having considered COG categories enriched per group, I focus next specifically on the presence of virulence factors, the subset of genes known to benefit host infections or microbe-microbe interactions. I found that 21 virulence factors were present in at least one isolate in all groups (Fig. 2.5). These fall into five categories: 1) flagellar and 2) pili genes related to cell motility, 3) genes involved in the type VI secretion system (T6SS), and genes involved in the synthesis of 4) alginate and 5) pyoverdine. The eight (38%) genes involved in flagellar synthesis include several genes that encode for structural proteins, including FlgG which encodes for a distal rod protein (Bouteiller et al. 2021), FlgH and FlgI which encode for L- and D- ring proteins respectively, FliP which encodes an export gate protein (Bouteiller et al. 2021), and FliI and FliM which encode for formation of the flagellar motor and switch (Dasgupta et al. 2000). The gene *fleN* regulates flagella number (Dasgupta et al. 2000). Additionally, four (19%) virulence factors shared by all groups were involved in T6SS structure and function. These include Hcp1, an extracellular component of the T6SS (Haapalainen et al. 2012), and ClpV1, which is an adenosine triphosphatase that provides essential energy for secretion of the Hcp1 protein (Mougous et al. 2006). In *P. aeruginosa*, it has been shown that HsiB1/VipA and hsiC1/vipB form a stable complex together, making a tail-like sheath that contracts to push the tube and spike of the T6SS through the bacterial envelope (Lossi et al. 2013). There were also three (14%) genes involved in alginate synthesis, AlgA, Alg8, and AlgU. Alg8 is an alginate polymerase (Oglesby et al. 2008) and AlgU is a regulator of alginate synthesis that has been shown to regulate other pathogenesis genes in DC3000 (Markel et al. 2016). Another three (14%) genes, MbtH-like, PvdH, and PvdS, are each required for expression of pyoverdine synthesis (Lamont and Martin 2003; Vandenende et




(B).

| | All groups |
|---|---|
| 1 | G - Carbohydrate transport and metabolism |
| 2 | I - Lipid transport and metabolism |
| 3 | L - Replication, recombination, and repair |
| 4 | Q - Secondary metabolites |
| 5 | R - Function predicted only |
| 6 | S - Function unknown |
| 7 | T - Signal transduction mechanisms |
| 8 | V - Defense mechanisms |
| | **At group** |
| 1 | A - RNA processing and modification |
| 2 | J - Translation |
| 3 | M - Cell membrane biogenesis |
| 4 | O - Posttranslational modification |
| | **Crop group** |
| 1 | F - Nucleotide transport and metabolism |
| 2 | X - Mobilome: prophages, transposons |

Figure 2.4: Significantly enriched functional categories (A) Venn diagram showing overlap of categories. (B) Categories per group.

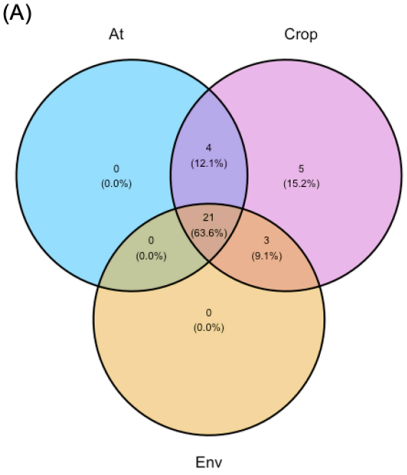
al. 2004; Drake et al. 2007). Pyoverdine is a siderophore that has been shown to contribute to virulence of *P. syringae* in tobacco (Taguchi et al. 2010). Finally, three (14%) genes are related to the type IV pilus, including PilG which is involved in pilus production (Darzins 1994), PilH which is involved in twitching motility (Darzins 1994), and PilT which encodes for the cytoplasmic motor complex (Tammam et al. 2013). These virulence factors reflect functions conserved across ecological sources and which are likely to be important for *P. syringae* with potentially divergent life histories.

The remaining virulence factors were either only present in the crop group or shared with the crop group (Fig. 2.5). Neither the At nor the env group had any specific virulence factors, but I found five virulence factors whose presence was specific to the crop group: AlgR, DotUI, FlhA, HsiG1, and WaaF. AlgR has been shown to be a response regulator of alginate synthesis (Peñaloza-Vázquez et al. 2004). WaaF is a known glycosyltransferase (Kievit and Lam 1997) found on the outer core of the cell-surface lipopolysaccharide (Lam et al. 2011). FlhA is a component of the flagellum assembly apparatus that is required for cell motility in *P. aeruginosa* (Fleiszig et al. 2001). DotU1 and HsiG1 both encode T3SS-associated proteins; in *P. aeruginosa*, dotU1 is involved in forming a cell envelope platform and HsiG1 is part of a complex that forms a bacteriophage-like tail (Lossi et al. 2013). Interestingly, each of these virulence factors is localized to the cell membrane, suggesting a role for ecological interactions in shaping their selective environment.

### *2.4.7 Presence of effectors*

Next, I focused on the presence of secreted effectors, which are a key component of *in planta* infection. I found a 'core' group of eight secreted effectors present in at least one strain in all groups, including AvrE, AvrPto, HopA, HopAG, HopAH, HopB, HopM, and HopO (Fig. 2.6). In fact, AvrE and HopB are part of the larger *P. syringae* core genome (Dillon et al. 2019B). HopA appears to be recently acquired by *P. syringae* (Baltrus et al. 2011)

(A)

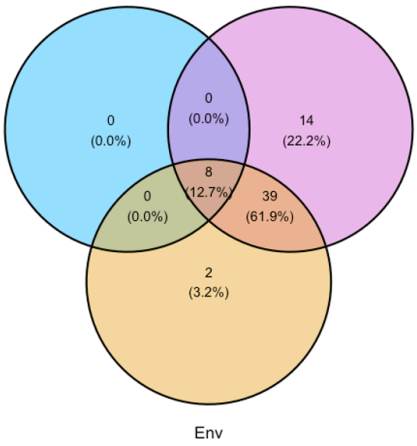



(B)

| | Crop group | | All groups | | |
|---|---|---|---|---|---|
| 1 | algR | 1 | alg8 | 15 | hsiC1/vipB |
| 2 | dotU1 | 2 | algA | 16 | mbtH-like |
| 3 | flhA | 3 | algU | 17 | pilG |
| 4 | hsiG1 | 4 | clpV1 | 18 | pilH |
| 5 | waaF | 5 | fleN | 19 | pilT |
| | Crop and At | 6 | flgG | 20 | pvdH |
| 1 | algB | 7 | flgH | 21 | pvdS |
| 2 | algC | 8 | flgI | | |
| 3 | fliA | 9 | fliG | | |
| 4 | motC | 10 | fliI | | |
| | Crop and env | 11 | fliM | | |
| 1 | algI | 12 | fliP | | |
| 2 | pilM | 13 | hcp1 | | |
| 3 | pilO | 14 | hsiB1/vipA | | |

Figure 2.5: Presence of general virulence factors. (A.) Venn diagram showing overlap of virulence factors. (B.) Virulence factor repertoires.

(A)




(B)

| | All groups | | Crop group | | Crop and env groups | | |
|---|---|---|---|---|---|---|---|
| 1 | AvrE | 1 | AvrB | 1 | AvrRpm | 21 | HopBO/HopX |
| 2 | AvrPto | 2 | AvrRpt | 2 | HopAA | 22 | HopBP |
| 3 | HopA | 3 | HopAD | 3 | HopAB/HopAY | 23 | HopBX |
| 4 | HopAG | 4 | HopAL | 4 | HopAF | 24 | HopC |
| 5 | HopAH | 5 | HopAQ | 5 | HopAI | 25 | HopD/HopAO |
| 6 | HopB/HopAC | 6 | HopAR | 6 | HopAM | 26 | HopF |
| 7 | HopM | 7 | HopBC | 7 | HopAS | 27 | HopG |
| 8 | HopO | 8 | HopBE | 8 | HopAT/HopAV | 28 | HopH |
| | | 9 | HopBI | 9 | HopAU | 29 | HopI |
| | Env group | 10 | HopBT | 10 | HopAW | 30 | HopK/AvrRps |
| 1 | HopAW | 11 | HopBU | 11 | HopAX | 31 | HopL |
| 2 | HopU | 12 | HopE | 12 | HopAZ | 32 | HopN |
| | | 13 | HopQ | 13 | HopBA | 33 | HopR |
| | | 14 | HopBO/HopX | 14 | HopBD | 34 | HopS |
| | | | | 15 | HopBF | 35 | HopT |
| | | | | 16 | HopBH | 36 | HopV |
| | | | | 17 | HopBK | 37 | HopW/HopAE |
| | | | | 18 | HopBL | 38 | HopY |
| | | | | 19 | HopBM | 39 | HopZ |
| | | | | 20 | HopBN | | |

Figure 2.6: Presence of secreted effectors. (A) Venn diagram showing overlap of categories. (B) Categories per group.

and has been shown to suppress PTI in many plant species (Dahale et al. 2021). HopAG is disrupted by a transposon in DC3000 (Vinatzer et al. 2005) and has been convergently lost in at least 3 pathovar *lachrymans* strains, suggesting it is recognized by cucumber (Baltrus et al. 2011). HopB is part of the conserved exchangeable effector locus (Xin et al. 2018) which contains effectors that contribute to virulence on many plants including tobacco and *A. thaliana* (Alfano et al. 2000). Thus, a number of secreted effectors are present in strains derived from both host and non-host environments.

One striking result was that the At group demonstrated a relative dearth of effectors, having only the core set shared by all groups, and none shared only with either the crop

or env isolates. One of two secreted effectors found only in the three isolates from aquatic sources (rain and snow), HopU (Fig. 2.6), has been shown to target several plant proteins (Nicaise et al. 2013). Thus, the environmentally-derived isolates contain genes representing functions that determine infection, even though these isolates were not from plant sources. The remainder of effectors detected were shared by the crop and env isolates, which shared more secreted effectors than were found in either group alone (Fig 2.5). These are numerous (39), and many have been shown to interact with plant hosts. For example, AvrRpm, HopAA, HopB, HopBA, HopF, HopO, and HopZ have been shown to elicit Effector Triggered Immunity (ETI) in *A. thaliana* (Laflamme et al. 2020), in addition to HopAS (McAtee et al. 2018). Evidence of extensive ETI in *A. thaliana* is in accordance with my finding of few effectors in genomes from that source, in that host recognition and an ETI response selects against the presence of the effector.

Finally, I consider the presence of antibiotic resistance genes across group of isolates. Of the five antibiotic resistance genes identified with MEGAres (Fig. 2.7), the three shared by all groups are associated with multi-drug resistance. Efflux pumps are a key part of bacterial defense against harmful chemicals often encountered in the environment (Tian et al. 2010). Resistance-nodulation-division (RNDs) systems are efflux pumps that are typically associated with resistance to multiple antimicrobial compounds, and have been shown to be involved in regulation of the T3SS in *P. aeruginosa* (Linares et al. 2005). The multidrug resistance loci MexE and MexF are part of a multidrug efflux pump transporter complex that regulates quorum-sensing molecules (Sawada et al. 2018). Finally, EmhC is part of a RND system that confers resistance to tetracycline among other antibiotics (Tian et al. 2010), which is a common component of agricultural soil (Zheng et al. 2020). The remaining two antibiotic resistance genes were shared by the env and crop groups. TtgB is part of a RND system; one example of an antibiotic it confers resistance to is toluene, an organic solvent emitted by plants under stressed conditions (Heiden et al. 1999). The other gene,

CpxAR, encodes a pair of proteins involved in resistance to multiple bactericidal antibiotics (Tian et al. 2016). Thus, the structure of the accessory genomes provides an indication of the selective pressures at play in the different groups.

## 2.5 Discussion

The genotypic and phenotypic variability of microbes, including *P. syringae*, has made them difficult to categorize them into distinct species (Gomila et al. 2017; Baltrus 2016; Berge et al. 2014). To what extent do groups of strains from different sources represent selection in different ecologies? I examined gene content variation among groups of *P. syringae* isolated from crop hosts, *A. thaliana*, and environmental sources in an attempt to uncover patterns of ecological distinction.

My single copy core gene phylogeny yielded two main results; one was that host association did not sync with pathogen phylogeny. This was not unexpected based on the previous findings of others (Dillon et al. 2019B; Baltrus et al. 2011; Sarkar and Guttman 2004), but indicated that the genes with host-specific functions are largely absent from the core genome, which suggests they are members of the variable genome. The second main result from Fig. 2.1 was that the A. thaliana-derived isolates formed a monophyletic group. In fact, I found that *P. syringae* isolated from *A. thaliana* are likely to represent an ecotype, as described by Cohan 2002. In particular, the At group meets each of 3 expectations for forming a stable ecotype (Cohan 2002; Ward et al. 2008). First, the core gene phylogeny (Fig. 2.1) shows that the At-derived strains fall into a single monophyletic clade. Second, Fig. 2.1 shows isolates sharing similar branch lengths, consistent with a star-phylogeny. Third, calculations of ANI revealed that the At group was more similar to itself than to either the env or crop groups, and in fact was more similar to an allopatric sample of *P. syringae* isolated from *A. thaliana* in North America than to either the env or crop groups from France. These findings reflect that *in planta* environments represented by crop vs. natural hosts impose

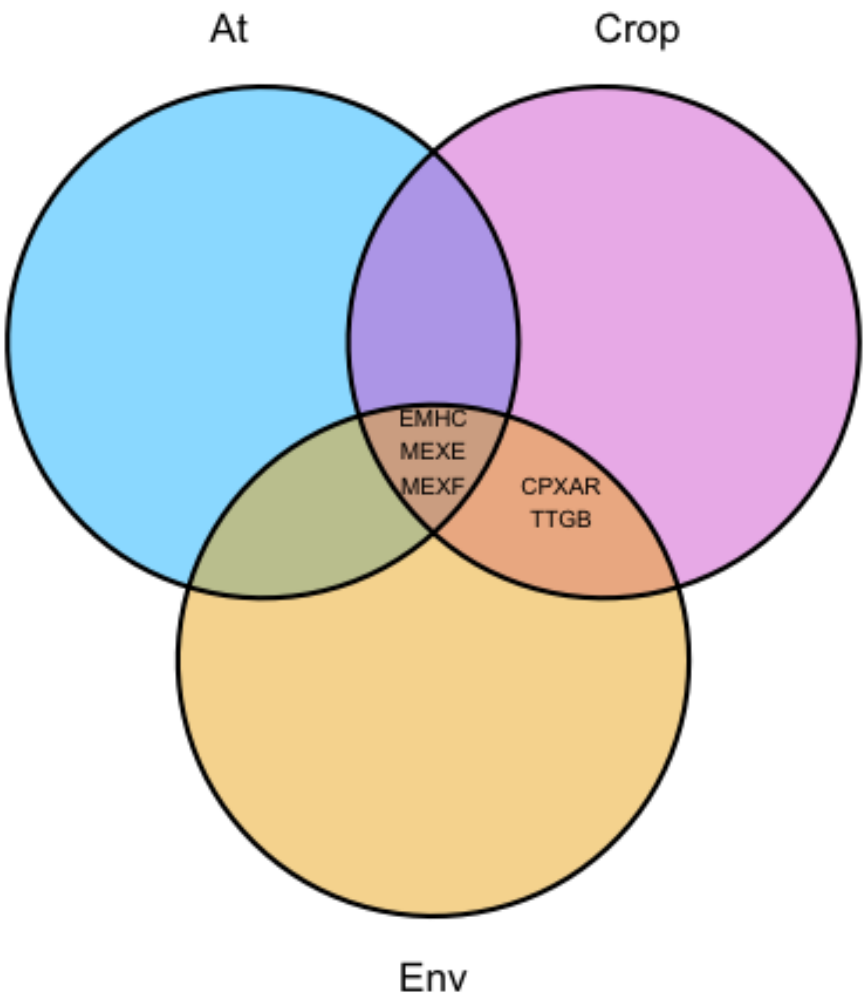



Figure 2.7: Venn diagram showing overlap of categories of antibiotic resistance genes.

fundamentally different selective pressures, such that many genes conserved and likely to be generally useful *in planta* (e.g. the Hop family of effector genes) are absent in the At group.

Previous studies have identified similar patterns in *A. thaliana*-isolated *P. syringae* from other locations. In a comparison of 18 *P. syringae* genomes from *A. thaliana* in the Midwest USA to 22 isolates from various crop hosts, the *A. thaliana* isolates had significantly fewer singletons than the crop-derived isolates (Karasov et al. 2017). Not only did I see the same pattern in in my comparison of French *P. syringae* isolates from crops vs. *A. thaliana*, I found that isolates from environmental sources also had significantly more singletons than the *A. thaliana*-derived isolates, in fact, to the same extent as did crop isolates. This was somewhat expected as closely related crop and environmental isolates have been identified previously (Bartoli et al. 2015; Monteil et al. 2013). Furthermore, some virulence factors underlie microbe-microbe interactions (Snelders et al. 2021), so it seems that at least some effectors should be present in the genomes of environmental isolates. Why these effectors are not present in At-derived isolates, however, is unclear. Karasov et al. (2017) found relatively few effectors were encoded by American *A. thaliana*-derived isolates and I found zero effectors specific to the French *A. thaliana* isolates in my comparisons. Interestingly, there is some evidence of a European ecotype amongst *A. thaliana*-derived isolates of *P. syringae*. Specifically, a large-scale study of such strains isolated across Germany found that a single, dominant Operational Taxonomic Unit within phylogroup 7 infects *A. thaliana*(Karasov et al. 2018). These strains are likely *P. viridiflava* (Karasov et al. 2018), and in fact the *P. viridiflava* reference genome, when included in the single-copy core gene phylogeny, clusters within the French At group (data not shown). Previous research has revealed the existence of two *P. viridiflava* phylogroups, 7 and 8 (Goss et al. 2005; Bartoli et al. 2014). Since the American *P. syringae* isolates from *A. thaliana* fall within phylogroup 2, it is possible that they are not *P. viridiflava*, or that *P. viridiflava* is more genetically diverse than previously thought.

Unlike to the At group, the environmental and crop isolates were not particular to any one phylogenetic cluster (Fig. 2.1). One group of environmental genomes corresponded to phylogroup 13, whereas many genomes from the env group were spread across phylogroups 1-3. Similarly, the crop group contains genomes that span phylogroups 1-4 and 7, plus a group of two genomes that form an early-branching, secondary phylogroup (Fig. 2.1). In addition to the core genome phylogeny, every individual gene trees built also fail to resolve the crop and environmental groups into their own clusters (Fig. 2.2). However, there is a clear delineation of environmental isolates that comprise phylogroup 1A in many trees, and the Unifrac distances of the crop-env comparisons often indicated that these lineages were distinctly evolving (Table 2.4). While the crop and env groups are too similar to represent distinct ecotypes, and discussion of these groups has focused on the fact that environmental and crop strains are found in the same phylogroup (Monteil et al. 2013), my analyses have revealed that there is ecological distinction between genes evolving in these two groups.

I predicted that strain-specific genes should be in the variable genome, and found evidence of selection (significant Tajima's D; Table 2.3) in the majority of variable genes that I checked. The variable genes also tended to have significant Unifrac values, indicating that they represented distinct lineages per each group (Table 2.4). This indicates that a good source of adaptive potential is the variable genome, even excluding singletons. Previous explorations of the variable genome have often focused on the adaptive potential of the singleton genome specifically (Fakhar et al. 2023; Wolf et al. 2016; Tautz et al. 2011; Wilson et al. 2005), which means that less is known about the variable content that is shared amongst isolates. My exploration of a few of these genes revealed some examples that show evidence of selection, many of which also had phylogenetic patterns consistent with distinct patterns of evolution for 'group' (At, core, or env). I focused specifically on virulence functions, but my analysis is far from exhaustive. Future work interested in the adaptive potential of the accessory genome should focus on elucidating the functions of those genes which are shared

Table 2.4: Unweighted Unifrac distances for a selection of 20 genes.

| | FUNCTION | GROUPS | UWSCORE | UWSIG | N_AT | N_CROP | N_ENV |
|---|---|---|---|---|---|---|---|
| **VARIABLE GENES** | | | | | | | |
| **ALKA** | virulence | At-crop | 0.954 | <0.0010 | 25 | 19 | 0 |
| **ALKA** | virulence | At-env | 0.981 | 0.092 | 25 | 0 | 4 |
| **ALKA** | virulence | crop-env | 0.729 | 0.246 | 0 | 19 | 4 |
| **FEPA** | virulence | At-crop | 0.716 | <0.0010 | 25 | 24 | 0 |
| **FEPA** | virulence | At-env | 0.983 | <0.0010 | 25 | 0 | 25 |
| **FEPA** | virulence | crop-env | 0.962 | 0.009 | 0 | 24 | 25 |
| **HIPA** | virulence | At-crop | 0.249 | 0.047 | 25 | 23 | 0 |
| **HIPA** | virulence | At-env | 0.744 | <0.0010 | 25 | 0 | 23 |
| **HIPA** | virulence | crop-env | 0.694 | 0.021 | 0 | 23 | 23 |
| **OMPV** | virulence | At-crop | 0.850 | <0.001 | 25 | 22 | 0 |
| **OMPV** | virulence | At-env | 0.970 | <0.001 | 25 | 0 | 22 |
| **OMPV** | virulence | crop-env | 0.954 | <0.001 | 0 | 22 | 22 |
| **RAYT** | virulence | At-crop | 0.663 | 0.004 | 22 | 20 | 0 |
| **RAYT** | virulence | At-env | 0.874 | <0.0010 | 22 | 0 | 24 |
| **RAYT** | virulence | crop-env | 0.746 | <0.0010 | 0 | 20 | 24 |
| **TRA5** | virulence | At-crop | 0.679 | 0.564 | 14 | 22 | 0 |
| **TRA5** | virulence | At-env | 0.598 | 0.665 | 14 | 0 | 22 |
| **TRA5** | virulence | crop-env | 0.556 | 0.76 | 0 | 22 | 22 |
| **UBIH** | virulence | At-crop | 0.985 | <0.0010 | 25 | 24 | 0 |
| **UBIH** | virulence | At-env | 0.992 | <0.0010 | 25 | 0 | 25 |
| **UBIH** | virulence | crop-env | 0.613 | 0.03 | 0 | 24 | 25 |
| **YOBV** | virulence | At-crop | 0.970 | <0.0010 | 25 | 12 | 0 |
| **YOBV** | virulence | At-env | 0.998 | <0.0010 | 25 | 0 | 19 |
| **YOBV** | virulence | crop-env | 0.971 | 0.002 | 0 | 12 | 19 |
| **CORE GENES** | | | | | | | |
| **GAPA** | housekeeping | At-crop | 0.302 | 0.106 | 25 | 24 | 0 |
| **GAPA** | housekeeping | At-env | 0.551 | 0.04401 | 25 | 0 | 25 |
| **GAPA** | housekeeping | crop-env | 0.506 | 0.248 | 0 | 24 | 25 |
| **GYRB** | housekeeping | At-crop | 0.989 | <0.0010 | 25 | 24 | 0 |
| **GYRB** | housekeeping | At-env | 0.996 | <0.0010 | 25 | 0 | 25 |
| **GYRB** | housekeeping | crop-env | 0.687 | 0.007 | 0 | 24 | 25 |
| **SCG CORE** | housekeeping | At-crop | 0.983 | <0.0010 | 25 | 24 | 0 |
| **SCG CORE** | housekeeping | At-env | 0.987 | <0.0010 | 25 | 0 | 25 |
| **SCG CORE** | housekeeping | crop-env | 0.683 | 0.008 | 0 | 24 | 25 |
| **FLHA** | virulence | At-crop | 0.437 | 0.02 | 25 | 24 | 0 |
| **FLHA** | virulence | At-env | 0.650 | 0.006 | 25 | 0 | 25 |
| **FLHA** | virulence | crop-env | 0.504 | 0.032 | 0 | 24 | 25 |
| **HCP** | virulence | At-crop | 0.752 | <0.0010 | 25 | 24 | 0 |
| **HCP** | virulence | At-env | 0.917 | <0.0010 | 25 | 0 | 25 |
| **HCP** | virulence | crop-env | 0.773 | 0.01 | 0 | 24 | 25 |
| **PILT** | virulence | At-crop | 0.992 | <0.001 | 25 | 24 | 0 |
| **PILT** | virulence | At-env | 0.993 | <0.001 | 25 | 0 | 25 |
| **PILT** | virulence | crop-env | 0.681 | 0.02501 | 0 | 24 | 25 |

but not universal.

This work informed our understanding of the adaptive patterns of *P. syringae*. My results indicate that *A. thaliana*-derived isolates of *P. syringae* are likely to function as a stable ecotype (Cohan 2001). This was not the case for isolates from crop and environmental sources, which exhibited a less rigid delineation from each other than either did from the At group. Nonetheless, the env and crop groups frequently exhibited evidence of evolving in distinct patterns per group, even though they share phylogroups. My results are largely consistent with previous findings but provide the first direct comparison of *P. syringae* isolates from *A. thaliana*, crops, and environmental sources.

# CHAPTER 3

# COSTS OF LOCAL ADAPTATION IN *PSEUDOMONAS SYRINGAE*

## 3.1 Abstract

Are pathovars truly adapted to their hosts-of-isolation? To answer this question, I performed a series of controlled infections of five plant-pathogens in their own, and in each other's, host-of-isolation. I first compared the effector repertoire of the pathogens, and predicted that for *in vitro* experiments, DC3000 would have a fitness cost relative to NP29, since DC3000 contained the most effectors and NP29 contained the fewest. Instead, I find that while NP29 does have an advantage across all environments, DC3000 does not demonstrate a cost. I thus predict that *in planta*, DC3000 will have an advantage due to its wealth of effectors, but I do not predict a reciprocal cost for NP29, since it appears from the *in vitro* experiments that pathogens have methods to mitigate costs of effectors. My predictions are in accordance with my results; I find that DC3000 is the best *in planta* grower overall, though NP29 did not demonstrate a cost relative to other pathogens. I also detected evidence of local adaptation in two of five host-pathogen pairs, in that they grew to significantly higher carrying capacity on the host-of-isolation vs. all other hosts. These results demonstrate that effector load is not the only determining factor for infection success; e.g., host ecology does make a difference. These results also demonstrate that pathovars are not necessarily adapted to their hosts-of-isolation, which suggests it would be wise to divorce strain classification from host phenotype altogether.

## 3.2 Introduction

Plant pathogens are often classified into pathovar systems based on their host range (Bull and Koike 2015). For example, the well-characterized *Xanthomonas* genus comprises several species that are further divided into pathovars with intra-specific race designations to describe the host ranges of groups of isolates (Jacques et al. 2019). Pathovar systems assume adaptation of the pathogen to its host-of-isolation, but this assumption typically remains untested. That host-of-isolation is an indicator of local adaptation, in which an isolate performs better on that host relative to others, should not be assumed. After all, where a microbe is found indicates how well it does relative to other microbes on that host rather than how well it does on a focal host relative to other hosts (Read and Taylor 2001).

The agriculturally important bacterium, *P. syringae*, is an opportunistic pathogen with a broad host range that includes crop plants, natural weeds such as *A. thaliana*, and non-host environments (Baltrus et al. 2016; Xin and He 2018). The species complex is divided into more than 60 pathovars (Dillon et al. 2019B), although design of a suitable classification system has proven difficult (Gomila et al. 2017; Baltrus et al. 2016; Bull and Koike 2015). First, there are several examples of pathovars infecting alternative hosts. For example, pathovar *P. syringae tomato* caused a recent outbreak of bacterial leaf spot disease on kale (Koike et al. 2017). *P. syringae* pathovar *tabaci*, rather than coffee pathovar *P. syringae garcae*, was found to be the causal agent of infections on coffee (Destefano et al. 2010) and, in 2012, the crucifer pathovar, *P. syringae* pathovar *alisalensis*, was first reported on *Avena storigosa* (bristle oat; Ishiyama et al. 2013). Second, pathogenic varieties of *P. syringae* have been isolated from several non-host sources including snow, rain, and irrigation waters (Monteil et al. 2013). Third, individual strains vary in their host breadth, with some strains appearing to be generalists (Morris et al. 2019). Finally, host association explains only a small portion of *P. syringae* core genome variation (Sarkar and Guttman 2004), indicating an incongruence between host range and phylogeny, and revealing a conflict between phenotypic

and genomic methods of classification. It thus remains an open question whether *P. syringae* strains are adapted to their hosts-of-isolation.

One way to test for local adaptation is through artificial selection experiments. Here the expectation is that selection on one host would lead to reduced performance on other hosts due to fitness tradeoffs (Bell and Reboud 1997; Jasmin and Kassen 2007). For example, Gould (1979) evolved populations of mites on cucumber and discovered a decline in fitness on lima bean (Gould 1979). Similarly, Fry (1990) evolved populations of mites on lima bean and found that as fitness on lima beans increased, fitness on tomato declined (Fry 1990). However, counterexamples exist. In particular, Meaden and Koskella (2017) evolved *P. syringae* pathovar *tomato* on both *A. thaliana* and tomato and found that selection on *A. thaliana* created generalists with higher relative fitness on both plants. Even in a highly simplistic, non-host environment, the response of *E. coli* to selection in a glucose-limited environment was positively correlated across replicate populations in 5 out of 6 environments (Ostrowski et al. 2005). That said, in the 6th (melibiose-limited) environment, a few mutants exhibited a trade-off in fitness, indicating context-dependent antagonistic pleiotropy as one molecular mechanism at play. Taken together, it appears that fitness tradeoffs are often, but not necessarily, present in adaptation to novel hosts. One common example of genes incurring fitness trade-offs are pathogen effectors, a type of secreted virulence factor that benefits the pathogen by influencing the microbes or hosts surrounding the cell (Barrett et al. 2009). The plant microbiome is dynamic, and ecological interactions between microbes can alter survival probability and the outcome of infection (Read and Taylor 2001). Plants and pathogens interact through a system in which hosts recognize secreted effector proteins, introducing the possibility of arms race dynamics in which pathogens evolve to escape recognition and hosts counter-adapt to again recognize the pathogen and trigger a resistance response (Boller and He 2009; Jones and Dangl 2006; Dangl and Jones 2001). Since binding to host factors represents the potential for infection success and for host recognition with an ensuing immune

response, pathogen effector genes can be considered 'double edged swords' (Russell et al. 2015). Still, conventional wisdom is that the pathogen should be winning the evolutionary arms race, due to its shorter generation times and larger population sizes (Kaltz and Shykoff 1998). Thus, one expects that locally adapted pathogens will achieve a higher, sustained titer within their preferred hosts.

Herein, I use a suite of hosts that include four crops and the model weed *A. thaliana* to test to what extent strains are specialized to their hosts-of-isolation vs. capable of thriving in multiple hosts. To generate hypotheses about the relative performance of strains, I performed an initial comparison of their effector repertoires. On the logic that effectors come with two inherent risks: (1) a risk of host recognition (Block and Alfano 2011) and (2) a metabolic cost of maintaining a gene that is only advantageous in some environments (Morris et al. 2012), I predict that the strain with the highest effector load should incur a relative cost when growing *in vitro*, while the strain with the lowest effector load should have a relative advantage. I make predictions about the outcome of the *in planta* experiments based on the outcome of the *in vitro* experiments.

I focus on carrying capacity $K$ achieved by each isolate in the designated condition, because this is a direct, calculable proxy for fitness. I first evaluate the patterns of growth of each strain in three abiotic (host-free) environments. I use three lab media, LB, KB, and a media constructed in-house that consists of macerated *A. thaliana* plants in sterile water (Arabidopsis Broth, or AB). I expect the pathovars to perform relatively consistently, with all strains performing better on KB, which is a selective media for Pseudomonads, than LB, which is typically used to cultivate *E. coli* (discussed in detail in Methods). I furthermore expect that NP29, the strain collected from *A. thaliana*, will outperform the others in the AB media.

I then perform controlled infections of each strain on each of five host plants, including their host-of-isolation. Here, I specifically address whether each pathogen performs best

on its host-of-isolation relative to the other hosts. This naïve expectation assumes local adaptation of the strains to their host-of-isolation, which I test with an ANCOVA with hosts as fixed effect and plate as covariate, followed by planned contrasts for host-of-isolation versus all other hosts. I also ask whether there is a 'best' pathogen that outperforms all others across associations, which I test with post-hoc contrasts. Finally, I consider infection trajectories in light of evolutionary theory. I expect to see a crash in pathogen population size if the host is winning the arms race or if the pathogen overexploits the host. If the pathogen is tolerating or has successfully evaded the host immune response, I expect to see the absence of such a crash, i.e., the 'maintenance' of a high pathogen titer *in planta*. I interpret the maintenance of prolonged infection as evidence of adaptation of the pathogen.

## 3.3 Methods

Strain and environment selection I selected virulent strains for which high-quality genomes were already publicly available, including pathovars of bean (1448A), radish (ES4326) and tomato (DC3000 and A9), plus a strain isolated from the weedy species, *A. thaliana* (NP29). 1448A was isolated from bean in Ethiopia in 1985 (Joardar et al. 2005). ES4326 is a streptomycin-resistant strain of PSM4326, isolated from diseased radish in Wisconsin in 1965 (Sarris et al. 2013). DC3000 was isolated from tomato in 1960 in Guernsey Island, UK (Buell et al. 2003). A9 was isolated from tomato in Yolo County, California in 1996 (Kunkeaw et al. 2010), and NP29 was isolated from *A. thaliana* in Michigan in 2002 (Barrett et al. 2011). Isolates were obtained by kind donations from colleagues.

I chose to use two abiotic environments that our strains would likely encounter during cultivation in the lab, Lysogeny Broth (LB) and King's Medium B (KB), plus one novel media created in-house. LB is a standard nutrient-rich medium containing the carbon sources tryptone, yeast, and salt. Yeast provides organic compounds while tryptone provides amino acids. KB is commonly used to culture Pseudomonads as it allows visual confirmation of

the fluorescent coronatine phenotype (Hwang et al. 2005). KB contains proteose peptone, glycerol, and salt. The proteose peptone in KB provides amino acids from bovine milk, while glycerol provides a carbon source. Finally, I made *Arabidopsis* Broth (AB) in-house from a collection of mixed *A. thaliana* plants collected from various sites in the Midwest, USA. Plants were washed with sterile water and ethanol to remove particulate and loosely associated microbes before maceration and filtering through a mesh of 0.2 microns using RO water. The resulting broth was then autoclaved twice for 60 minutes to ensure sterility and stored at 4°C before equilibration to room temperature for experimentation. This broth thus contains a suite of chemical elicitors present in the *Arabidopsis* host environment.

In an attempt to pair my pathogens with their native host, I made an effort to obtain the cultivar of isolation or, if this cultivar was no longer available, I obtained a phenotypically similar cultivar grown in the same geographic region in the same years (Table 2.1). For example, 1448A was isolated from *Phaseolus vulgaris* in Ethiopia in the 1960s but no cultivar data was available; I therefore found a bean cultivar from Kenya that was grown in Kenya in the 1960s (*P. vulgaris* cv. African Premiere). RRS10 is a representative Midwest *A. thaliana* ecotype, which would be very similar to NP29's native host (Platt et al. 2010).

Identification of secreted effectors A recent study gathered sequences from all known *P. syringae* effectors, cataloging 14,613 sequences spanning 70 gene families and 523 alleles as the PsyTec Compendium (LaFlamme et al. 2020). They published the sequences from this database in the supplemental methods for the paper, which means it is freely available for download. I downloaded and used their supplemental data to build a custom BLAST database using BLAST+ for the command line (Camacho et al. 2009). I used the command makeblastdb to assemble the PsyTec database and then blastn to compare my genomes against it. I limited my analysis to matches with at least a 99% identity to the query sequence.

### *3.3.1 Growth* in vitro

To estimate growth rate $r$ and carrying capacity $K$ of each pathogen in LB, KB and AB media types, I measured optical density over 20 hours of incubation. Bacteria were brought to comparable physiological conditions with a period of acclimation in which a single colony was picked from KB agar and grown in liquid KB to stationary phase the night before the assay. In the morning, overnight cultures were diluted 1/10 into fresh KB and grown for 6 more hours. Finally, I randomly positioned a 1/200 dilution of each acclimated culture into a well of a sterile, optically transparent 96-well plate containing 200 $\mu$L of liquid media. I incubated the plates at 28°C in a Tecan plate reader set to shake at an amplitude of 2 mm for 10 seconds, then measured absorbance at 600 nm with 25 flashes, every 10 minutes for 20 hours. I repeated this assay for a total of six replicates. Each well was normalized to its own initial optical density, removing the contribution of the media to optical density estimates.

Statistical analyses were performed in R v. 4.2.2 (R Core Team 2021) using the growthcurver package (Sprouffske and Wagner 2016). Growthcurver is an R package that fits growth curve data to a standard form of the logistic equation with the following parameters: the intrinsic growth rate of the population, $r$, the initial population size, *N0*, and the carrying capacity, $K$. The intrinsic growth rate of the population, $r$, is the growth rate that would occur if there were no restrictions imposed by population size. Carrying capacity $K$ is defined as the maximum optical density reached over the course of the experiment. The logistic equation describes the population size $Nt$ at time $t$ as: $Nt = K/1 + ((K - N0)/N0) * e - rt$ . Sample curves for each pathogen are pictured below (Fig 1).

To calculate the number of generations, I used the standard equation (OD600final – OD600initial)/ln(2), where OD600final is the carrying capacity $K$. Time was transformed from seconds to hours for all calculations.

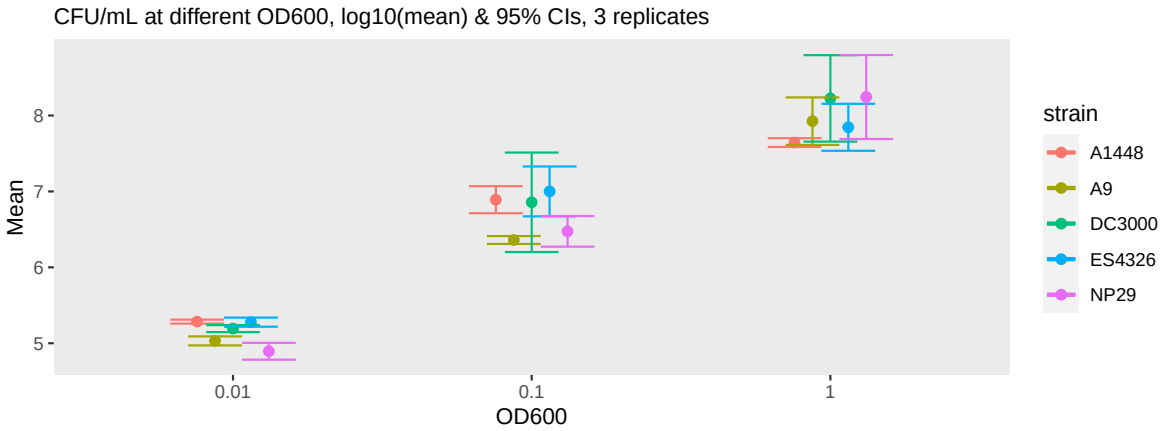



Figure 3.1: Mean and 95% CI of CFU/mL over 3 replicates of pathogens grown at three optical densities

### *3.3.2 OD600 correlation to CFU/mL*

My measurements of strain performance were calculated on the optical density data for each of the 5 strains. First, I sought to confirm that optical density could serve as a proxy for CFU/mL in the same way across strains. To do this, I grew overnight cultures of each strain, seeded by a single colony, pelleted cells to remove media, and resuspended them in sterile buffer to common optical densities of 1, 0.1, and 0.01. I then diluted the cultures for spread-plating, and enumerated colonies after incubation at 28°C for 2 days. The mean and 95% CIs for CFU/mL are shown for each of the 5 isolates in Fig. 3.1. I verified that there was no effect of strain on CFU/mL using ANOVA ($F = 0.821$, $p = 0.516$) in which strains were treated as fixed effects. Thus, the population sizes of each strain were indistinguishable at a particular optical density.

### *3.3.3 Growth* in planta

The *in planta* growth experiments were performed using all five hosts and all five strains simultaneously, in each of four fully randomized blocks that included one replicate of each host-strain combination. Experiments were performed on hosts synchronized to the same

developmental stage. This was accomplished by planting bean and radish 12 days after *Arabidopsis* and tomato and using *Arabidopsis* and tomato seedlings that were 21 days post emergence (thus, radish and bean that were 9 days post emergence). Seeds were sewn into autoclaved potting soil four inch 18-cell flats and covered with clear plastic domes until germination. I watered as needed by soaking and supplemented seedlings with Jack's 15-30-15 fertilizer one week prior to infection. Plants were grown with a 16-hour photoperiod at 22°C. Flats were rotated within and among shelves every 48 hours to minimize effects of growth chamber position.

Each of the five pathogens were grown for plant inoculation by selecting single clones from KB agar and seeding 5 mL liquid KB cultures grown at 28°C, shaking at 175 rpm for 12 hours. Overnight cultures were diluted 1/1000 into 200 mL fresh KB and grown for 5 hours until normalizing to OD600=0.2. Pathogens were applied to plants by spray inoculating leaves until they were dripping wet, using sterilized amber glass bottles with plastic spray fixtures. I allowed plants to dry for 12 hours before sorting into random order for the duration of the experiment. To discourage cross-contamination, a checkerboard pattern was used in which every-other cell was skipped so that plants would not touch.

Plants were harvested one, three, and five days after spray inoculation. For each plant, two randomly selected whole leaves (whole rosettes for *A. thaliana*) were picked and washed in 70% ETOH and sterile water for 30 seconds before taking two sets of six standard (six-mm) hole punches. Samples were placed directly into ice-cold 50% glycerol before storage at -20°C. One set of samples was saved as a backup, and total DNA was extracted from the other using the protocol of Mayjonade et al. (2016). First, glycerol was removed by pipetting. Then sterile, two mm ceramic beads and 300 $\mu$L of lysis buffer supplemented with RNAseA at 0.3 mg/mL was added to leaf material before maceration in a Genogrinder at 1750 rpm for three minutes. After homogenizing, I added another 300 $\mu$L of lysis buffer and incubated samples for one hour at 65°C, mixing by inversion every 10 minutes. After

lysis, I added 1/3 volume of 5 M potassium acetate to the samples, mixed by inversion, and centrifuged for 10 minutes to precipitate cellular debris. Finally, I purified the extracted DNA with a solution of homemade SPRI beads at 0.4%.

Titers were determined with qPCR on purified DNA. I first quantitated DNA in a 96-well plate reader using Picogreen and diluted to approximately 30 ng/$\mu$L after confirming purity with a Nanodrop. qPCR was performed using PowerUP SYBR MasterMix (catalog number A25741) and a reaction volume of 10 $\mu$L in 384-well plates. I used the primers for the single copy gene *oprF* (Ross and Somssich 2016) to quantify the abundance of my *P. syringae* strains after verifying specificity by confirming that only a single product was amplified. In addition, I confirmed that the standard curves were consistent for each of the five strains by generating standard curves, doubly replicated for each pathogen, consisting of 1/10 dilutions from 100 ng/$\mu$L to a concentration of $<$ 10 ng/$\mu$L (no amplification). Each of these correlation coefficients were above 0.99. To confirm there was no effect of pathogen on CT, I used a two-way ANOVA (OD $\sim$ CFU/mL * pathogen). I found a significant main effect of CFU/mL (F = 129.138, $p <$ 2e-16), but no effect of pathogen (F = 0.313, $p = 0.868$) or the interaction between CFU/mL and pathogen (F = 0.917, $p = 0.459$). To generate the zero time-point, which represents the baseline titer given no infection, I took the mean per cultivar of the mock treatment at Day 1.

Average coefficients from the standard curves used for titer calculations were $b = 34.9$ and $m = -3.7$. To calculate absolute titer, I used the standard formula: Log Quantity $= 10^{Ct-b/m}$. I took carrying capacity $K$ as the highest titer reached over the course of an infection and the calculated number of generations using the same methods used for the abiotic experiments. I estimated growth rates by fitting Baranyi curves (Baranyi and Roberts 1995) to infection curves using the R package growthrates (Mira et al. 2017). All infections were fit with $r^2 > 0.82$, and 20 out of 25 had $r^2 > 0.90$.

### *3.3.4 Phylogenetic Analysis*

I sought to determine the phylogenetic relationship among pathogens in order to consider its influence on infection patterns. To do this, I used anvi'o v. 7.1 (Eren et al. 2015). In brief, I built a pangenome by: (1) simplifying FASTA definition lines, (2) generating anvi'o contigs databases per each genome, (3) identifying open reading frames using Prodigal (Hyatt et al. 2010) v2.60X, (4) annotating genes with functions using the NCBI's Clusters of Orthologous Groups (Tatusov et al. 2000), (5) identifying single copy core genes using HMMER (Eddy 2008) v3.2.1X and an included collection of HMM profiles spanning bacteria and archaea, (6) building an anvi'o genomes storage database, and finally (7) building the pangenome using the script anvi-pan-genome. After building the pangenome using the steps described above, I extracted the 3,548 single copy core gene clusters shared by each genome and used trimAl (Capella-Gutierrez et al. 2009) to remove nucleotide positions that were gap characters in more than half of the sequences. I then used IQ-TREE (Nguyen et al. 2015) with the general matrix model 'WAG' to infer a maximum likelihood tree with 10,000 bootstraps.

### *3.3.5 Statistical Analyses*

For both *in vitro* and *in planta* experiments, I used two-way ANCOVA with plate as covariate, and plant and pathogen as fixed effects and including their interactions. The models for all ANCOVAs were the same ($K \sim$ environment * pathogen + plate). In all cases I confirmed that the assumptions of the ANCOVA were met prior to calculation of the model. Specifically, I confirmed that there was linearity of regression slopes, and a lack of interaction between the covariate "plate" and either grouping variable "plant" or "pathogen" ($p > 0.05$). To check Akaike Information Criteria and perform tests for homoscedasticity, I used the R packages AICcmodavg (Mazerolle 2023) and performance (Lüdecke et al. 2021) respectively. Data were log transformed before ANCOVA. For the *in vitro* experiments, I followed the two-way ANCOVA with post hoc contrasts, i.e., a Tukey's Honest Distance test on all pair-

wise comparisons. Additionally, I performed ANOVAs to test for an effect of pathogen on $K$ in the AB environment ($K \sim$ pathogen), and to test for the variability of NP29 across environments ($K \sim$ environment). To test for a best/worst pathogen and for local adaptation of the pathogens in the *in planta* experiments, I followed the two-way ANCOVA with posthoc contrasts to determine the relative performance of DC3000 and NP29, and planned contrasts comparing performance on the native host vs. all other hosts for each pathogen.

## 3.4 Results

### *3.4.1 Effector repertoires*

I compared effector repertoires of the five strains in order to develop hypotheses about their relative performance (Fig 2.6). Though effectors are frequently considered to be determinates of a strain's host range (Dillon et al. 2019B), it is unclear to what extent they affect growth *in vitro*. One simple expectation is that maintenance of effectors is metabolically costly (Iwasaki and Medzhitov 2015), so that a strain with more effectors would experience a greater cost when growing in environments lacking a susceptible host and thus, a possible benefit. The strain with the largest effector load in this experimental system is DC3000 (Fig. 2.6A), suggesting that DC3000 should grow relatively poorly in abiotic environments. Conversely, NP29 should grow relatively well in abiotic environments as it contains few effectors (Fig. 4A), a trait previously reported for many *P. syringae* strains isolated from *A. thaliana* (Karasov et al. 2017).

The five pathogen strains contained numbers of effector families ranging from six to 36 (Fig. 4A). Two effector families were present in all five strains: AvrE and HopM (Fig. 4B). These two families are broadly distributed across *P. syringae* and were each identified as 'core' effectors in a larger pangenome analysis of 494 strains (Dillon et al. 2019B). AvrE was the first identified Type III secretion effector (Napoli and Staskawicz 1987) and it, along

| strain | # effector families |
|---|---|
| A9 | 23 |
| DC3000 | 36 |
| ES4326 | 29 |
| 1448A | 22 |
| NP29 | 6 |

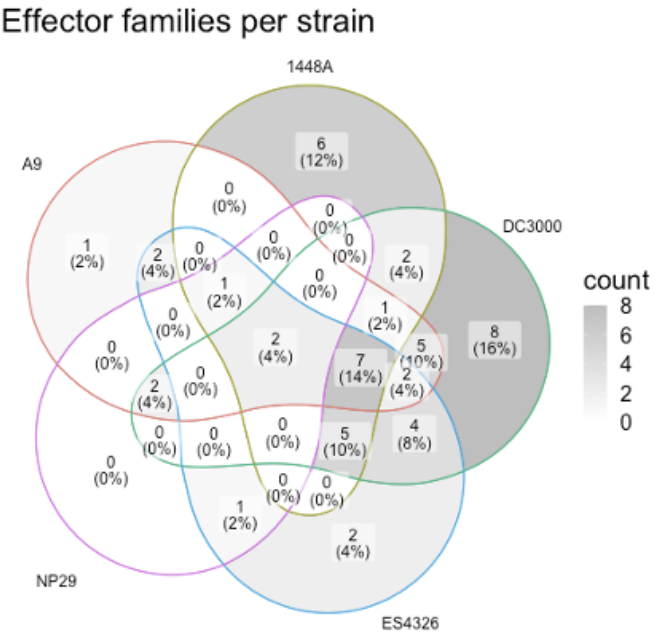



Figure 3.2: Effector repertories. (A.) Number of effector families per strain. (B.) Venn diagram showing overlap of effector family identity.

with HopM, are part of the canonical Type III secretion system (T3SS) pathogenicity island (Dillon et al. 2019B). There were seven effectors shared by all but NP29. (Fig. 4B), including HopAA, HopAB, HopAF, HopAH, HopAS, HopI, HopR. These are additional widespread effectors that have generally been well-characterized and function to increase pathogenicity during infection; for example, HopR has been shown to suppress the plant immune system by blocking callose deposition (Kvitko et al. 2009), and HopI has been shown to suppress accumulation of the defense compound salicylic acid by the plant (Jelenska et al. 2010).

Genes specific to individual strains include any that have been maintained because they are beneficial in the strain's selective environment. DC3000, which had the highest number of effector families at 36, contained eight specific effector families: AvrPto, HopAI, HopAM, HopC, HopE, HopH, HopN, HopU. Most of these have known virulence functions *in planta*; for example, AvrPto suppresses the basal immune response in tomato and *Arabidopsis* (Deslandes and Rivas 2012), HopN suppresses cell death during the second phase of the plant

immune response (Lopez-Solanilla et al. 2004), and HopAM has been shown to increase the virulence of the pathogen under stressful, drought conditions *in planta* (Goel et al. 2008). Strains 1448A and ES4326 each contained 22 and 29 effector families, with six and two specific effector families, respectively. ES4326 contains HopZ, which comprises a diverse family of effectors that cause disease in many plants including *A. thaliana* (Lewis et al. 2008), and 1448A contains AvrB, which is particularly interesting because though it is recognized by *A. thaliana* and soybean, molecular recognition is suppressed by an additional effector, AvrRpt2 (Russell et al. 2015). Strain A9 contained 23 effector families, one that was strain-specific: HopAZ, which have been shown to elicit Effector Triggered Immunity in *A. thaliana* (Nikolić et al. 2023). Note that there were no effector families specific to NP29. These effector repertoires reveal a diverse but strain-specific suite of strategies used by *P. syringae* during infections.

### *3.4.2 Growth* in vitro

This set of experiments consisted of growing five pathogens in three abiotic environments over 18 hours. Strains grew repeatably, with similar patterns of lag phase, exponential growth and then a plateau as carrying capacity was reached. $K$ ranged from OD600=0.27-0.77 (Table 3.1), with mean $K$ of 0.418, 0.529, and 0.362 for the AB, KB, and LB environments respectively.

There were significant differences in $K$ across environments and strains (Fig. 3.2). The results of a two-way ANCOVA indicated significant main effects of environment ($F = 4.177$, $p = 0.02047$), pathogen ($F = 5.333$, $p = 0.001$), and plate ($F = 3.066$, $p= 0.024$). There was no significant interaction between pathogen and environment ($F = 1.002$, $p = 0.445$), indicating that the differences among pathogens were not dependent upon the environment in which they were grown. Tukey Honest Significant Difference tests revealed a significant difference between the environments LB and KB (-0.31, adjusted $p = 0.015$). Over all environments,

Table 3.1: Main results of in vitro experiments

| Environment | Pathogen | K | Generations | Rate |
|---|---|---|---|---|
| AB | A9 | 0.4758920 | 1.5808789 | 0.3250892 |
| KB | A9 | 0.4252324 | 1.4125913 | 0.3175542 |
| LB | A9 | 0.3099950 | 1.0297810 | 0.3122457 |
| AB | 1448A | 0.3957976 | 1.3148113 | 0.3862845 |
| KB | 1448A | 0.5114521 | 1.6990071 | 0.3438171 |
| LB | 1448A | 0.4347746 | 1.4442901 | 0.4724533 |
| AB | DC3000 | 0.3435305 | 1.1411837 | 0.4250605 |
| KB | DC3000 | 0.4369375 | 1.4514749 | 0.3883002 |
| LB | DC3000 | 0.2795898 | 0.9287771 | 0.4083306 |
| AB | ES4326 | 0.3124702 | 1.0380035 | 0.3601332 |
| KB | ES4326 | 0.4733373 | 1.5723924 | 0.3267870 |
| LB | ES4326 | 0.3658353 | 1.2152784 | 0.3578362 |
| AB | NP29 | 0.5641046 | 1.8739150 | 0.3315395 |
| KB | NP29 | 0.7749478 | 2.5743207 | 0.3588800 |
| LB | NP29 | 0.4196724 | 1.3941214 | 0.3848911 |

NP29 differed significantly from several strains: NP29-A9 (0.41, adjusted $p = 0.033$), NP29-DC3000 (0.58, adjusted $p = 0.001$) and NP29-ES4326 (0.45, adjusted $p = 0.013$). NP29 had the highest median $K$ across all strains in all environments (Fig. 3.2) and furthermore was the only strain whose $K$ significantly varied across environments (ANOVA, F = 9.945, $p = 0.003$). Interestingly, the variability of NP29's performance across environments was not driven by the expected advantage in AB, an environment composed of NP29's host-of-isolation (ANOVA F = 2.254, $p = 0.099$). Instead, NP29 demonstrated a general tendency to outperform other strains across all environments (Fig. 3.2). This is consistent with an advantage of carrying relatively few effectors.

However, the inverse prediction that DC3000, which had the highest effector load, would perform poorly in abiotic environments was not borne out. Though NP29 was the clear best grower of the strains, there was not a clear worst grower; the other four strains grew to the same approximate titer in each environment. The absence of a theoretical cost for the strain with the highest effector load reveals that any metabolic cost of maintaining a high number of effectors is ameliorated in practice. For example, secondary mutations can compensate for the fixation of an allele with a deleterious effect (Levin et al. 2000). Due to the absence of a cost of high effector load under even simplified abiotic conditions, I do not predict a cost of effectors for the *in planta* experiments. Instead, since effectors are likely to provide an *in planta* advantage, I predict that the strain with the highest effector load will be the best grower in general. Contrariwise, NP29 should suffer a disadvantage *in planta* due to low effector load.

### *3.4.3 Growth* in planta

The next set of experiments consisted of growing five pathogens within five hosts over five days. All strains grew on all hosts (Fig. 3.3), reaching $K$ ranging from mean log values of 4.5 – 5.6 and completing 5 – 12 generations. This performance indicates the suitability of these

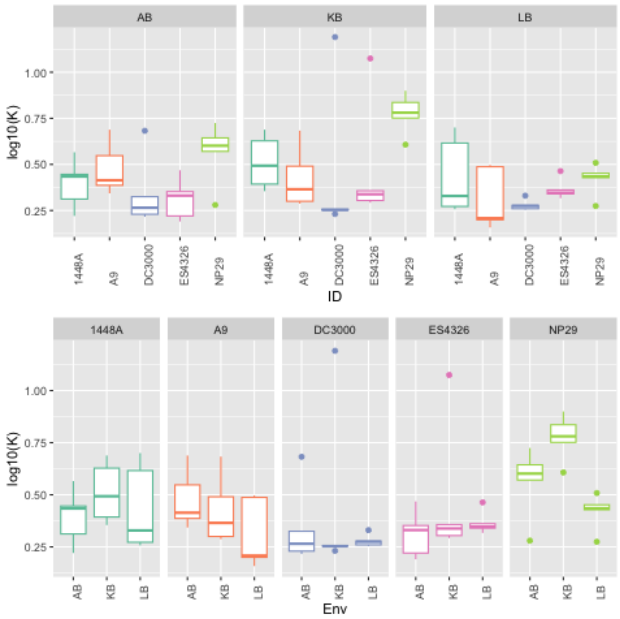



Figure 3.3: Boxplots for *in vitro* experiments. The x-axis represents plant or pathogen as indicated; the y-axis shows $K$. Boxes show the median and lower and upper quartiles of the data with whiskers representing the range. Points reflect outlying data points. Colors are consistent for each pathogen.

hosts for these pathogens. I test whether the best grower is DC3000, and whether NP29, with its relative dearth of effectors, will have a disadvantage relative to the other strains. I also test whether strains are locally adapted by asking whether they grew best on their hosts of isolation, and examine the trajectory of infection in context of local adaptation.

To test the simple prediction that pathogens grew best on their hosts of isolation, I performed a two-way ANCOVA ($K \sim$ plant * pathogen + plate) to explore whether there is a significant interaction between pathogens and the plants in which they are grown. The ANCOVA revealed significant main effects of host (F = 18.938, $p$ = 3.25e-12) and pathogen (F = 4.975, $p = 0.009$), but no effect of plate (F = 0.0, $p = 0.993$) or the interaction between plant and pathogen (F = 0.691, $p = 0.798$). Thus, differences among pathogens and hosts appeared to be independent of which plant a particular pathogen was grown in. Despite this lack of statistical interaction, two pathogens, 1448A and NP29, achieved their highest $K$ on their hosts-of-isolation (Table 3.2). Planned contrasts between $K$ on the host-of-isolation

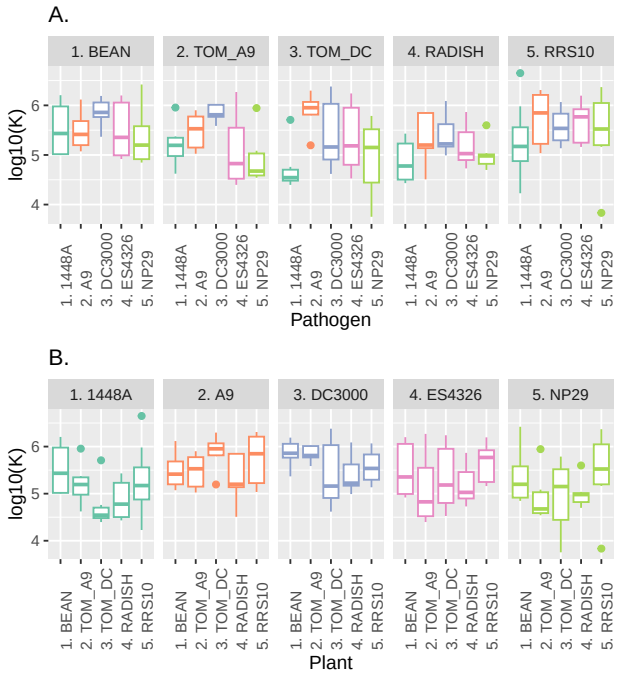



Figure 3.4: Boxplots for *in planta* experiments. The x-axis represents plant or pathogen as indicated; the y-axis shows $K$. Boxes show the median and lower and upper quartiles of the data with whiskers representing the range. Points reflect outlying data points. Colors are consistent for each pathogen.

vs. all other hosts were significant in these two cases (1448A $p = 0.035$; NP29 $p = 0.029$), supporting the idea that these pathogens were locally adapted. For the remaining strains, patterns were haphazard with respect to host: ES4326 reached its highest $K$ on RRS10 and the tomato pathovars reached theirs on bean, but these contrasts were not significant ($p > 0.05$).

To elucidate the 'best' and 'worst' pathogens, I performed post-hoc contrasts on all possible comparisons. I found only two significant contrasts, both for DC000 and indicating an advantage (1448A, $p = 0.021$; NP29, $p = 0.006$). Thus, the strain with the largest effector repertoire was evidently the best performer in this set of experiments. In contrast, I did not find evidence of a relative disadvantage for the strain with the lowest effector repertoire, NP29, which had zero significant contrasts. Rather, the results indicate that, with the exception of DC3000, the pathogens grew to the same approximate $K$ across hosts, such

Table 3.2: Main results of in planta experiments

| Plant | Pathogen | K | Generations | Rate |
|---|---|---|---|---|
| 1. BEAN | 1448A | 5.393123 | 6.921250 | 1.0584440 |
| 2. TOM_A9 | 1448A | 5.058357 | 10.437717 | 1.7342471 |
| 3. TOM_DC | 1448A | 4.533729 | 5.499470 | 0.5907821 |
| 4. RADISH | 1448A | 4.694460 | 5.528116 | 0.4762454 |
| 5. RRS10 | 1448A | 5.099821 | 8.178174 | 1.0815596 |
| 1. BEAN | A9 | 5.349666 | 6.776887 | 1.0418571 |
| 2. TOM_A9 | A9 | 5.339828 | 11.372743 | 1.8863936 |
| 3. TOM_DC | A9 | 5.751461 | 9.544689 | 1.0180566 |
| 4. RADISH | A9 | 5.164183 | 7.088502 | 0.6239570 |
| 5. RRS10 | A9 | 5.609360 | 9.870827 | 1.2895347 |
| 1. BEAN | DC3000 | 5.732722 | 8.049373 | 1.1880619 |
| 2. TOM_A9 | DC3000 | 5.724585 | 12.650876 | 2.0943700 |
| 3. TOM_DC | DC3000 | 5.294024 | 8.025115 | 0.8575522 |
| 4. RADISH | DC3000 | 5.263515 | 7.418474 | 0.6551933 |
| 5. RRS10 | DC3000 | 5.438491 | 9.303212 | 1.2197922 |
| 1. BEAN | ES4326 | 5.362369 | 6.819086 | 1.0467056 |
| 2. TOM_A9 | ES4326 | 4.927412 | 10.002725 | 1.6634658 |
| 3. TOM_DC | ES4326 | 5.190258 | 7.680412 | 0.8211431 |
| 4. RADISH | ES4326 | 5.029417 | 6.640819 | 0.5815778 |
| 5. RRS10 | ES4326 | 5.532722 | 9.616241 | 1.2582539 |
| 1. BEAN | NP29 | 5.221442 | 6.350938 | 0.9929169 |
| 2. TOM_A9 | NP29 | 4.714792 | 9.296418 | 1.5485364 |
| 3. TOM_DC | NP29 | 4.779370 | 6.315471 | 0.6769718 |
| 4. RADISH | NP29 | 4.825215 | 5.962473 | 0.5173632 |
| 5. RRS10 | NP29 | 5.321040 | 8.913047 | 1.1718529 |

that there was not a truly poor performing pathogen or one that failed to grow.

### *3.4.4 Growth* in planta*: infection trajectories*

In another attempt to explore patterns of local adaptation, I classified infection trajectories after reaching high titers as having either maintained the high titer over the course of infection, or as having experienced a subsequent crash in population size. Local adaptation of the pathogen to the host is required to avoid a crash, in that the pathogen is able to subvert or tolerate the host immune response. Unfortunately, one cannot disentangle the possibilities of the pathogen subverting or tolerating the host immune response based on phenotypic data, but one indication is that a slow increase in population size is consistent with the pathogen needing time to subvert the host response. Thus, I noted whether the increase in population size was fast or slow in my analysis.

Growth curves *in planta* tended to follow one of two trajectory types (Fig. 3.4), in contrast to the abiotic experiments which shared one common mode of growth. The first and most common trajectory was one of 'maintenance', such that high titers were reached and then maintained over the course of infection. The absence of a crash in population size suggests the absence of a plant immune system response to shut-down pathogen growth, either because there was no strong response or because the pathogen was able to tolerate the plant's response. Most of these maintenance infections demonstrated quick increases in population size, such that $K$ was reached by 24 hours-post infection with no indication of the host slowing the pathogen's growth. This occurred for 13 infections, including four of five ES4326 infections (the exception being radish, the native host) and four of five pathogens in Brandywine tomato (Tom-A9), excepting A9, the native pathovar. A few infections used a different mode of infection: a slow, steady increase with eventual plateau as $K$ is reached, as is the case for 1448A in its native host, bean. This pattern is similar to the slow increase and maintenance of high titer we see for DC3000 in its native host, Golden Sunrise tomato

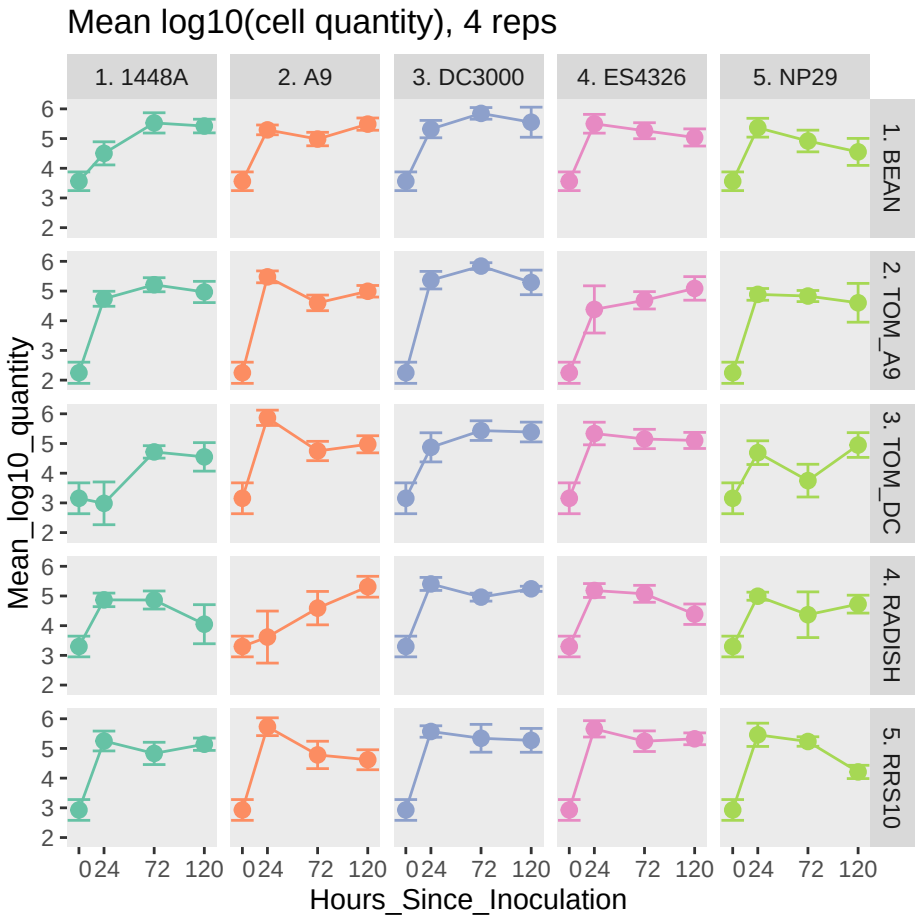



Figure 3.5: Infection trajectories over the average of 4 replicates. The x-axis shows Hours since inoculation and the y-axis shows log pathogen titer. Points show log10 average titer, and error bars show 95% confidence intervals.

(Tom-DC), as well as 1448A in Tom-DC, and A9 in radish. Thus, four infections fit this 'slow-and-steady' pattern, two of which are pathovars within their native host.

The other type of growth trajectory visible across experiments (Fig. 3.5) is one of an increase followed by a 'crash' or significant decrease in population size. These infections were able to reach high titers by 24 hours post-inoculation, but not to maintain such titers over the course of five? days. This trajectory is consistent with a host immune response stopping the pathogen from growing, or with the pathogen overexploiting all available resources. This pattern was evident for seven infections, including ES4326, A9, and NP29 in their native hosts. Each of these declines entailed a significant decrease in titer after 72 to 120 hours

(Student's t-tests, $p < 0.05$). Interestingly, three of five native-pair infections followed this trajectory.

There was one infection that had a singular pattern: a significant decrease followed by a significant increase (NP29 in Golden Sunrise tomato, i.e., Tom-DC). This infection is the only one that follows its trajectory, and so I do not consider it in the following comparison. In summary, I observed 54% maintenance patterns (13/24), 29% grow and crash (7/24), and 17% (4/24) slow and steady (Table 3.2).

## 3.5 Discussion

*P. syringae*, along with many microbial plant pathogens, has a dynamic pangenome that includes a repertoire of molecular 'weapons' that benefit the pathogen in different ecologies. Secreted effectors represent one such weapon that can determine infection success, with patterns of presence/absence that are strain-specific (Dillon et al. 2019A). Another determinate of infection success is adaptation of the pathogen to the plant, which is implied by pathovar classification systems but not often empirically tested. In this study, I compared the effector repertoires and tested the growth of a suite of five pathogens using a series of controlled growth experiments in abiotic vs. host environments. I tested questions including 1) whether number of effectors was a good predictor of performance *in vitro* and *in planta*, 2) whether all pathogens performed better in KB vs. LB, 3) whether NP29 would have an advantage in the AB environment, 4) whether there was evidence of a 'best' and 'worst' pathogen, and 5) whether there was evidence for local adaptation in this experimental system.

The range of effector loads across the five strains (8-36, Fig. 2.6) is not surprising, considering that effector genes are considered to be highly dynamic, with frequent gene loss at play (Ochman and Moran 2001; Morris et al. 2012). I explored the possibility that selection in alternative environments may explain variation in effector repertoires among strains. I expected that NP29, with its reduced effector load, would grow best in abiotic

treatments due to minimal costs of carrying effectors, and this was exactly what I found. At the same time, I expected the minimal effector load of NP29 would have a deleterious effect *in planta*, and it did not. This indicates that the effector load itself is not the determining factor of infection success; that is, even with few effectors, a strain can successfully grow in multiple host species. This is likely the case for most *P. syringae* isolates of *A. thaliana*, where there is selection against many effector families due to host recognition (Laflamme et al. 2020). As for how a strain can be a successful pathogen with few effectors, there is evidence that the *P. syringae* effector repertoire functions redundantly, such that an expanded repertoire does not necessarily mean an equally expanded set of functions (Bundalovic-Torma et al. 2022). This implies the existence of a minimal repertoire, which has been defined empirically in DC3000 (Cunnac et al. 2011). Additionally, some effectors are key determinates of microbe:microbe interactions (Snelders et al. 2020), so it is possible that the effectors that are maintained in NP29 provide a benefit outside of the host environment.

I performed the *in planta* experiments with the aim of elucidating local adaptation in this experimental system, as well as testing my predictions that NP29 would have a relative disadvantage due to encoding few effectors, and that DC3000 with the highest effector load of the pathogens used in this study would be the best *in planta* grower. I did this by infecting pathogens on their own, and on each other's, host-of-isolation. I found that the tomato pathovar DC3000 was the best *in planta* grower as evidenced by it having achieved significantly higher carrying capacity in two of four comparisons (planned contrasts following ANOVA). DC3000 is a race 0 pathovar of tomato, and known for being highly virulent (Dong et al. 1991). This study confirmed that its virulence is not specific to tomato, but rather that this strain appears to be a true generalist, capable of infecting multiple plant species to roughly the same extent. The same set of *in planta* experiments revealed patterns of local adaptation for two of five pathogens. That is, I found that two of five pathogens achieved significantly higher carrying capacity in their hosts-of-isolation than in all other

hosts. These strains were 1448A, isolated from bean, and NP29, isolated from *A. thaliana*. 1448A is an important genetic model of *P. syringae* that is known to be recognized by its host-of-isolation (Arnold et al. 2011). 1448A has been shown to have a generalist, but not ubiquitous host range (Morris et al. 2019). My results are not inconsistent with previous findings, but reveal that 1448A has a strong preference for its host-of-isolation despite being able to grow across a suite of hosts. Interestingly, NP29 has been classified as 'maladapted' to *A. thaliana* based on its induction of host recognition (Kniskern et al. 2011). Here, I found that NP29 still performed best in *A. thaliana* relative to all other hosts, despite any response from the plant. That NP29 was able to outperform all other pathogens on this host reveals a benefit of NP29's lack of effectors in an *A. thaliana* environment, consistent with the findings of others that *A. thaliana* recognition selects against many effectors (Laflamme et al. 2020).

Finally, I considered the trajectory of infection as it pertains to local adaptation. I found maintenance trajectories for the majority of infections, including 1448A in its native host, which demonstrated a slow increase, consistent with the pathogen needing time to subvert the host immune response. This was not the trajectory experienced by NP29 in its native host; instead, NP29 experienced a crash in population size in RRS10. In fact, three of five native-pairings exhibited this crash pattern, making it tempting to speculate that hosts may have a relative advantage in their ability to shut-down infections with their native pathogen. Future studies may focus on the genetics underlying the patterns of growth seen here, revealing whether there are specific genes responsible for the patterns of local adaptation observed.

One caveat to my approach is that I focused exclusively on carrying capacity as a proxy for fitness, when there is also rate of growth to consider. I feel confident in this choice as repeating the analysis with rate in place of $K$ generally made the results less informative. In the *in vitro* experiments for example, repeating the ANCOVA (rate $\sim$ environment *

pathogen + plate) with rate in place of $K$ revealed a significant main effect of pathogen (F = 6.591 , $p = 0.000209$) and plate (F = 4.162, $p = 0.005134$), but no effect of environment or the interaction between environment and pathogen. This differs from the original ANCOVA with $K$, which uncovered significant effects of plant and pathogen, indicating that $K$ was a more informative choice than rate. For the *in planta* experiments, ANCOVA with rate (rate $\sim$ plant * pathogen + plate) yielded results that were entirely non-significant, indicating that pathogens grew at approximately the same rate across experiments, again indicating that $K$ was a more appropriate choice.

In conclusion, I find that signatures of local adaptation are present, but not universal, in this experimental system. Our findings indicate a caution against assuming local adaptation based on host-of-adaptation. They also reveal that predicting growth based on effector load is not always straightforward. For NP29, maintaining only a small effector load seems to be an effective strategy for growth in abiotic and host environments. However, the strain with the largest effector load did not suffer a disadvantage because of it in these experiments. Instead, I found that the pathogen with the highest effector load, DC3000, was the best pathogen across plants. Thus, successful prediction of infection outcomes likely relies on a deep understanding of the molecular underpinnings of the virulence factors per a pathogen, plus the ecology of the interaction itself.

# CHAPTER 4

# TNSEQ WITH FIVE HOST-PATHOGEN PAIRS

## 4.1 Abstract

The genes involved in disease spread and emergence are of central interest to the field of evolutionary biology. TnSeq is one method by which genotype can be linked directly to phenotype and even to selective coefficient in a given environment. Much has been learned from TnSeq applied to microbial pathogens; for example, the essential genome has been elucidated for many pathogens in common lab cultivation media. There have been a few important attempts to apply TnSeq to a host-pathogen system *in planta*; these have revealed the suggestion of a host-specific accessory genome in a single of *Pseudomonas syringae*. I designed a five-by-five factorial infection experiment to generate a TnSeq dataset that would have allowed me to learn whether, for each of five pathogens, there exists a host-specific accessory genome for each of give hosts. However, I was ultimately unable to generate enough sequencing material for statistical hypothesis testing. I present growth assays on my Tn-libraries and as much explication as possible in hopes that someone may benefit from it in future.

## 4.2 Introduction

*Pseudomonas syringae* is an agriculturally important pathogen with a host range that includes many crop plants (Preston et al. 2000; Lindeberg et al. 2009; Baltrus et al. 2017). *P. syringae* has a global distribution and can cause devastating pandemics; for example, an outbreak of disease on kiwifruit devastated production across multiple countries for years (McCann et al. 2017). Thus, there is a great deal of interest in identifying the genetic basis of disease in this pathogen. Like many bacterial phytopathogens, *P. syringae* has a large pangenome that is still increasing in size as new strains are added (Dillon et al. 2019).

There is a wealth of rare and even singleton genetic variants (Karasov et al. 2017), which are unlikely to be strictly neutral in fitness effect (Wolf et al. 2016). The observation of rare genetic variants in high numbers makes it tempting to speculate that such variants are maintained by selection on various host plants.

One way to link molecular pattern to function is through forward genetic screens, which involve knocking-out the genes in a single genetic backbone in order to observe the fitness effects of the knock-out relative to the wildtype gene (Shuman and Silhavy 2003). More specifically, TnSeq is a forward genetic screen that uses sequencing to obtain data comparing the relative allele frequencies of knock-outs grown in various environmental conditions (Chao et al. 2016). There are numerous examples of TnSeq applied to bacteria grown *in vitro* (Coe et al. 2019; Langridge et al. 2019; Hentchel et al. 2018; Higgins et al. 2017; Lee et al. 2015; Manoharan et al. 2015; Gallagher et al. 2010). For example, essential genomes in common lab media have been identified for some well-studied pathogenic taxa, including *Burkholderia cenocepacia* (Higgins et al. 2017), *Pantoea stewartia* (Duong et al. 2018), and *Pseudomonas aeruginosa* (Poulsen et al. 2019).

Some studies have focused on *P. syringae*, including one on the common cultivation environment King's B (KB) for strain 1448A (Manoharan et al. 2015) and another on KB vs. *Arabidopsis thaliana* for strain ES4326 (Schreiber et al. 2012). There are two additional key studies by Helmann et al. (2019 and 2020) that considered a Tn-library of *P. syringae* isolate B728A in common bean, lima bean, and pepper. In addition to identifying genes specific to each host, an interesting result from their study was that toxin production was beneficial in a host-specific pattern (less beneficial in lima bean than in common bean or pepper). This was not expected and demonstrates the ability of TnSeq to reveal differential patterns of selection across hosts. Other than this example, we know little about the specificity of genes required for growth on different hosts, and it remains an open question to what extent pathogens use specific genes to infect their hosts-of-isolation.

Herein I describe my attempts to use TnSeq to uncover the molecular mechanisms that cause pathogenicity in *P. syringae*. Unfortunately, I failed to generate significant results, for reasons I describe in detail below. With this approach, I attempted to generate a selective coefficient for each accessory gene, which would have enabled me to infer essentiality of those genes which cannot withstand transposon-insertion. For an initial set of experiments, I chose environments that represent typical laboratory growth conditions for *P. syringae*. Next, I paired the pathogens with their hosts-of-isolation (or as close as possible), in order to identify host-specific accessory genes, if any are present in this system.

## 4.3 Methods

Strain and environment selection In this chapter I repeat a series of growth assays that include growth in three abiotic and growth in five plant host environments (Table 4.1). Instead of using the wildtype pathogen as in Chapter 2, I performed these assays on transposon-mutant libraries of the pathogens (generation of these libraries is described in detail below). Details on strain and environment selection and composition are described in detail in Chapter 2; in brief, I chose virulent pathogens isolated from a range of hosts, including crop plants (bean, radish, and tomato) and the natural, weedy host *A. thaliana*. I chose two abiotic environments that the pathogens are likely to encounter during cultivation, King's Broth (KB) and Lysogeny Broth (LB), plus I used a third media that was generated in-house by maceration and filtration of *A. thaliana* plants. For *in planta* assays, I chose to assay each transposon-mutant-library in each pathogen's host-of-isolation (or a phenotypically similar cultivar that was grown in the same location at the time of isolation). Thus, I first grew libraries in abiotic enviroments with the goal of identifying phenotypic patterns of growth and the genes required for growth of each strain in each environment. Then, I grew libraries in their own, and in each other's host-of-isolation with the goal of asking whether each pathogen contains genes that are specifically beneficial in the host-of-isolation vs. all other

hosts.

### *4.3.1 Tn-library construction*

I generated transposon-mutant-libraries of each of five *P. syringae* donors via tri-parental mating with *Escherichia coli* donor SM10λpIT2, containing the transposon used for mutagenesis, and helper RK600, containing the conjugative plasmid, both kindly provided by Dr. Colin Manoil. Transposon Tn8 (ISlacZ/hah-tc) mutagenizes through Cre-Lox recombination, randomly inserting the "hah-tc" sequence containing a tetracycline-resistance cassette into the recipient genome (Jacobs et al. 2003).

I used the following antibiotics as required: for SM10λ/pIT2, ampicillin at 100 mg/mL; for RK600, chloramphenicol at 25 mg/mL, nitrofurantoin for Pseudomonads except for ES4326, at 100 mg/mL, and streptomycin for ES4326 at 30 mg/mL. In all cases, cultures were incubated at 28°C and liquid cultures were grown shaking at 120 rpm. Starting with single colonies grown on plain KB or LB, I inoculated 5 mL liquid cultures with the appropriate antibiotics and grew them overnight, shaking at 28°C for 12 hours. The following morning, I diluted cultures 1/10 into fresh media and antibiotics and grew for 4-6 more hours to mid-stationary phase. I then combined equal quantities of each parent, pelleted and washed the mixture twice in liquid KB, resuspended it in 100 μL liquid LB, then spotted the suspension onto plain KB agar. Matings were incubated at 28°C for two days, at which point I scraped them into sterile buffer, then pelleted and washed the cells twice in buffer to remove residual agar. I then spread-plated the transformants onto KB agar supplemented with two selective antibiotics (tetracycline to select for the transposon and nitrofurantoin or streptomycin to select for the Pseudomonad). These were incubated for 2 days or until individual colonies were visible, with a minimum of 50,000 transformants per strain. I scraped colonies into fresh liquid KB and grew for four hours to separate transformants from residual agar, then I normalized the libraries to OD600=1.4 in plain KB before diluting to a final

Table 4.1: Host-pathogen pairs

| Strain | Source | Length | CDS | Effectors | GC |
|---|---|---|---|---|---|
| A9 | tomato cv. Brandwine | 6295740 | 5744 | 23 | 58.8 |
| DC3000 | tomato cv. Guernsey | 6538260 | 5858 | 36 | 58.3 |
| ES4326 | radish cv. Cherry Belle | 6606571 | 6054 | 29 | 58.3 |
| 1448A | bean cv. African Premiere | 6112448 | 5573 | 26 | 57.9 |
| NP29 | A. thaliana ecotype RRS10 | 5992538 | 5146 | 8 | 59.2 |

concentration of 15% glycerol and storing in 2 mL aliquots at -80°C.

### *4.3.2* In vitro *growth assays*

To estimate growth of the Tn-libraries in abiotic environments, I measured optical density over 20 hours of incubation. To allow acclimation before this assay, I diluted flash-thawed library aliquots 1/10 into fresh KB and grew for three hours. I chose three hours to allow the library sufficient time to revive and expand but to preclude competition due to nutrient-limitation. I diluted these pre-grown cultures 1/200 into 199 $\mu$L of media in sterile, optically clear 96-well plates. I then incubated plates at 28°C in a Tecan plate reader set to shake at an amplitude of 2 mm for 10 seconds, and to measure absorbance at 600 nm with 25 flashes, every 10 minutes for 20 hours. I repeated this assay for a total of six replicates. Each well was normalized to its own initial optical density, removing the contribution of the media to optical density estimates. To generate material for sequencing to elucidate the genes essential for growth *in vitro*, I grew the libraries in three liquid media in 50 mL. I inoculated 200 $\mu$L of flash-thawed freezer culture into 20 mL of either LB, KB, or AB, and grew cultures for four hours at 28°C and shaking at 120 rpm before DNA extraction.

### *4.3.3* In planta *growth assay*

Plants were grown using the same conditions described in Chapter 2; briefly, all five hosts and five Tn-libraries were included simultaneously, in each of four fully-randomized blocks that included one replicate of each host-library combination. Plants were synchronized to the same approximate developmental stage; I infected 21-day old (*Arabidopsis* and tomato) or 9-day old (bean and radish) seedlings. To prepare Tn-libraries for infection, I inoculated 10 mL of flash-thawed freezer culture into 200 mL KB and grew for 4-6 hours to, at which point I pelleted and washed the cells with sterile buffer, before resuspending in buffer to an OD=1.1. Plants were sprayed until dripping wet and allowed to dry for 24 hours before

being randomized in a checkerboard pattern to ensure plants did not touch. I harvested plants on day six of infection by taking two sets of 18 (six mm) hole punches from surface-sterilized leaves (dipped in 70% then sterile water for 30 s each). One set of holepunches was used to obtain titers while the other was used to obtain sequencing material. Hole punches were reserved in 50% glycerol at -20°C until processing. To obtain titers, hole punches were macerated, diluted in sterile buffer, and spread-plated for colony enumeration. To obtain sequencing material, hole punches were macerated in buffer, spread-plated, and harvested (scraped into sterile buffer) after two days of incubation. Thus, my TnSeq input material was plant-selected bacteria. DNA extraction was performed using the Qiagen DNeasy Blood and Tissue kit (catalog 69504) with a final elution of 60 $\mu$L. DNA was kept in Milli-Q water at 4°C to avoid degradation from salt and freeze/thaw cycles, and quantitated with Nanodrop or Picogreen prior to library prep.

### *4.3.4 Statistical analyses*

Statistical analyses were performed in R v. 4.2.2 (R Core Team 2021) using the growthcurver package (Sprouffske and Wagner 2016). Growthcurver is an R package that fits growth curve data to a standard form of the logistic equation with the following parameters: the intrinsic growth rate of the population, $r$, the initial population size, $N0$, and the carrying capacity, $K$. The intrinsic growth rate of the population, $r$, is the growth rate that would occur if there were no restrictions imposed by population size. Carrying capacity $K$ is defined as the maximum optical density reached over the course of the experiment. The logistic equation describes the population size $Nt$ at time $t$ as: $Nt = K/1 + ((K - N0)/N0) * e - rt$ . To calculate the number of generations, I used the standard equation (OD600final – OD600initial)/ln(2), where OD600final is the carrying capacity $K$. Time was transformed from seconds to hours for all calculations. To analyze the *in vitro* growth experiments, I used an ANCOVA ($K \sim$ environment * library + plate), followed by Tukey's post hoc tests

on all possible comparisons. For the *in planta* experiments, I used an ANOVA ($K \sim$ plant * library) followed by planned contrasts of growth on the host-of-isolation vs. all other hosts.

### *4.3.5 Efforts related to sequencing*

Before large-scale sequencing, I performed some quality control qPCR reactions to check the efficacy of primers designed to amplify the transposon sequence. This was primarily due to the fact that I was planning to try amplifying transformants directly from plant material, meaning that the primers needed to amplify transformants and not plant material. The initial TnSeq primer G-43899 (inherited from TK) amplified the reverse complement of bases 53 through 33 of Tn8. This happened to contain a partial sequence match to the *A. thaliana* genome, so I selected a few alternate primer sequences to verify (T8-3: reverse complement of bases 46 to 13; T8-4: reverse complement of bases 193 to 158). I Sanger sequenced amplicons generated with each of the experimental primers to confirm orientation of the transposon was as expected. I also performed qPCR to compare the percent of on-target hits per each experimental primer. All experimental primers, designed to exclude plant material, performed better than the initial primer; the example shown below (Fig. 4.1) compares accuracy of T8-4 to G-43899 (TK) in a single sample. Thus, I used experimentally verified transposon sequences in the following primer designs.

I performed all sequencing preparations on ice and kept working dilutions of DNA in Milli-Q water at 4°C to avoid degradation from salt and freeze/thaw cycles. I used Picogreen to quantitate and a Nanodrop to check purity of DNA, and homemade SPRI beads for purification and size selection. I obtained all primers by custom order from IDT (Table 4.2).

### *4.3.6 Sequencing scheme 1*

An initial sequencing scheme (2016) was inherited from Talia Karasov. This involved prep with a Nextera kit (Illumina) to "tagment", or fragment and ligate adapters to the fragment

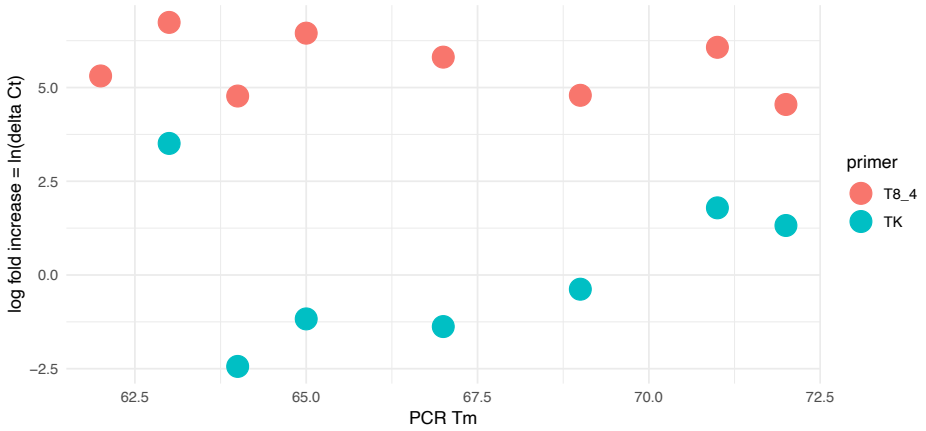



Figure 4.1: Log fold increase in on/off-target PCR product. T8-4 generally outperforms the initial primer.

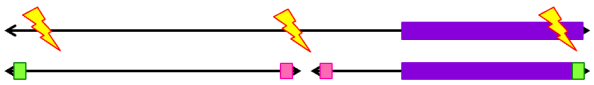


Figure 4.2: Tagmentation in this system. Double-stranded breaks are indicated by lightening bolts. The transposon is shown in purple; the standard Illumina adapters are indicated in pink and green. A 'successful' fragment of interest contains the transposon sequence.

ends in one reaction (Fig. 4.2). After tagmentation, there was a two-step, nested PCR to enrich for the transposon while adding sequencing primers and indexed-adapters.

Fig. 4.3 shows the PCR protocol used after tagmentation to generate sequence-ready amplicons. The first PCR used a standard N700 primer on one end (Illumina), and T8-4, a primer that amplified the 5' end of the Tn8-transposon (Table 4.2). This PCR was followed by a SPRI bead reaction to purify and remove residual adapters. Finally, a second PCR used a 'general' N700 primer (G-28018) that amplifies any N700 adapter on one end, and a modified P500 primer that contained part of the transposon sequence on the other end (Table 4.2). This generated sequence ready amplicons with indexed adapters flanking a region of *P. syringae* donor DNA and the junction with the transposon insertion.

I prepared approximately 60 samples in triplicate for sequencing on a Hiseq 4000. Sequencing with this scheme failed completely and lead to zero data being generated. This was unexpected as library concentration and Bioanalyzer traces were in accordance with typi-

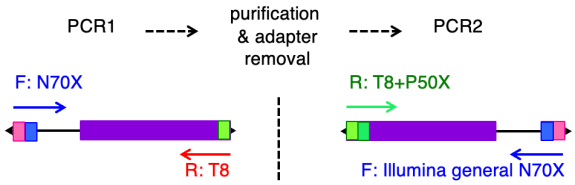



Figure 4.3: Sequencing scheme 1 included nested PCRs. The transposon is shown in purple; the standard Illumina adapters are indicated in pink, green, and blue.

cally successful library preparations. The sequencing was done at the UChicago core facility, where it seems the standard step of performing qPCR with Illumina sequencing primers was not performed. This was unexpected and left me without confirmation that my adapters were in the correct orientation. With the information I had, I reasoned something must be fundamentally wrong with the sequencing scheme. I consulted with my lab's sequencing expert, Hannah Whitehurst, who theorized that an intermediate product containing the transposon sequence could be interfering with priming in the second PCR. This is theoretically possible, though not directly testable, and it was our best guess for what mechanism led to a failed sequencing run. Thus, I decided to abandon this sequencing scheme and come up with an alternative method.

### *4.3.7 Sequencing scheme 2*

The second sequencing scheme was the most successful but still did not generate enough data for my purposes. The design is essentially the inverse of the first, in that it attached the transposon sequence to the opposite adapter, but I also tried to reduce complexity and combine all steps into one PCR (Fig. 4.4). Thus, the second scheme included tagmentation (same exact protocol as the first), but used different PCR primers to generate a sequence-ready amplicons in one step (Table 4.2). I prepared six samples for an initial pilot experiment that generated good-looking quality control metrics, including concentration and Bioanalyzer results. I submitted them for sequencing on a Miseq with a 1.2 Gb kit (v2 Micro). This sequencing run generated high quality reads, but only approximately 50% of them contained

Table 4.2: Primers used in this study

| Scheme | Primer Name | Primer Sequence |
|---|---|---|
| Scheme 1 | | |
| PCR1 FOR | N701 | CAAGCAGAAGACGGCATACGA TCGCCTTA GTCTCGTGGGCTCGG |
| PCR 1 REV | T8_4 | GGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCA |
| PCR2 FOR | G-28018 | CAAGCAGAAGACGGCATACGA |
| PCR2 REV | S501-Tn-Amend | AATGATACGGCGACCACCGAGATCTACAC TAGATCGC TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG CAGGACGCTACTTGTGTATAAGA |
| Scheme 2 | | |
| PCR FOR | N701-Tn-Amend | CAAGCAGAAGACGGCATACGAGAT TCGCCTTA GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG CGTTAACGATAACTTCGTATAGCAT |
| PCR REV | S501 | AATGATACGGCGACCACCGAGATCTACAC TAGATCGC TCGTCGGCAGCGTC |
| Scheme 3 | | |
| PCR1 FOR | RS_SemiDeg-1 | GAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNTAGATC |
| PCR 1 REV | T8_3 | AACGGGAAAGGTTCCGTCCAGGACGCTACTTGTG |
| PCR2 FOR | RS_PCR2-T8_3-i64 | AATGATACGGCGACCACCGAGATCTACAC TCTTTCCC TACACGACGCTCTTCCGATCT GGTTCGAT AGGACGCTACTTGTGTATAAGAGTCAG |
| PCR2 REV | PCR2-CSD-1 | CAAGCAGAAGACGGCATACGAGAT AACGATGG GTGACTGGAGTTCAGACGTGTGCTCTT |
| PCR2 REV | Corrected by Illumina | CAAGCAGAAGACGGCATACGAGAT AAGCGCAA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |

the transposon sequence, which should have been in virtually every read. Additionally, there were a high proportion of duplicates, which I addressed by reducing number of cycles from 30 to 18. With this adjustment, I proceeded to prepare and sequence approximately 100 samples in triplicate for sequencing on a Hiseq 4000. Though these preparations also yielded good-looking quality control metrics, unfortunately, again, about 50% of the reads contained the transposon sequence, when it should have been found in all reads. Additionally, the quality of the runs was very poor as indicated by FastQC and MultiQC, with most of the sequences being duplicates even though I had adjusted PCR cycle number. There was also a large amount of adapter contamination, which is my fault as I failed to confirm an additional adapter removal step with the sequencing facility after initial quality control was returned.

Thankfully, Hanna Maerkle realized that the reads not mapping to the transposon were actually mapping to the pIT2 transposon vector. That is, in sequences that did contain the 5' end of the transposon as expected, often in place of *P. syringae* donor DNA there was pIT2 vector DNA. Thus, it seemed that with some frequency, the vector had recombined into the transformant's genetic backbone after transformation. I confirmed that this does occur (personal communication with Colin Manoil), though it appears the frequency of this occurrence is unknown. It seems the majority of our reads were comprised of vector contamination instead of donor gDNA, severely limiting my ability to generate statistically significant TnSeq results. I was able to generate some non-significant TnSeq results with the sequencing data I did generate, which I present in the Results section.

### *4.3.8 Sequencing scheme 3*

In a final attempt to generate data, I thought of a new approach that would allow me to sequence around vector contamination while eliminating the use of tagmentation entirely. I thought this might be beneficial as the tagmentation relies on Tn5 transposition; Tn5 is closely related to Tn8 and I reasoned this could lead to unforeseen interactions. The original

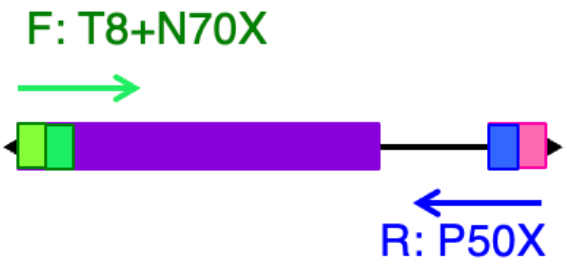



Figure 4.4: The second sequencing scheme involved a single PCR. The transposon is shown in purple; the standard Illumina adapters are indicated in pink, green, and blue.

TnSeq method used semi-degenerate primers (Manoil 2000) with complementary adapters to comprise a two-step, nested PCR to generate sequence-ready fragments. This approach had been modernized with the addition of TruSeq adapters by Anzai et al. (2017). The point of using semi- instead of fully-degenerate primers is that there are 'blocker' sequences at one end of the degenerate bases that limit binding to a reasonable number of genomic locations (instead of literally anywhere, which can lead to the amplification of extremely short sequences). I reasoned that I could use this same approach, but with modified semi-degenerate primers that would selectively exclude pIT2 (Fig. 4.5). I replaced the 'blocker' sequences GATAT and ACGC with TAGATC and AACGCTA, which are not found in pIT2 so would not amplify pIT2 if present. Thus, the scheme involved first amplifying out of the transposon on one end vs. a semi-random location in the donor genome on the other end. Then, a second PCR added indexed-adapters and sequencing primers. Critically, the final amplicon contains indexed-TruSeq adapters flanking the first 27 bases at the 5' end of the transposon, plus a region of the donor genome.

I prepared three samples for sequencing on a Novaseq 6000 using this scheme, and while the sequencing failed to generate data, I have every reason to think that the protocol worked to generate amplicons as expected. This is because 1) amplicons produced a smear via gel electrophoresis, as expected, 2) amplicons produced Bioanalyzer traces consistent with the distribution seen on the gel, and 3) qPCR results produced approximately the same

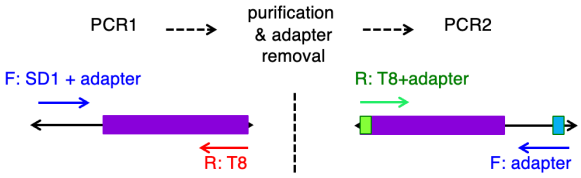



Figure 4.5: The transposon is shown in purple; the Illumina Truseq adapters are indicated in green and blue.

concentrations as the Bioanalyzer. Based on these results, I reasoned it must be an incompatibility of the sequencing adapters with the sequencer, even though I had used published sequences. I contacted Illumina customer support and worked with them to discover that this sequencing scheme is missing part of Read 2 sequencing primer, which is required to appropriately anneal amplicons to the sequencer. Note that the corrected primer sequence from Illumina, indicating the missing bases in bold, is listed in Table 4.2. Either there was an error in the published sequence (Anzai et al. 2017) or their method relies on out-of-date Illumina sequences. Thus, the method of using modified semi-degenerate primers for sequencing preparation appears to be a solid one, but an updated adapter scheme is required for successful sequencing.

### *4.3.9 TnSeq analysis with TRANSIT*

Hanna Maerkle did the majority of computational hacking of pipelines described in this section, while Rebecca Satterwhite completed the sequencing and analysis. We analyzed our TnSeq dataset by adjusting the publicly available TPP and TRANSIT pipelines (DeJesus et al. 2015) to account for the specifics of our Tn8 amplification and sequencing scheme. TPP cleans, maps, and tallies raw TnSeq reads to a reference genome. TRANSIT also requires standard reference genome annotations be converted to the '.prot-table' file format, which links genomic location to protein annotation. For statistical analyses of our Tn8 data, we selected the Tn5Gaps statistical method (DeJesus et al. 2015), which uses the distribution of 'gaps', i.e., sequence lacking Tn-insertions, across the whole genome (including introns), to

predict the expected maximum length of such gaps, and to calculate a p-value per each actual gap. Genes are then classified as essential based on the p-value of the largest overlapping gap; that is, the probability under a Gumbel distribution that the protein product is disrupted.

We modified TPP to allow tallying of our Tn8 reads to reference genomes. Modifications include 1) we added the Tn8 primer sequence to the program, 2) adjusted the coded location of the transposon to match the format of our data, and 3) omitted a superfluous step to check for unused barcodes. We used the standard settings to run TPP, and then combined all .wig files for a sample in a particular environment using the TRANSIT script convertToCombinedWig with normalization method set to Trimmed Total Reads (TTR). TTR normalizes by total read-count per isolate per environment and trims the top and bottom 5% of read-counts. This scales each sample to have the same mean over all read-counts, and is recommended for resampling analyses.

## 4.4 Results

The first set of experiments I performed with my Tn-libraries was to ask how they grow in standard laboratory environments (Fig 6). I asked whether there are significant differences in $K$ across libraries and environments with an ANCOVA ($K \sim$ environment * pathogen + plate). I found significant effects of environment (F = 9.295, $p = 0.000418$), plate (F = 5.082, $p = 0.001823$), and the interaction between library and environment (F = 2.428, $p = 0.028385$). Thus, the performance of each library depends, to at least some extent, on what environment it is in. To determine the drivers of these differences, I performed a Tukey's post hoc test of all possible comparisons, and found two significant environment contrasts indicating that $K$ in KB was higher than $K$ in AB ($p = 0.0005$) and in LB ($p = 0.008$). From this I learned that the libraries had significant growth advantages in KB relative to the other media. This is consistent with what I found for the pathogens themselves (Chapter 1). However, there were no significant library contrasts, consistent with the results

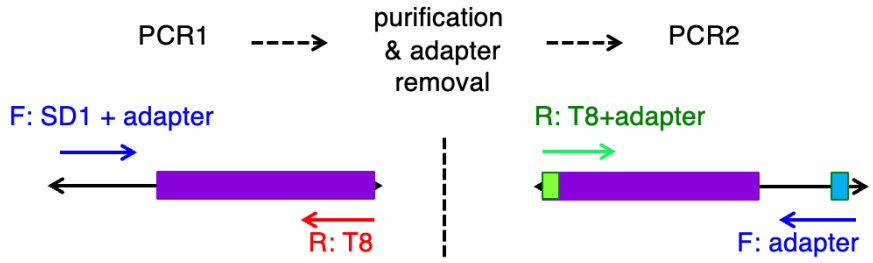



Figure 4.6: Median and 95% CIs of carrying capacity in abiotic environments

of the ANCOVA. Thus, the libraries had a significant advantage in KB relative to the other environments, but there was not a best or worst grower. This is somewhat surprising considering that NP29 had a significant advantage in prior assays (Chapter 1). However, in this case I am estimating growth of a pooled mutant library where allele frequencies are likely to change over the course of the assay.

Having assessed phenotypic patterns of growth, I next asked to what extent the genes essential for growth in each environment are exclusive using the sequencing I obtained with the second scheme. Unfortunately, and for reasons unknown, results from the A9-Tn-library were exceptionally low in number and quality, so I excluded them from these examples. The results produced a wide range of predicted numbers of significant (essential) genes (Table 4.2). In order to explore these results, I considered the number of overlapping essential genes in the AB vs. KB environment (Table 4.3). I confirmed that at least a few of the genes identified as essential are housekeeping genes, such as the RpoA, RpoB, and RpoC. Similarly, I confirmed that at least a few of the genes identified as non-essential are known to be non-essential, including virulence sensor protein BvgS and multidrug efflux pump subunit AcrA.

The second set of experiments I performed with my Tn-libraries was to grow them in a series of five plant hosts (Fig. 4.6). The ultimate goal of this experiment was to ask whether there is a host-specific accessory genome; however, I am limited by poor data generation from being able to answer this question. First, I ask what the phenotypic patterns of growth

are across the experiment; specifically, I ask whether there are significant differences in $K$ across libraries grown in their own, and in each other's, hosts-of-isolation ($K \sim$ plant * library). I found significant effects of plant (F = 40.19, $p < 2e-16$), library (F = 16.25, $p = 4.99e-12$), and the interaction between plant and library (F = 12.91, $p <$ 2e-16). This indicates that library growth depended on which plant the library was grown in, and demonstrates that the hosts represented distinct selective environments for the libraries. To determine whether the interaction between plant and library was driven by libraries growing most on their host-of-isolation, I performed planned contrasts of growth on the native host vs. all other hosts. I found that in only one of five cases (DC3000-Tom-DC, $p = 0.0453$), growth on the host-of-isolation was significantly greater than on all other hosts. In two other cases, growth on the host-of-isolation was actually significantly worse than on all other hosts (ES4236-Radish, $p = 0.0047$ and 1448A-Bean, $p < 0.0001$).

## 4.5 Discussion

Elucidating the molecular basis of disease has long been a central goal in the field of evolutionary biology (Suárez-Díaz 2019; Croll and Laine 2016; Rosenberg and Queitsch 2014; Monteil et al. 2013; Lederberg 1997; Pauling 1964). An often-used approach are forward genetic screens, for example TnSeq, which enables the fitness effect of any and all genes to be estimated after growth under controlled selective conditions. With this chapter I attempted to ask whether there exists a host-specific accessory genome in *P. syringae*. Though I was unable to answer this question specifically, I was able to generate and assay five Tn-libraries for growth phenotypes in multiple environments, revealing patterns largely different from those obtained with assaying the wildtype pathogen under the same conditions.

Consistent with assays of the pathogen *in vitro* from Chapter 2, I found a significant advantage for the libraries when grown in KB vs. LB and AB. This is not surprising as KB was designed for the cultivation of Pseudomonads. However, I did not find the same relative

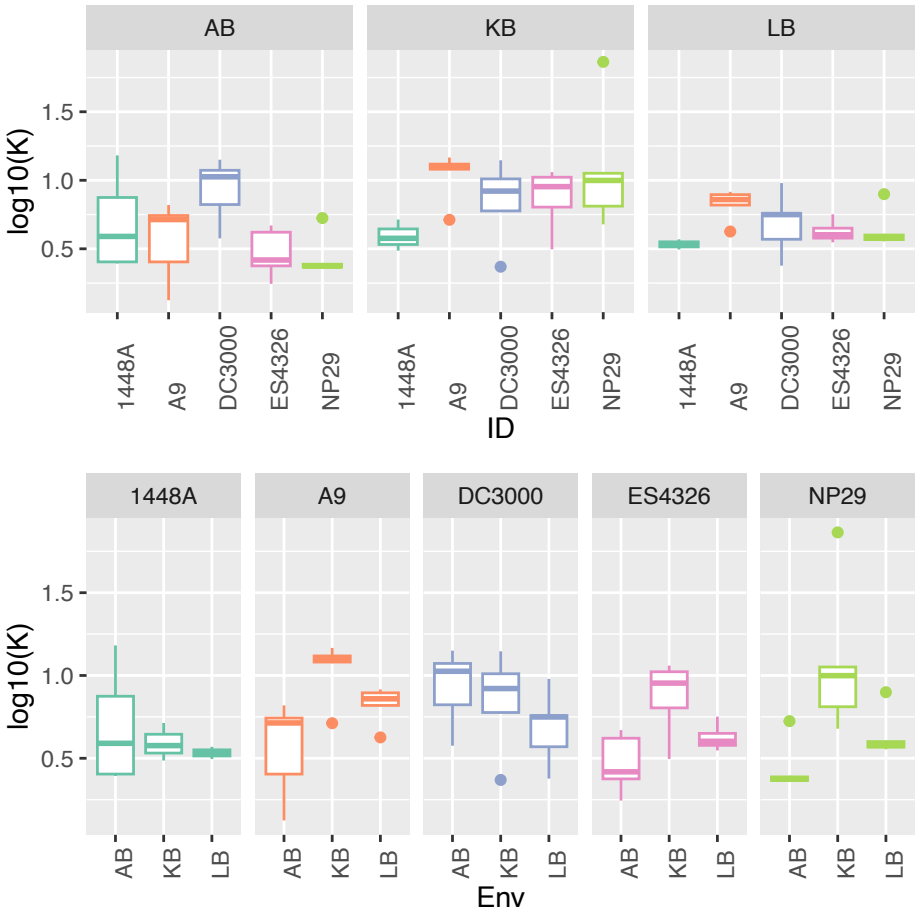



Figure 4.7: Median and 95% CIs of carrying capacity in abiotic environments

Table 4.3: Overlap in essential functions in the AB (A) and KB (B) environments.

| strain | environment | replicates | insertion density | significant genes | mean run length |
|---|---|---|---|---|---|
| ES4326 | KB | 3 | 0.005 | 214 | 253.6 |
| ES4326 | AB | 3 | 0.001 | 60 | 1477.5 |
| 1448A | AB | 3 | 0.001 | 22 | 1355.0 |
| 1448A | KB | 3 | 0.001 | 76 | 1080.9 |
| 1448A | LB | 3 | 0.001 | 105 | 1012.1 |
| DC3000 | KB | 1 | 0.000 | 58 | 4845.3 |
| DC3000 | AB | 3 | 0.001 | 656 | 1573.8 |
| NP29 | KB | 1 | 0.002 | 58 | 528.1 |
| NP29 | AB | 3 | 0.002 | 68 | 488.1 |
| NP29 | LB | 1 | 0.002 | 57 | 513.5 |

advantage for the NP29-library as for the wildtype pathogen (Chapter 2). Thus, library performance is likely not predictable based on the performance of the pathogen. This is also true for the *in planta* experiments, where only one of the libraries grew most on its host-of-isolation (DC3000). For the pathogen *in planta* experiments, not only were two pathogens found to grow most on their hosts-of-isolation, but neither were DC3000 (Chapter 2).

Though I was unable to produce significant TnSeq results, we have learned a lot from the previous work of others. For example, Duong et al. (2018) found the genes conditionally essential for growth of *Pantoea stewartii* in LB and corn xylem, finding approximately 500 genes that were beneficial for growth in corn xylem relative to LB. Another study by Helmann et al. (2019) found that for *P. syringae* strain B728A, 60 genes had significant competitive fitness estimates in the common bean apoplast relative to KB. It will be interesting to compare the results of the B728A experiment to the results of *in planta* TnSeq on other *P. syringae* strains, in particular with differing effector repertoires. For example, I would be curious to compare the fitness effects of genes used by a strain with a minimal effector repertoire vs. one a maximal effectors repertoire. B728A has at least 22 secreted effectors (Vinatzer et al. 2006), vs. a strain like NP29 that has only 6; thus, I would expect that such strains use different molecular strategies to infect.

If I were to repeat these experiments again, I would attempt to use the semi-degenerate primer scheme as I think it worked well and was the most straight-forward approach. The central change I would make is to use a barcoded Tn-library. Even at the cost of the novelty of generating a library in-house (that is, even limited to the existing barcoded libraries available), I think the utility of having each transposon insertion be barcoded such that unambiguous tracking and comparison of read enrichment is available is key to the strength of the TnSeq approach. Additionally, barcoded libraries provide the utility of having individual knock-outs in-hand to allow performance of fine-tuned experimentation. Future work could be done with the *P. syringae* library generated by Helmann et al. (2019) grown on additional

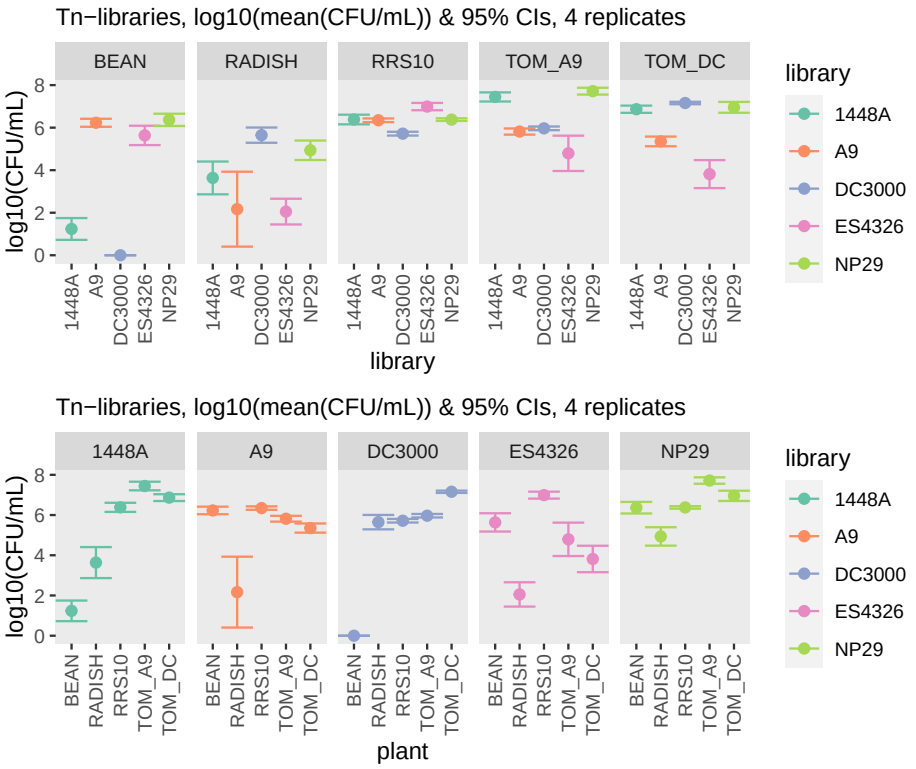



Figure 4.8: Median and 95% CIs of carrying capacity in planta

plants to expand the search for a host-specific accessory genome.

TnSeq is a powerful tool that can link phenotype and genotype directly with selective coefficient (van Opijnen et al. 2009). I attempted an ambitious TnSeq experiment to generate and compare the genomes of five *P. syringae* pathogens from various hosts. This would have enabled me to search broadly for a host-specific accessory genome, though unfortunately this was not possible given the data I was able to generate. Instead, I was able to observe patterns of growth *in vitro* and *in planta* for the pooled Tn-libraries. I offer my best explication of the experiments in hope that someone can benefit in future.

# CHAPTER 5

# CONCLUSIONS

Microbial isolates comprising an ecotype are adapted to their shared ecology, which is borne out in their patterns of relatedness; these works expand our knowledge of adaptation and ecotype evolution in *Pseudomonas syringae*. Chapter 1 examines the relatedness of three groups of strains from different sources (*Arabidopsis thaliana*, crop hosts, and environmental reservoirs), and supports them as distinctly evolving lineages. Additionally, I find that the *A. thaliana*-derived group represents a stable ecotype. However, host is a poor proxy for strain relatedness (outside of the *A. thaliana*-derived group). Chapter 2 compares the effector repertoires and growth of five isolates from different hosts in their own, and in each other's, host-of-isolation. Findings indicate that NP29 has a benefit across abiotic environments, and that local adaptation is present, but not ubiquitous, in this system. Chapter 3 describes attempts to generate a Tn-seq database that would have enabled elucidation of the pathogen genes underlying the infections described in Chapter 2. In lieu of these results, which I was unable to produce, I presented growth assays and compared the results of an insignificant TnSeq analysis. In summary, these findings suggest that *A. thaliana*-derived isolates act as an evolutionary ecotype, that host-of-isolation is not a reliable predictor of pathogen relatedness, and that pathovars are not necessarily adapted to the hosts from which they were isolated. These works also indicate a caution for the assumption of local adaptation implied by the use of pathovar systems for classification.

More generally, these results reflect that molecular signatures of host-specific adaptation are more likely to be in the variable than core genomes. It is possible that a focus on singletons (reviewed in Taut et al. 2011) and effectors (e.g., Collmer et al. 2009) specifically has been at the expense of other members of the variable genome that represent adaptation to particular host niches. Even core genes can come in variable copy numbers per strain, indicating the differential influence of selection and/or horizontal transfer even in genes

present in all members of a species. To fully understand ecological adaptation, a more comprehensive view of the variable genome is required.

### *5.0.1 Chapter 2 Caveats and future work*

Several points of evidence support the designation of *A. thaliana* derived isolates of *P. syringae* as a stable ecotype, but these are only correlative. For example, the At group achieves all three predictions of a stable ecotype defined by Cohan (2002). These include 1) that they form a monophyletic group, 2) that they form a star clade with individuals approximately equally related, and 3) that relatedness is greater within an ecotype than between. Each of these were true for the At group, in contrast to the crop and env groups, neither of which belong to one single monophyletic clade. Additionally, my findings of few effectors for the At group, along with similar, previous findings of others (Karasov et al. 2017) indicate the presence of a selective force for loss of effectors in *A. thaliana*. The At group pattern was again in contrast to the env and crop groups, which shared the majority of effectors discovered in this study. The fact that few effectors were also detected in *A. thaliana*-derived isolates of *P. syringae* from Michigan, USA (Karasov et al. 2017) indicates that having few effectors is not specific to just one *P. syringae* population. In fact, if the selective pressure comes from *A. thaliana*, it is unlikely that the effect is only present in *P. syringae*, as *A. thaliana* has several natural pathogens. It remains unclear why, relative to other plants, *A. thaliana* seems to contain so much potential for effector recognition. It may not be something unique to *A. thaliana* but simply a byproduct of the plant's selective environment; that is, sampled populations of *A. thaliana* often grown adjacent to agricultural fields (François et al. 2008) where they are likely to interact with at least some of the pathogens that infect crops. Thus, it may be that *A. thaliana* and similar weedy hosts have the evolutionary advantage over crop cultivars, in that they could have, over millions of years, naturally built-up extensive resistance to crop-infecting pathogens. It has been shown

that diffuse interactions with multiple pathogens can maintain resistance polymorphism in *A. thaliana* for millions of years (Karasov et al. 2015), so perhaps there is a case for a 'slow and steady' accumulation of resistance in this natural host.

Future work should compare *A. thaliana*-derived isolates of *P. syringae* from France to those from other geographic regions to test how well the ecotype finding holds against an expanded sample. Additionally, it would be of interest to know whether isolates of other gram-negative phytopathogens of *A. thaliana*, like *X. campestris*, also contain few effectors relative to isolates from other sources. This would indicate the presence of a powerful selective force represented by *A. thaliana* and acting against pathogen effector alleles beyond just the specific suite found in *P. syringae*.

### *5.0.2 Chapter 3 conclusions and caveats*

One caveat to the approach used in Chapter 3 is that I focused exclusively on carrying capacity as a proxy for fitness, when there is also rate of growth to consider. I feel confident in this choice as repeating the analysis with rate in place of $K$ generally made the results less informative. In the *in vitro* experiments for example, repeating the ANCOVA (rate $\sim$ environment * pathogen + plate) with rate in place of $K$ revealed a significant main effect of pathogen (F = 6.591 , $p = 0.000209$) and plate (F = 4.162, $p = 0.005134$), but no effect of environment or the interaction between environment and pathogen. This differs from the original ANCOVA with $K$, which uncovered significant effects of plant and pathogen, indicating that $K$ was a more informative choice than rate. For the in planta experiments, ANCOVA with rate (rate $\sim$ plant * pathogen + plate) yielded results that were entirely non-significant, indicating that pathogens grew at approximately the same rate across experiments, again indicating that $K$ was a more appropriate choice.

Another caveat to this work is that any direct test of local adaptation is limited by the number of plants tested. It is possible that I would have found different results if we had used

a different set of cultivars, but I did my best to find a cultivar either identical to or as similar as possible to the host-of-isolation. It seems unlikely that my finding of local adaptation of 1448A to bean and NP29 to *A. thaliana* would change, considering the number of hosts already tested. Future experiments should focus on testing for local adaptation of additional pathovars, since assuming local adaptation based on host-of-isolation is evidently not reliable. Also of note: a classification system completely divorced from pathogen phenotype, and instead based on whole genome sequences has been suggested (Baltrus 2016). This has the advantage of application to even unculturable isolates and does not require phenotypic testing for proper classification. Perhaps such a system should be adopted in the future.

### *5.0.3 Chapter 3 conclusions and caveats*

Chapter 3 was built on a solid question that I still wish I knew the answer to: is there a host-specific accessory genome in *P. syringae*? There is only one previously published example of a Tn-library of a *P. syringae* strain has been grown in more than one host (Helmann et al. 2020). This important study demonstrated the presence of a host-specific accessory genome for B728A in common bean, lima bean, and pepper grown in the apoplast. However, these results have yet to be replicated in another strain or even another lab, so the extent of their generality remains to be known. The obvious caveat to this chapter is that I was unable to generate significant TnSeq results, but I do not believe there was a fundamental problem with the experimental design. One lesson learned is to make full use of Illumina customer service when planning a sequencing experiment, since they will help make sure your scheme is compatible with the specific sequencer. Another is not to trust published sequences without verification. The extent to which *P. syringae* has a host-specific genome remains to be fully explored.

# CHAPTER 6

# REFERENCES

1. Achtman, M. & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. Nature Reviews Microbiology, 6(6), 431–440.

2. Alfano, J. R., Charkowski, A. O., Deng, W.-L., Badel, J. L., Petnicki-Ocwieja, T., Dijk, K. van & Collmer, A. (2000). The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. Proceedings of the National Academy of Sciences, 97(9), 4856–4861.

3. An, S., Potnis, N., Dow, M., Vorhölter, F.-J., He, Y., Becker, A., Teper, D., Li, Y., Wang, N., Bleris, L. & Tang, J. (2019). Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. FEMS Microbiology Reviews, 44(1), 1–32.

4. Arnold, D. L., LOVELL, H. C., Jackson, R. W. & Mansfield, J. W. *Pseudomonas syringae* pv. *phaseolicola*: from 'has bean' to supermodel. Molecular Plant Pathology 12, 617–627 (2011).

5. Baltrus, D. A., Guttman, D. S., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., Roach, J., Grant, S. R., Jones, C. D. & Dangl, J. L. (2011). Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 *Pseudomonas syringae* Isolates. PLoS Pathogens, 7(7), e1002132-22.

6. Baltrus, D. A. Divorcing Strain Classification from Species Names. Trends Microbiol.

24, 431–439 (2016).

7. Baltrus, D. A., McCann, H. C. & Guttman, D. S. (2016). Evolution, genomics and epidemiology of *Pseudomonas syringae*. Molecular Plant Pathology, 18(1), 152–168.

8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455-77.

9. Baranyi, Jozsef, and Terry A. Roberts. 1995. "Mathematics of Predictive Food Microbiology." International Journal of Food Microbiology 26 (2): 199–218.

10. Barrett, L. G., Bell, T., Dwyer, G. & Bergelson, J. (2011). Cheating, trade-offs and the evolution of aggressiveness in a natural pathogen population. Ecology Letters, 14(11), 1149–1157.

11. Bartoli, C., Berge, O., Monteil, C. L., Guilbaud, C., Balestra, G. M., Varvaro, L., Jones, C., Dangl, J. L., Baltrus, D. A., Sands, D. C. & Morris, C. E. (2014). The *Pseudomonas viridiflava* phylogroups in the *P. syringae* species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. Environmental Microbiology, 16(7), 2301–2315.

12. Bartoli, C., Lamichhane, J. R., Berge, O., Guilbaud, C., Varvaro, L., Balestra, G. M., Vinatzer, B. A. & Morris, C. E. (2015). The epidemic potential of plant pathogens. Molecular Plant Pathology, 16(2), 137–149.

13. Bell, G. & Reboud, X. Experimental evolution in Chlamydomonas. II. Genetic variation in strongly contrasted environments. Heredity (1997).

14. Bender, C. L., Alarcon-Chaidez, F. & Gross, D. C. (1999). Pseudomonas syringae Phytotoxins Mode of Action, Regulation, and Biosynthesis by Peptide and Polyketide Synthetases. Microbiology and Molecular Biology Reviews, 63(2), 266–292.

15. Berge, O., Vos, M., Monteil, C. L., Bartoli, C., Chandeysson, C., Guilbaud, C., Sands, D. C. & Morris, C. E. (2014). 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, 9(9), e105547-15.

16. Bergelson, J., Dwyer, G. & Emerson, J. J. (2001). Models and data on plant-enemy coevolution. Annual Review of Genetics, 35(1), 469–499.

17. Block, A. & Alfano, J. R. (2011). Plant targets for *Pseudomonas syringae* type III effectors: virulence targets or guarded decoys? Current Opinion in Microbiology, 14(1), 39–46.

18. Bobay, L.-M. & Ochman, H. (2018). Factors driving effective population size and pan-genome evolution in bacteria. BMC Evolutionary Biology, 1–12.

19. Boller, T. & He, S. Y. Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. Science 324, 742–744 (2009).

20. Nathalie Bonin, Enrique Doster, Hannah Worley, Lee J Pinnell, Jonathan E Bravo, Peter Ferm, Simone Marini, Mattia Prosperi, Noelle Noyes, Paul S Morley, Christina Boucher, MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing, Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D744–D752

21. Bouteiller, M., Dupont, C., Bourigault, Y., Latour, X., Barbey, C., Konto-Ghiorghi, Y. & Merieau, A. (2021). *Pseudomonas* Flagella: Generalities and Specificities. International Journal of Molecular Sciences, 22(7), 3337.

22. Buell, C. R. et al. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae pv. tomato* DC3000. Proceedings of the National Academy of Sciences of the United States of America 100, 10181–10186 (2003).

23. Bull, C. T. & Koike, S. T. (2015). Practical Benefits of Knowing the Enemy: Modern Molecular Tools for Diagnosing the Etiology of Bacterial Diseases and Understanding the Taxonomy and Diversity of Plant Pathogenic Bacteria. Annual Review of Phytopathology, 53(1), 1–24.

24. Burdman, S., Bahar, O., Parker, J. K. & Fuente, L. D. L. (2011). Involvement of Type IV Pili in Pathogenicity of Plant Pathogenic Bacteria. Genes, 2(4), 706–735.

25. Bushnell, Brian. BBMap: A Fast, Accurate, Splice-Aware Aligner. United States.

26. Cai, R., Guttman, D. S., Lewis, J., Yan, S., Liu, H., Clarke, C. R., Campanile, F., Almeida, N. F., Studholme, D. J., Lindeberg, M., Schneider, D., Zaccardelli, M., Setubal,

J. C., Morales-Lizcano, N. P., Bernal, A., Coaker, G., Baker, C., Bender, C. L., Leman, S. & Vinatzer, B. A. (2011). The Plant Pathogen *Pseudomonas syringae pv. tomato* Is Genetically Monomorphic and under Strong Selection to Evade Tomato Immunity. 7(8), e1002130-15.

27. Camacho, C., Coulouris, G., Avagyan, V. et al. BLAST+: architecture and applications. BMC Bioinformatics 10, 421 (2009).

28. Cameron, A. & Sarojini, V. (2014). *Pseudomonas syringae pv. actinidiae*: chemical control, resistance mechanisms and possible alternatives. Plant Pathology, 63(1), 1–11.

29. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug 1;25(15):1972-3.

30. Chao, M. C., Abel, S., Davis, B. M. & Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. Nature Reviews. Microbiology, 14(2), 119–128.

31. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. Nucleic Acids Res. 2016 Jan 4;44(D1):D694-7.

32. Choi, S., Jayaraman, J., Segonzac, C., Park, H.-J., Park, H., Han, S.-W. & Sohn, K. H. (2017). *Pseudomonas syringae pv. actinidiae* Type III Effectors Localized at Multiple Cellular Compartments Activate or Suppress Innate Immune Responses in *Nicotiana benthamiana*. Frontiers in Plant Science, 8, 2157.

33. Coe, K. A., Peschel, A., Lee, W., Stone, M. C., Komazin-Meredith, G., Meredith, T. C., Grad, Y. H. & Walker, S. (2019). Multi-strain Tn-Seq reveals common daptomycin resistance determinants in *Staphylococcus aureus*. PLoS Pathogens, 15(11), e1007862-26.

34. Cohan, F. M. (2001). Bacterial species and speciation. Systematic Biology.

35. Cohan, F. M. (2002). What are Bacterial Species? Annual Review of Microbiology, 56(1), 457–487.

36. Cohan, F. M. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1475), 1985–1996.

37. Collmer, A., Badel, J. L., Charkowski, A. O., Deng, W.-L., Fouts, D. E., Ramos, A. R., Rehm, A. H., Anderson, D. M., Schneewind, O., Dijk, K. van & Alfano, J. R. (2000). *Pseudomonas syringae* Hrp type III secretion system and effector

38. Collmer, A., Schneider, D. J. & Lindeberg, M. (2009). Lifestyles of the Effector Rich: Genome-Enabled Characterization of Bacterial Plant Pathogens. Plant Physiology, 150(4), 1623–1630.

39. Craig, K., Johnson, B. R. & Grunden, A. (2021). Leveraging *Pseudomonas* Stress Response Mechanisms for Industrial Applications. Frontiers in Microbiology, 12, 660134.

40. Croll, D. & Laine, A.-L. (2016). What the population genetic structures of host and

pathogen tell us about disease evolution. New Phytologist, 212(3), 537–539.

41. Cunnac, S., Chakravarthy, S., Kvitko, B. H., Russell, A. B., Martin, G. B. & Collmer, A. (2011). Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in *Pseudomonas syringae*. Proceedings of the National Academy of Sciences, 108(7), 2975–2980.

42. Dahale, S. K., Ghosh, D., Ingole, K. D., Chugani, A., Kim, S. H. & Bhattacharjee, S. (2021). HopA1 Effector from *Pseudomonas syringae pv syringae* Strain 61 Affects NMD Processes and Elicits Effector-Triggered Immunity. International Journal of Molecular Sciences, 22(14), 7440.

43. Dai, Z., Wu, T., Xu, S., Zhou, L., Tang, W., Hu, E., Zhan, L., Chen, M. & Yu, G. (2022). Characterization of toxin-antitoxin systems from public sequencing data: A case study in *Pseudomonas aeruginosa*. Frontiers in Microbiology, 13, 951774.

44. Dangl, J. L. & Jones, J. D. Plant pathogens and integrated defence responses to infection. Nature 411, 826–833 (2001).

45. Darzins, A. (1994). Characterization of a *Pseudomonas aeruginosa* gene cluster involved in pilus biosynthesis and twitching motility: sequence similarity to the chemotaxis proteins of enterics and the gliding bacterium *Myxococcus xanthus*. Molecular Microbiology, 11(1), 137–153.

46. Dasgupta, N., Arora, S. K. & Ramphal, R. (2000). fleN , a Gene That Regulates Flagellar Number in *Pseudomonas aeruginosa*. Journal of Bacteriology, 182(2), 357–364.

47. Dawkins R, Krebs JR. 1979. Arms races between and within species. Proc. R. Soc. London Ser. B 202:489–511

48. Destéfano, S. A. L., Rodrigues, L. M. R., Beriam, L. O. S., Patrício, F. R. A., Thomaziello, R. A., & Rodrigues Neto, J. (2010). Bacterial leaf spot of coffee caused by *Pseudomonas syringae pv. tabaci* in Brazil. New Disease Reports, 22.

49. (A) Dillon, M. M., Almeida, R. N. D., Laflamme, B., Martel, A., Weir, B. S., Desveaux, D. & Guttman, D. S. (2019). Molecular Evolution of *Pseudomonas syringae* Type III Secreted Effector Proteins. Frontiers in Plant Science, 10, 418.

50. (B) Dillon, M. M., Thakur, S., Almeida, R. N. D., Wang, P. W., Weir, B. S. & Guttman, D. S. (2019). Recombination of ecologically and evolutionarily significant loci maintains genetic cohesion in the *Pseudomonas syringae* species complex. Genome Biology, 1–28.

51. Dong, X., Mindrinos, M., Davis, K. R. & Ausubel, F. M. Induction of *Arabidopsis* defense genes by virulent and avirulent *Pseudomonas syringae* strains and by a cloned avirulence gene. The Plant Cell (1991).

52. Doyle, T. B., Hawkins, A. C. & McCarter, L. L. (2004). The Complex Flagellar Torque Generator of *Pseudomonas aeruginosa*. Journal of Bacteriology, 186(19), 6341–6350.

53. Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM. The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. J Biol Chem. 2007 Jul 13;282(28):20425-34.

54. Duong, D. A., Jensen, R. V. & Stevens, A. M. (2018). Discovery of *Pantoea stewartiis sp. stewartii* genes important for survival in corn xylem through a Tn-Seq analysis. Molecular Plant Pathology, 19(8), 1929–1941.

55. Eddy SR. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput. Biol. 4:e1000069 (2008)

56. S. R. Eddy. Accelerated profile HMM searches. PLOS Comp. Biol., 7:e1002195, 2011

57. Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792-97.

58. Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L. & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ, 3, e1319.

59. Ertesvåg H. Alginate-modifying enzymes: biological roles and biotechnological uses. Front Microbiol. 2015 May 27;6:523.

60. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048.

61. Fakhar, A. Z., Liu, J., Pajerowska-Mukhtar, K. M. & Mukhtar, M. S. (2023). The Lost and Found: Unraveling the Functions of Orphan Genes. Journal of Developmental

Biology, 11(2), 27.

62. Feehily, C. & Karatzas, K. A. G. (2013). Role of glutamate metabolism in bacterial responses towards acid and other stresses. Journal of Applied Microbiology, 114(1), 11–24.

63. Fleiszig, S. M. J., Arora, S. K., Van, R. & Ramphal, R. (2001). FlhA, a Component of the Flagellum Assembly Apparatus of *Pseudomonas aeruginosa*, Plays a Role in Internalization by Corneal Epithelial Cells. Infection and Immunity, 69(8), 4931–4937.

64. Franklin, M. J. & Ohman, D. E. (2002). Mutant Analysis and Cellular Localization of the AlgI, AlgJ, and AlgF Proteins Required for O Acetylation of Alginate in *Pseudomonas aeruginosa*. Journal of Bacteriology, 184(11), 3000–3007.

65. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. Science. 2007 Jan 26;315(5811):476-80. doi: 10.1126/science.1127573.

66. Fry, J. D. Trade-Offs in Fitness on Different Hosts: Evidence from a Selection Experiment with a Phytophagous Mite. Am Nat 136, 569–580 (1990).

67. Galazka, J. M. & Freitag, M. (2014). Variability of chromosome structure in pathogenic fungi—of 'ends and odds.' Current Opinion in Microbiology, 20, 19–26.

68. Gallagher, L. A., Shendure, J. & Manoil, C. (2010). Genome-Scale Identification of Resistance Functions in *Pseudomonas aeruginosa* Using Tn-seq. mBio, 2(1), e00315-10-e00315-10.

69. Gandon, S. Local adaptation and the geometry of host–parasite coevolution. Ecol Lett 5, 246–256 (2002).

70. Gimenez-Ibanez, S., Boter, M., Fernández-Barbero, G., Chini, A., Rathjen, J. P. & Solano, R. (2014). The Bacterial Effector HopX1 Targets JAZ Transcriptional Repressors to Activate Jasmonate Signaling and Promote Infection in *Arabidopsis*. PLoS Biology, 12(2), e1001792.

71. Fabian Giska, Małgorzata Lichocka, Marcin Piechocki, Michał Dadlez, Elmon Schmelzer, Jacek Hennig, Magdalena Krzymowska, Phosphorylation of HopQ1, a Type III Effector from *Pseudomonas syringae*, Creates a Binding Site for Host 14-3-3 Proteins, Plant Physiology, Volume 161, Issue 4, April 2013, Pages 2049–2061

72. Francois, M. T., Naturalist, L. F. T. A. & 2019. The Relationship between Spatial Structure and the Maintenance of Diversity in Microbial Populations.

73. Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, Dangl JL, Grant SR. The *Pseudomonas syringae* type III effector HopAM1 enhances virulence on water-stressed plants. Mol Plant Microbe Interact. 2008 Mar;21(3):361-70.

74. Goldberg, J. B. & Dahnke, T. (1992). *Pseudomonas aeruginosa* AlgB, which modulates the expression of alginate, is a member of the NtrC subclass of prokaryotic regulators. Molecular Microbiology, 6(1), 59–66.

75. Gomila, M., Busquets, A., Mulet, M., García-Valdés, E. & Lalucat, J. (2017). Clarification of Taxonomic Status within the *Pseudomonas syringae* Species Group Based on a

Phylogenomic Analysis. Frontiers in Microbiology, 8, 2422.

76. Goss, E. M., Kreitman, M. & Bergelson, J. (2005). Genetic Diversity, Recombination and Cryptic Clades in *Pseudomonas viridiflava* Infecting Natural Populations of *Arabidopsis thaliana*. Genetics, 169(1), 21–35.

77. Gould, F. RAPID HOST RANGE EVOLUTION IN A POPULATION OF THE PHYTOPHAGOUS MITE *TETRANYCHUS URTICAE* KOCH. Evolution 33, 791–802 (1979).

78. Greenwald, J. W., Greenwald, C. J., Philmus, B. J., Begley, T. P. & Gross, D. C. (2012). RNA-seq Analysis Reveals That an ECF $\sigma$ Factor, AcsS, Regulates Achromobactin Biosynthesis in *Pseudomonas syringae pv. syringae* B728a. PLoS ONE, 7(4), e34804.

79. Guo, M., Kim, P., Li, G., Elowsky, C. G. & Alfano, J. R. (2016). A Bacterial Effector Co-opts Calmodulin to Target the Plant Microtubule Network. Cell Host & Microbe, 19(1), 67–78.

80. Guo, M., Tian, F., Wamboldt, Y. & Alfano, J. R. (2009). The Majority of the Type III Effector Inventory of *Pseudomonas syringae pv. tomato* DC3000 Can Suppress Plant Immunity. Molecular Plant-Microbe Interactions®, 22(9), 1069–1080.

81. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29(8), 1072–1075.

82. Haapalainen, M., Mosorin, H., Dorati, F., Wu, R.-F., Roine, E., Taira, S., Nissinen, R., Mattinen, L., Jackson, R., Pirhonen, M. & Lin, N.-C. (2012). Hcp2, a Secreted Protein

of the Phytopathogen *Pseudomonas syringae pv. Tomato* DC3000, Is Required for Fitness for Competition against Bacteria and Yeasts. Journal of Bacteriology, 194(18), 4810–4822.

83. Harmsen, M., Yang, L., Pamp, S. J. & Tolker-Nielsen, T. (2010). An update on *Pseudomonas aeruginosa* biofilm formation, tolerance, and dispersal. FEMS Immunology & Medical Microbiology, 59(3), 253–268.

84. Heiden, A. C., Kobel, K., Komenda, M., Koppmann, R., Shao, M. & Wildt, J. (1999). Toluene emissions from plants. Geophysical Research Letters, 26(9), 1283–1286.

85. Helmann, T. C., Deutschbauer, A. M. & Lindow, S. E. (2019). Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. Proceedings of the National Academy of Sciences of the United States of America, 116(38), 18900–18910.

86. Helmann, T. C., Deutschbauer, A. M. & Lindow, S. E. (2020). Distinctiveness of genes contributing to growth of *Pseudomonas syringae* in diverse host plant species. PloS One, 15(9).

87. Hentchel, K. L., Ruiz, L. M. R., Curtis, P. D., Fiebig, A., Coleman, M. L. & Crosson, S. (2018). Genome-scale fitness profile of *Caulobacter crescentus* grown in natural freshwater. The ISME Journal, 1–14.

88. Hershberg, R. & Petrov, D. A. (2010). Evidence That Mutation Is Universally Biased towards AT in Bacteria. PLoS Genetics, 6(9), e1001115.

89. Higgins, S., Sanchez-Contreras, M., Gualdi, S., Carbó, M. P., Carlier, A. & Eberl, L. (2017). The Essential Genome of *Burkholderia cenocepacia* H111. Journal of Bacteriology, 199(22).

90. Hockett, K. L., Renner, T. & Baltrus, D. A. (2015). Independent Co-Option of a Tailed Bacteriophage into a Killing Complex in **Pseudomonas**. mBio, 6(4), e00452-15.

91. Hwang, M. S. H., Morgan, R. L., Sarkar, S. F., Wang, P. W. & Guttman, D. S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. Applied and Environmental Microbiology, 71(9), 5182–5191.

92. Hyatt, D., Chen, GL., LoCascio, P.F. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119 (2010).

93. Ishiyama, Y., Yamagishi, N., Ogiso, H., Fujinaga, M. & Takikawa, Y. Bacterial brown spot on *Avena storigosa* Schereb. caused by *Pseudomonas syringae pv. alisalensis*. J Gen Plant Pathol 79, 155–157 (2013).

94. Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M. V. & Manoil, C. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences of the United States of America, 100(24), 14339–14344.

95. Jacques, M.-A. et al. Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*: French Network on Xanthomonads (FNX). Annu. Rev. Phy-

topathol. 54, 1–25 (2015).

96. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications, 9(1), 7200–7208.

97. Jasmin, J.-N. & Kassen, R. On the experimental evolution of specialization and diversity in heterogeneous environments. Ecology Letters 10, 272–281 (2007).

98. Jelenska, J., Yao, N., Vinatzer, B. A., Wright, C. M., Brodsky, J. L. & Greenberg, J. T. (2007). A J Domain Virulence Effector of *Pseudomonas syringae* Remodels Host Chloroplasts and Suppresses Defenses. Current Biology, 17(6), 499–508.

99. Joardar, V. et al. Whole-Genome Sequence Analysis of P*seudomonas syringae pv. phaseolicola* 1448A Reveals Divergence among Pathovars in Genes Involved in Virulence and Transposition. Journal of bacteriology 187, 6488–6498 (2005).

100. Jones, J. D. G. & Dangl, J. L. (2006). The plant immune system. Nature, 444(7117), 323–329.

101. Kaltz, O. & Shykoff, J. A. (1998). Local adaptation in host–parasite systems. Heredity.

102. Karasov, T. L., Horton, M. W. & Bergelson, J. (2014). Genomic variability as a driver of plant–pathogen coevolution? Current Opinion in Plant Biology, 18, 24–30.

103. Karasov, T. L., Kniskern, J. M., Gao, L., DeYoung, B. J., Ding, J., Dubiella, U.,

Lastra, R. O., Nallu, S., Roux, F., Innes, R. W., Barrett, L. G., Hudson, R. R. & Bergelson, J. (2015). The long-term maintenance of a resistance polymorphism through diffuse interactions. Nature, 512(7515), 436–440.

104. Karasov, T. L., Kuo, C.-H., Barrett, L., Hershberg, R. & Bergelson, J. (2017). Similar levels of gene content variation observed for *Pseudomonas syringae* populations extracted from single and multiple host species. PloS One, 12(9), e0184195.

105. Karasov, T. L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D. S., Neumann, M., Regalado, J., Neher, R. A., Kemen, E. & Weigel, D. (2018). *Arabidopsis thaliana* and *Pseudomonas* Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host and Microbe, 24(1), 168-179.e4.

106. Keller, L. M., Colman, D. R. & Boyd, E. S. (2023). An active microbiome in Old Faithful geyser. PNAS Nexus, 2(3), pgad066.

107. Kievit, T. R. de & Lam, J. S. (1997). Isolation and characterization of two genes, waaC (rfaC) and waaF (rfaF), involved in *Pseudomonas aeruginosa* serotype O5 inner-core biosynthesis. Journal of Bacteriology, 179(11), 3451–3457.

108. Knight, T. J., Durbin, R. D. & Langston-Unkefer, P. J. (1986). Role of glutamine synthetase adenylylation in the self-protection of *Pseudomonas syringae* subsp. "tabaci" from its toxin, tabtoxinine-beta-lactam. Journal of Bacteriology, 166(1), 224–229.

109. Kniskern, J. M., Barrett, L. G. & Bergelson, J. MALADAPTATION IN WILD POPULATIONS OF THE GENERALIST PLANT PATHOGEN PSEUDOMONAS SYRINGAE.

Evolution 65, 818–830 (2011).

110. Koike, S. T., Alger, E. I., Sepulveda, L. R. & Bull, C. T. First Report of Bacterial Leaf Spot Caused by *Pseudomonas syringae pv. tomato* on Kale in California. Plant Dis 101, 504–504 (2017).

111. Konstantinidis, K. T. & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences, 102(7), 2567–2572.

112. Kunkeaw, S., Tan, S. & Coaker, G. Molecular and evolutionary analyses of *Pseudomonas syringae pv. tomato* race 1. Molecular plant-microbe ... 23, 415–424 (2010).

113. Laflamme, B., Dillon, M. M., Martel, A., Almeida, R. N. D., Desveaux, D. & Guttman, D. S. (2020). The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. Science, 367(6479), 763–768.

114. Lam, J. S., Taylor, V. L., Islam, S. T., Hao, Y. & Kocíncová, D. (2011). Genetic and Functional Diversity of *Pseudomonas aeruginosa* Lipopolysaccharide. Frontiers in Microbiology, 2, 118.

115. Lamont, I. L. & Martin, L. W. (2003). Identification and characterization of novel pyoverdine synthesis genes in *Pseudomonas aeruginosa*. Microbiology, 149(4), 833–842.

116. Lan, R., Reeves, P. R., Lan, R. & Reeves, P. R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends in Microbiology, 8(9),

396–401.

117. Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J. & Turner, A. K. (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Research, 19(12), 2308–2316.

118. Lederberg J. Infectious disease as an evolutionary paradigm. Emerg Infect Dis. 1997 Oct-Dec;3(4):417-23.

119. Lee, S. A., Gallagher, L. A., Thongdee, M., Staudinger, B. J., Lippman, S., Singh, P. K. & Manoil, C. (2015). General and condition-specific essential functions of *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences, 112(16), 5189–5194.

120. Lelliott, R. A., Billing, E. & Hayward, A. C. (1966). A Determinative Scheme for the Fluorescent Plant Pathogenic Pseudomonads. Journal of Applied Bacteriology, 29(3), 470–489.

121. Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research, 49(W1), W293–W296.

122. Levin, B. R., Perrot, V. & Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics, 154(3), 985–997.

123. Lewis, J. D., Abada, W., Ma, W., Guttman, D. S. & Desveaux, D. (2008). The

HopZ Family of *Pseudomonas syringae* Type III Effectors Require Myristoylation for Virulence and Avirulence Functions in *Arabidopsis thaliana*. Journal of Bacteriology, 190(8), 2880–2891.

124. Li, W., Chiang, Y.-H. & Coaker, G. (2013). The HopQ1 Effector's Nucleoside Hydrolase-Like Domain Is Required for Bacterial Virulence in *Arabidopsis* and Tomato, but Not Host Recognition in Tobacco. PLoS ONE, 8(3), e59684.

125. Li G, Froehlich JE, Elowsky C, Msanne J, Ostosh AC, Zhang C, Awada T, Alfano JR. Distinct *Pseudomonas* type-III effectors use a cleavable transit peptide to target chloroplasts. Plant J. 2014 Jan;77(2):310-21.

126. Lim, M. T. S. & Kunkel, B. N. (2004). Mutations in the *Pseudomonas syringae* avrRpt2 gene That Dissociate Its Virulence and Avirulence Activities Lead to Decreased Efficiency in AvrRpt2-Induced Disappearance of RIN4. Molecular Plant-Microbe Interactions®, 17(3), 313–321.

127. Linares JF, López JA, Camafeita E, Albar JP, Rojo F, Martínez JL. Overexpression of the multidrug efflux pumps MexCD-OprJ and MexEF-OprN is associated with a reduction of type III secretion in *Pseudomonas aeruginosa*. J Bacteriol. 2005 Feb;187(4):1384-91.

128. LINDEBERG, M., CUNNAC, S. & COLLMER, A. (2009). The evolution of *Pseudomonas syringae* host specificity and type III effector repertoires. Molecular Plant Pathology, 10(6), 767–775.

129. Lindow, S. E. (1983). The role of bacterial ice nucleation in frost injury to plants.

Annual Review of Phytopathology.

130. Liu, J., Yu, M., Chatnaparat, T., Lee, J. H., Tian, Y., Hu, B. & Zhao, Y. (2020). Comparative transcriptomic analysis of global gene expression mediated by (p) ppGpp reveals common regulatory networks in *Pseudomonas syringae*. BMC Genomics, 21(1), 296.

131. Lo, T., Koulena, N., Seto, D., Guttman, D. S. & Desveaux, D. (2017). The HopF family of *Pseudomonas syringae* type III secreted effectors. Molecular Plant Pathology, 18(3), 457–468.

132. Lopez, V.A., Park, B.C., Nowak, D., Sreelatha, A., Zembek, P., Fernadez, J. et al. (2019) A bacterial effector mimics a host HSP90 client to undermine immunity. Cell, 179, 205–218.e21.

133. Lossi, N. S., Manoli, E., Förster, A., Dajani, R., Pape, T., Freemont, P. & Filloux, A. (2013). The HsiB1C1 (TssB-TssC) Complex of the *Pseudomonas aeruginosa* Type VI Secretion System Forms a Bacteriophage Tail Sheathlike Structure. Journal of Biological Chemistry, 288(11), 7536–7548.

134. LOVELL, H. C. et al. In planta conditions induce genomic changes in *Pseudomonas syringae pv. phaseolicola*. Molecular Plant Pathology 12, 167–176 (2010).

135. Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D (2021). "performance: An R Package for Assessment, Comparison and Testing of Statistical Models." Journal of Open Source Software, 6(60), 3139.

136. Manoharan, B., Lu, H., Neale, H. C., Hancock, J. T., Jackson, R. W. & Arnold, D. L. (2015). The Identification of Genes Important in *Pseudomonas syringae pv. phaseolicola* Plant Colonisation Using In Vitro Screening of Transposon Libraries. PloS One, 10(9), e0137355-19.

137. Marcelletti S, Ferrante P, Petriccione M, Firrao G, Scortichini M. Pseudomonas syringae pv. actinidiae draft genomes comparison reveal strain-specific features involved in adaptation and virulence to Actinidia species. PLoS One. 2011;6(11):e27297.

138. Markel, E., Stodghill, P., Bao, Z., Myers, C. R. & Swingle, B. (2016). AlgU Controls Expression of Virulence Genes in *Pseudomonas syringae pv. tomato* DC3000. Journal of Bacteriology, 198(17), 2330–2344.

139. Mayjonade, B. et al. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. BioTechniques 61, 1–3 (2016).

140. Mazerolle MJ (2023). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c).

141. McAtee, P. A., Brian, L., Curran, B., Linden, O. van der, Nieuwenhuizen, N. J., Chen, X., Henry-Kirk, R. A., Stroud, E. A., Nardozza, S., Jayaraman, J., Rikkerink, E. H. A., Print, C. G., Allan, A. C. & Templeton, M. D. (2018). Re-programming of *Pseudomonas syringae pv. actinidiae* gene expression during early stages of infection of kiwifruit. BMC Genomics, 19(1), 822.

142. McCann, H. C., Li, L., Liu, Y., Li, D., Pan, H., Zhong, C., Rikkerink, E. H. A.,

Templeton, M. D., Straub, C., Colombi, E., Rainey, P. B. & Huang, H. (2017). Origin and Evolution of the Kiwifruit Canker Pandemic. Genome Biology and Evolution, 9(4), 932–944.

143. Meaden, S. & Koskella, B. Adaptation of the pathogen, *Pseudomonas syringae*, during experimental evolution on a native vs. alternative host plant. Molecular ecology 26, 1790–1801 (2017).

144. Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. Y. (2006). Plant Stomata Function in Innate Immunity against Bacterial Invasion. Cell, 126(5), 969–980.

145. Portia Mira and others, Statistical Package for Growth Rates Made Easy, Molecular Biology and Evolution, Volume 34, Issue 12, December 2017, Pages 3303–3309

146. Monteil, C. L., Cai, R., Liu, H., Llontop, M. E. M., Leman, S., Studholme, D. J., Morris, C. E. & Vinatzer, B. A. (2013). Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. New Phytologist, 199(3), 800–811.

147. Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Méric, G., Swingle, B., Morris, C. E., Vinatzer, B. A. & Sheppard, S. K. (2016). Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas syringae* pathogens. 2(10), 1–16.

148. Morlot C, Uehara T, Marquis KA, Bernhardt TG, Rudner DZ. A highly coordinated cell wall degradation machine governs spore morphogenesis in *Bacillus subtilis*. Genes Dev. 2010 Feb 15;24(4):411-22.

149. Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H. & Thompson, B. M. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. The ISME Journal, 2(3), 321–334.

150. Morris, C. E., Monteil, C. L. & Berge, O. (2013). The Life History of *Pseudomonas syringae*: Linking Agriculture to Earth System Processes. 51(1), 85–104.

151. Morris, C. E., Lamichhane, J. R., Nikolić, I., Stanković, S. & Moury, B. (2019). The overlapping continuum of host range among strains in the *Pseudomonas syringae* complex. Phytopathology Research, 1–16.

152. Morris, J. J., Lenski, R. E. & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio, 3(2), e00036-12-e00036-12.

153. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordoñez CL, Lory S, Walz T, Joachimiak A, Mekalanos JJ. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science. 2006 Jun 9;312(5779):1526-30.

154. Morschhäuser, J., Köhler, G., Ziebuhr, W., Blum–Oehler, G., Dobrindt, U. & Hacker, J. (2000). Evolution of microbial pathogens. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 355(1397), 695–704.

155. Munkvold KR, Russell AB, Kvitko BH, Collmer A. *Pseudomonas syringae pv. tomato* DC3000 type III effector HopAA1-1 functions redundantly with chlorosis-promoting factor PSPTO4723 to produce bacterial speck lesions in host tomato. Mol Plant Microbe Interact.

2009 Nov;22(11):1341-55.

156. Nagendran, R. & Lee, Y. H. (2019). SrfC of *Pseudomonas cichorii* JBC1 affects its attachment to the host surface and host tissue infection. Plant Pathology, 68(6), 1099–1108.

157. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution, 32(1), 268–274.

158. Nicaise, V., Joe, A., Jeong, B., Korneli, C., Boutrot, F., Westedt, I., Staiger, D., Alfano, J. R. & Zipfel, C. (2013). *Pseudomonas* HopU1 modulates plant immune receptor levels by blocking the interaction of their mRNAs with GRP7. The EMBO Journal, 32(5), 701–712.

159. Nikolić, I., Glatter, T., Ranković, T., Berić, T., Stanković, S. & Diepold, A. (2023). Repertoire and abundance of secreted virulence factors shape the pathogenic capacity of *Pseudomonas syringae pv. aptata*. Frontiers in Microbiology, 14, 1205257. 160. Ochman, H. & Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science.

161. Oglesby, L. L., Jain, S. & Ohman, D. E. (2008). Membrane topology and roles of *Pseudomonas aeruginosa* Alg8 and Alg44 in alginate polpolymerization. Microbiology, 154(6), 1605–1615

162. Opijnen, T. van, Bodi, K. L. & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature Methods,

6(10), 767–772.

163. Ostrowski, E. A., ROZEN, D. E. & Lenski, R. E. Pleiotropic effects of beneficial mutations in *Escherichia coli*. Evolution (2005).

164. Paradis E (2010). "pegas: an R package for population genetics with an integrated–modular approach." Bioinformatics, 26, 419-420.

165. Park SY, Chao X, Gonzalez-Bonet G, Beel BD, Bilwes AM, Crane BR. Structure and function of an unusual family of protein phosphatases: the bacterial chemotaxis proteins CheC and CheX. Mol Cell. 2004 Nov 19;16(4):563-74.

166. PAULING, L. (1964). MOLECULAR DISEASE. Bulletin of the New York Academy of Medicine, 40, 334–342.

167. Peñaloza-Vázquez, A., Fakhr, M. K., Bailey, A. M. & Bender, C. L. (2004). AlgR functions in algC expression and virulence in *Pseudomonas syringae pv. syringae*. Microbiology, 150(8), 2727–2737.

168. Petitjean, M., Martak, D., Silvant, A., Bertrand, X., Valot, B. & Hocquet, D. (2017). Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395. Microbial Genomics, 3(10), e000129.

169. Platt, A. et al. The Scale of Population Structure in *Arabidopsis thaliana*. PLoS Genetics 6, e1000843 (2010).

170. Poulsen, B. E., Yang, R., Clatworthy, A. E., White, T., Osmulski, S. J., Li, L., Penaranda, C., Lander, E. S., Shoresh, N. & Hung, D. T. (2019). Defining the core essential genome of *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences of the United States of America, 116(20), 10072–10080.

171. Preston, G. M. (2000). *Pseudomonas syringae pv. tomato*: the right pathogen, of the right plant, at the right time. Molecular Plant Pathology, 1(5), 263–275.

172. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

173. Read, A. F. & Taylor, L. H. (2001). The Ecology of Genetically Diverse Infections. Science, 292(5519), 1099–1102.

174. Rosenberg, S. M. & Queitsch, C. (2014). Medicine. Combating evolution to fight disease. Science, 343(6175), 1088–1089.

175. Ross, A. & Somssich, I. E. A DNA-based real-time PCR assay for robust growth quantification of the bacterial pathogen *Pseudomonas syringae* on *Arabidopsis thaliana*. Plant Methods 12, 473–8 (2016).

176. Russell, A. R., Ashfield, T. & Innes, R. W. (2015). *Pseudomonas syringae* Effector AvrPphB Suppresses AvrB-Induced Activation of RPM1 but Not AvrRpm1-Induced Activation. Molecular Plant-Microbe Interactions®, 28(6), 727–735.

177. Sarkar, S. F. & Guttman, D. S. (2004). Evolution of the Core Genome of *Pseu-*

*domonas syringae*, a Highly Clonal, Endemic Plant Pathogen. Applied and Environmental Microbiology, 70(4), 1999–2012.

178. Sarris, P. F. et al. Comparative Genomics of Multiple Strains of *Pseudomonas cannabina pv. alisalensis*, a Potential Model Pathogen of Both Monocots and Dicots. PloS one 8, e59366-16 (2013).

179. Sawada, T., Eguchi, M., Asaki, S., Kashiwagi, R., Shimomura, K., Taguchi, F., Matsui, H., Yamamoto, M., Noutoshi, Y., Toyoda, K. & Ichinose, Y. (2018). MexEF-OprN multidrug efflux pump transporter negatively controls N-acyl-homoserine lactone accumulation in *Pseudomonas syringae pv. Tabaci* 6605. Molecular Genetics and Genomics, 293(4), 907–917.

180. Schreiber, K. J., Semsey, S., Ye, D., Fich, E., Jian, A., Lo, T. & Desveaux, D. (2012). A High-Throughput Forward Genetic Screen Identifies Genes Required for Virulence of *Pseudomonas syringae pv. maculicola* ES4326 on Arabidopsis. 7(8), e41461-9.

181. Seemann T, Abricate, Github https://github.com/tseemann/abricate

182. Shapiro, B. J. (2016). How clonal are bacteria over time? Current Opinion in Microbiology, 31, 116–123.

183. Shen, W., Le, S., Li, Y. & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS ONE, 11(10), e0163962.

184. Shuman, H. A. & Silhavy, T. J. (2003). The art and design of genetic screens: *Es-*

*cherichia coli*. Nature Reviews Genetics, 4(6), 419–431.

185. Siew, N. & Fischer, D. (2003). Analysis of singleton ORFans in fully sequenced microbial genomes. Proteins: Structure, Function, and Bioinformatics, 53(2), 241–251.

186. Snelders, N. C., Petti, G. C., Berg, G. C. M. van den, Seidl, M. F. & Thomma, B. P. H. J. (2021). An ancient antimicrobial protein co-opted by a fungal plant pathogen for *in planta* mycobiome manipulation. Proceedings of the National Academy of Sciences, 118(49), e2110968118.

187. Snelders, N. C., Rovenich, H., Petti, G. C., Rocafort, M., Berg, G. C. M. van den, Vorholt, J. A., Mesters, J. R., Seidl, M. F., Nijland, R. & Thomma, B. P. H. J. (2020). Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. Nature Plants, 6(11), 1365–1374.

188. Spiers, A. J., Buckling, A. & Rainey, P. B. (2000). The causes of Pseudomonas diversity. Microbiology.

189. Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. Bmc Bioinformatics 17, 172 (2016).

190. Stahl EA, Bishop JG. Plant-pathogen arms races at the molecular level. Curr Opin Plant Biol. 2000 Aug;3(4):299-304.

191. Starnbach MN, Lory S. The fliA (rpoF) gene of *Pseudomonas aeruginosa* encodes an alternative sigma factor required for flagellin synthesis. Mol Microbiol. 1992 Feb;6(4):459-69.

192. Stringfellow JM, Turpin B, Cooper RA. Sequence of the *Escherichia coli* C homoprotocatechuic acid degradative operon completed with that of the 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase-encoding gene (hpcH). Gene. 1995 Dec 1;166(1):73-6.

193. Studholme, D. J., Downie, J. A. & Preston, G. M. (2005). Protein domains and architectural innovation in plant-associated Proteobacteria. BMC Genomics, 6(1), 17.

194. Suárez-Díaz, E. (2019). The Molecular Basis of Evolution and Disease: A Cold War Alliance. Journal of the History of Biology, 52(2), 325–346.

195. Taguchi, F., Shimizu, R., Inagaki, Y., Toyoda, K., Shiraishi, T. & Ichinose, Y. (2003). Post-Translational Modification of Flagellin Determines the Specificity of HR Induction. Plant and Cell Physiology, 44(3), 342–349.

196. 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.

197. F Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism., Genetics, Volume 123, Issue 3, 1 November 1989, Pages 585–595.

198. Tammam, S., Sampaleanu, L. M., Koo, J., Manoharan, K., Daubaras, M., Burrows, L. L. & Howell, P. L. (2013). PilMNOPQ from the *Pseudomonas aeruginosa* Type IV Pilus System Form a Transenvelope Protein Interaction Network That Interacts with sPilA. Journal of Bacteriology, 195(10), 2126–2135.

199. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000 Jan 1;28(1):33-6.

200. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). A Genomic Perspective on Protein Families. Science, 278(5338), 631–637.

201. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. Nat. Rev. Genet. 12, 692–702 (2011).

202. Tian, T., Wu, X.-G., Duan, H.-M. & Zhang, L.-Q. (2010). The resistance-nodulation-division efflux pump EmhABC influences the production of 2,4-diacetylphloroglucinol in *Pseudomonas fluorescens* 2P24. Microbiology, 156(1), 39–48.

203. Tian ZX, Yi XX, Cho A, O'Gara F, Wang YP. CpxR Activates MexAB-OprM Efflux Pump Expression and Enhances Antibiotic Resistance in Both Laboratory and Clinical nalB-Type Isolates of *Pseudomonas aeruginosa*. PLoS Pathog. 2016 Oct 13;12(10):e1005932.

204. Trunk, K., Coulthurst, S. J. & Quinn, J. (2019). A New Front in Microbial Warfare—Delivery of Antifungal Effectors by the Type VI Secretion System. Journal of Fungi, 5(2), 50–13.

205. Tsiamis G, Mansfield JW, Hockenhull R, Jackson RW, Sesma A, Athanassopoulos E, Bennett MA, Stevens C, Vivian A, Taylor JD, Murillo J. Cultivar-specific avirulence and virulence functions assigned to avrPphF in *Pseudomonas syringae pv. phaseolicola*, the cause

of bean halo-blight disease. EMBO J. 2000 Jul 3;19(13):3204-14.

206. Vandenende, C. S., Vlasschaert, M. & Seah, S. Y. K. (2004). Functional Characterization of an Aminotransferase Required for Pyoverdine Siderophore Biosynthesis in *Pseudomonas aeruginosa* PAO1. Journal of Bacteriology, 186(17), 5596–5602.

207. Vanneste, J. L. (2016). The Scientific, Economic, and Social Impacts of the New Zealand Outbreak of Bacterial Canker of Kiwifruit (*Pseudomonas syringae pv. actinidiae*). Annual Review of Phytopathology, 55(1), 1–23.

208. Vinatzer, B. A., Jelenska, J. & Greenberg, J. T. (2005). Bioinformatics Correctly Identifies Many Type III Secretion Substrates in the Plant Pathogen *Pseudomonas syringae* and the Biocontrol Isolate P. fluorescens SBW25. Molecular Plant-Microbe Interactions®, 18(8), 877–888.

209. Vinatzer, B. A., Teitzel, G. M., Lee, M., Jelenska, J., Hotton, S., Fairfax, K., Jenrette, J. & Greenberg, J. T. (2006). The type III effector repertoire of *Pseudomonas syringae pv. syringae* B728a and its role in survival and disease on host and non-host plants. Molecular Microbiology, 62(1), 26–44.

210. Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kühl, M. & Grossman, A. (2008). Genomics, environmental genomics and the issue of microbial species. Heredity, 100(2), 207–219.

211. Wei, H.-L., Chakravarthy, S., Mathieu, J., Helmann, T. C., Stodghill, P., Swingle, B., Martin, G. B. & Collmer, A. (2015). Pseudomonas syringae pv. tomato DC3000 Type

III Secretion Effector Polymutants Reveal an Interplay between HopAD1 and AvrPtoB. Cell Host & Microbe, 17(6), 752–762.

212. Weiler, E. W., Kutchan, T. M., Gorba, T., Brodschelm, W., Niesel, U. & Bublitz, F. (1994). The *Pseudomonas* phytotoxin coronatine mimics octadecanoid signalling molecules of higher plants. FEBS Letters, 345(1), 9–13.

213. Wilson, G. A., Bertrand, N., Patel, Y., Hughes, J. B., Feil, E. J. & Field, D. (2005). Orphans as taxonomically restricted and ecologically important genes. Microbiology, 151(8), 2499–2501.

214. Wiser, M. J. & Lenski, R. E. A Comparison of Methods to Measure Fitness in *Escherichia coli*. PLoS ONE 10, e0126210 (2015).

215. Wolf, Y. I., Makarova, K. S., Lobkovsky, A. E. & Koonin, E. V. (2016). Two fundamentally different classes of microbial genes. Nature Microbiology, 2(3), 16208.

216. Xin, X.-F., Kvitko, B. & He, S.-Y. (2018). *Pseudomonas syringae*: what it takes to be a pathogen. Nature Reviews. Microbiology, 16(5), 316–328.

217. Yao, J., Ericson, M. E., Frank, M. W. & Rock, C. O. (2016). Enoyl-Acyl Carrier Protein Reductase I (FabI) Is Essential for the Intracellular Growth of Listeria monocytogenes. Infection and Immunity, 84(12), 3597–3607.

218. Yi, L. & Dalbey, R. E. (2009). Oxal/Alb3/YidC system for insertion of membrane proteins in mitochondria, chloroplasts and bacteria (Review). Molecular Membrane Biology,

22(1–2), 101–111.

219. Young J.M., Dye D.W., Bradbury J.F., Panagopoulos C.G., Robbs C.F., 1978. A proposed nomenclature and classification for plant pathogenic bacteria. New Zealand Journal of Agricultural Research 21: 153-177.

220. Yu, X., Lund, S. P., Scott, R. A., Greenwald, J. W., Records, A. H., Nettleton, D., Lindow, S. E., Gross, D. C. & Beattie, G. A. (2013). Transcriptional responses of *Pseudomonas syringae* to growth in epiphytic versus apoplastic leaf sites. Proceedings of the National Academy of Sciences, 110(5), E425-34.

221. Zhang, J., Shao, F., Li, Y., Cui, H., Chen, L., Li, H., Zou, Y., Long, C., Lan, L., Chai, J., Chen, S., Tang, X. & Zhou, J.-M. (2007). A *Pseudomonas syringae* Effector Inactivates MAPKs to Suppress PAMP-Induced Immunity in Plants. Cell Host & Microbe, 1(3), 175–185.

222. Zhang, L., Li, X., Zhang, X., Li, Y. & Wang, L. (2021). Bacterial alginate metabolism: an important pathway for bioconversion of brown algae. Biotechnology for Biofuels, 14(1), 158.

223. Zhang Z, Wu Y, Gao M, Zhang J, Kong Q, Liu Y, Ba H, Zhou J, Zhang Y. Disruption of PAMP-induced MAP kinase cascade by a *Pseudomonas syringae* effector activates plant immunity mediated by the NB-LRR protein SUMM2. Cell Host Microbe. 2012 Mar 15;11(3):253-63.

224. Zheng, J., Zhang, J., Gao, L., Kong, F., Shen, G., Wang, R., Gao, J. & Zhang, J.

(2020). The Effects of Tetracycline Residues on the Microbial Community Structure of Tobacco Soil in Pot Experiment. Scientific Reports, 10(1), 8804.

225. Zheng X, Zhou Z, Gong Z, Hu M, Ahn YJ, Zhang X, Zhao Y, Gong G, Zhang J, Zuo J, Han GZ, Hoon SK, Zhou JM. Two plant NLR proteins confer strain-specific resistance conditioned by an effector from *Pseudomonas syringae pv. actinidiae*. J Genet Genomics. 2022 Aug;49(8):823-832.