Title
Associations between plant hosts and the pathogen X. fastidiosa

Author
Kahn, Alexandra Katz

Publication Date
2023

Peer reviewed|Thesis/dissertation

Associations between plant hosts and the pathogen *X. fastidiosa*

By

Alexandra Katz Kahn

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Rodrigo Almeida, Chair
Professor Rachel Brem
Professor Britt Koskella

Summer 2023

Associations between plant hosts and the pathogen *X. fastidiosa*

By

Alexandra Katz Kahn

Abstract

Associations between plant hosts and the pathogen *X. fastidiosa*

by

Alexandra Katz Kahn

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Rodrigo P.P. Almeida, Chair

Xylella fastidiosa was the first bacterial plant pathogen to have its whole genome sequenced. Nonetheless, the scientific community remains uncertain about the mechanism by which this pathogen causes disease. The determinants of its host range are largely unknown. While as a species, *X. fastidiosa* can infect hundreds of plants, any individual strain's potential hosts cannot be predicted. This dissertation is an interrogation of host specificity at three vastly different scales. The research is ordered from broad to narrow, beginning with a global analysis. I computationally modeled pathogen host range using 23 plant genera. Through this breadth, I was able to find correlations between hosts and genome content, as well as make estimates for likely ancestral hosts at internal phylogenetic nodes. My second chapter compares strains from Central America with strains introduced into the United States. I show that the introduced pathogen population is poor at infecting an ancestral host. I also found a suite of genes in the introduced population associated with the host jump. Finally, I dive into disease progression in one host species grown in Northern California, *Vitis vinifera*. Over the course of three years, I documented how disease development can vary among grapevine cultivars. This final chapter provides suggestions for disease management through more accurate detection *of X. fastidiosa* in an agricultural setting. The overarching goal of these works is to improve our understanding of the determinants of host specificity and disease progression of the economically and environmentally important pathogen species *Xylella fastidiosa.*

DEDICATION

This dissertation is dedicated to my family who inspire me to try to take care of the world while also taking care of myself.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

As I lay awake at night thinking about what to include in this section I started to cry. There are so many souls who have contributed to getting me and this work to where I am today. I feel overwhelmingly grateful to those who have held out a hand to me along my path.

To start with my family, without whom I would not exist, but more so who have loved me with all their hearts; my parents Sheri and Asher, my partner Kirk, my brother Zach, sister-in-law Jody, niblings Lev and Dylan, and my dearest Sophie. My fabulous cousins, aunts and uncles who inspire me and have offered me a seemingly infinite number of places around the country where I can call home because they are there to welcome me. My dear friends: from New York, I am so lucky to have found so many fierce, kind, and quirky people from day 1. From Oberlin, the ones I got to grow up with, I am so lucky to have been able to do it with you. From Berkeley, although I might not really be a California person, I have been so lucky to meet people who have come here to enjoy the beauty and the vegetables. I am so grateful to have friends who have helped me to push the conceived boundaries of what a life should and can look like.

From a more professional perspective, I want to thank my Advisor Rodrigo for not only all of the teaching and support but for helping cultivate a community that has been supportive and fun throughout my time here. I feel lucky to have been able to count on you these last six years. To all of my lab mates, past and present, what a pleasure to sit next you all every day and get to smell your microwaved fish, hear the metal seeping from your headphones, make infinite coffees, and get excited to come to work every day to spend time together. I have been so lucky to be part of such a fantastic community. Thank you to my dissertation committee, Britt Koskella and Rachel Brem for your thoughtful suggestions on my research and excellent teaching. While I won't thank every teacher that I ever had, I certainly would not have gotten to where I am now with Professor Yolanda Cruz, who gave me the push and the opportunities to find my groove in the sciences.

Finally, I want to thank all of the collaborators for this work, my coauthors, as well as those I have worked with on other projects. Your work has been educational, fun, inspiring, and most of all has offered me insights into the type of future I want for myself, which is more challenging to imagine than any paper.

Chapter 1
We acknowledge and thank Andreina Castillo Siri, Mathieu Vanhove, Adam Zeilinger, Michael Voeltz, Dylan Beal, Anne Sicard, Aidee Guzman, and Kirk Pearson for helpful discussions, ideas, edits, and support. We also thank Monica Donegan, Elizabeth Clark, and Isabel Bojanini who have helped with final revisions of the manuscript. This work was supported by the California Department of Food and Agriculture PD/ GWSS Research Program and the European Union's Horizon 2020 research program (XF-ACTORS). A.K. received support from the UC Berkeley William Carroll Smith Plant Pathology Fellowship.

Chapter 2
We acknowledge and thank Monica Donegan, Isabel Bojanini, Andrea Brown, Karina Gadkari, Giovanna Figueroa, Liz Clark, Patrick Lee, Karyn Newman, Christina Wistrom and the Oxford Tract greenhouse staff, Juan Medrano, and Marta Matvienko for their contributions to the greenhouse work and conversations about this paper. We also thank Frinj for their generous donation of coffee plants to support this work. A.K. received support from the UC Berkeley William Carroll Smith Plant Pathology Fellowship, the ESPM Spring 2023 Research Fellowship, and the ESPM Summer 2023 Research Funding Grant.

Chapter 3
We acknowledge and thank Michael Voeltz, Amielia Adams, Karina Gadkari, Emily Donaldson, Clarisse Dantas Lima, Juliet Pavy, Akila Ganpathi, Andreina Castillo Siri, Kirk Pearson, Sheri Katz, and Dylan Beal for contributions to the field, lab, and editing work for this project. This work was supported by the California Department of Food and Agriculture PD/GWSS Research Program and the European Union's Horizon 2020 research program (XF-ACTORS). A.K. received support from the UC Berkeley William Carroll Smith Plant Pathology Fellowship. A.S. received support from a Marie Skłodowska-Curie Fellowship (European Union's Horizon 2020 Research and Innovation Program, grant agreement 707013).

INTRODUCTION

Pathogen outbreaks due to international trade and environmental degradation have become a theme of our daily lives. From a management perspective, scientists try to distill and disseminate information that will be sufficiently protective without being overwhelmingly complex. In the world of the bacterial plant pathogen *Xylella fastidiosa*, we are still struggling to find that balance. From a policy perspective, we need to draw lines in the sand in regard to naming pathogens that are dangerous to particular crops in particular regions. But where do we draw them? This dissertation is focused understanding these ambiguities. I attempt to address questions such as:

- Which crops have likely broad host ranges or have been repeatedly infected with diverse pathogen strains?
- Are strains adapted to particular host plants after new introduction events?
- Which symptoms after a new infection will be common among several grape cultivars and which will be unique to one?

Science and policy can sometimes feel like two magnets affixed to strings—dancing between cohesion and dissonance. The questions that I focused on here are my attempt at unifying the evolutionary process of host specificity and its complex real-world implications.

The theme of my dissertation is the dynamic relationship between a bacterial pathogen, *X. fastidiosa*, and its various plant hosts. The work encompassed in my dissertation is a combination of computational work focusing on whole genomes of bacteria, greenhouse infections of plants, and field studies in Napa Valley. This has allowed me to approach this system from a multiplicity of perspectives and learn many methods that will strengthen my skills as a researcher. *X. fastidiosa* is a bacterial plant pathogen found across the Americas and more recently, worldwide. The bacterial species is able to infect at least 679 plant species belonging to 82 botanical families (Delbianco *et al.*, 2023). While the species *X. fastidiosa* infects many plants, particular strains have a narrower plant specificity (Sicard *et al.*, 2018; Nunney *et al.*, 2019). My work all circulates around this distinction.

Chapter 1, Phylogenetics of Historical Host Switches in a Bacterial Plant Pathogen, was published in *Applied and Environmental Microbiology* on March 21, 2022 (https://doi.org/10.1128/aem.02356-21). This chapter focuses on the genetic relationships of a set of 349 whole *Xylella* genomes that were sequenced largely by the Almeida lab over the last several years. Using a set of phylogenetic techniques, I compared how these different techniques recreated the ancestral relationships between these strains. The strains are very diverse, as they were isolated from ten countries and 32 plant species. I used this distribution of contemporary hosts to predict the most likely hosts at various important historical timepoints, for example, what the most likely host would have been at the time of an introduction of these pathogens from Central America into California. I also identified genes that have significant connections to strains isolated from a particular type of plant, allowing us to detect convergent evolution, where two organisms separately evolve similar characteristics to survive in similar environments.

Chapter 2, A Pathogen of Good Taste: a Bacterial Host Shift from Coffee to Wine Grapes, is not yet published. Chapter two is a project of intermediate host scale: the analysis of the introduction event from Central America into the United States. This chapter seeks to address the host adaptation of the introduced strains from their former hosts to a common U.S. host--*Vitis vinifera*--through a combination of greenhouse and computational methods. In this paper I experimentally tested the introduced ability of the introduced strain to infect an ancestral host, *Coffee arabica*, to elucidate the process of a documented and economically important host jump. Using a set of 15 whole genome sequences from Costa Rica and 292 whole genome sequences from the introduced clade, 232 isolated from grapevines and 60 from alternative hosts, I tested for traces of adaptation to *Vitis spp.* I demonstrate that the introduced strains do not, for the most part, persistently infect *C. arabica*. Furthermore, by testing the hypothesis of hypervirulence post-introduction, I did not observe an overall increase in the virulence of the introduced strains towards coffee. I also found both genes and single nucleotide polymorphisms (SNPs) that are associated with the host shift to *Vitis.* These results support the hypothesis that genetic adaptation to a novel host occurred following this pathogen's introduction.

Chapter 3, Progression of *Xylella fastidiosa* infection in grapevines under field conditions, was published in *Phytopathology* in 2023. In this chapter, I performed an analysis of a field infection experiment in a vineyard in Napa Valley that was conducted in 2017-2019. This pathogen is of particular interest in California, where it causes Pierce's disease of grapevine. This pathogen was introduced to California in the late 1800s (Vanhove *et al.*, 2020), but our study is the first to document the progression of symptoms in field conditions of mature commercial plants. This is largely due to the exceptional collaborations with the University of California Cooperative Extension Team, who have worked hard to build trust between the growers and researchers, allowing us to infect vines in a working vineyard, which is unprecedented using such high-value crops. Forty-five experimental vines were infected with *X. fastidiosa* in Napa County in April of 2017, and symptoms and bacterial populations were quantified throughout the growing seasons of 2017-2019. Our findings identify previously undetected initial symptoms significantly associated with bacterial populations at the inoculation site. Symptoms were also observable earlier than expected, and pathogen spread through the plants was tracked. The number of plants that recovered during the winter was also much higher than expected, resulting in vital implications for grape growers in California.

The results of these three projects are useful from both fundamental and applied science perspectives. In the first chapter I identified hosts that are likely susceptible to *X. fastidiosa* infections by a diverse set of strains. I also showed substantial differences between the results of different phylogenetic methods, demonstrating the importance of whole genome sequencing for taxonomy. In chapter 2, I tested the susceptibility of coffee to introduced strains isolated from grapevines, and showed that the introduced strains are not highly virulent towards this important ancestral host. Through this work we gained insight into the undescribed process of a host jump in *X. fastidiosa*. In both chapters 1 and 2 I also identified target genes that are correlated with host specificity, allowing for later research that functionally tests these genes to broaden our understanding of the genetics of host specificity in this system. In the final chapter, I described the progression of Pierce's disease of grapevines in field conditions and identified important symptoms for improved disease detection. Diseases caused by the pathogen *X. fastidiosa* can be economically, culturally, and environmentally devastating. My hope is that this

dissertation will add a small piece to the larger puzzle as we attempt to both understand and control the damage caused by *X. fastidiosa.*

All coauthors for the three papers have given their approval for the inclusion of these works in my doctoral dissertation.

CHAPTER ONE

Phylogenetics of Historical Host Switches in a Bacterial Plant Pathogen

Alexandra Kahn[1], Rodrigo P. P. Almeida[1]
[1]Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA 94720

Abstract
Xylella fastidiosa is an insect-transmitted bacterial plant pathogen found across the Americas and more recently, worldwide. *X. fastidiosa* infects plants of at least 563 species belonging to 82 botanical families. While the species *X. fastidiosa* infects many plants, particular strains have increased plant specificity. Understanding the molecular underpinnings of plant host specificity in *X. fastidiosa* is vital for predicting host shifts and epidemics. While there may exist multiple genetic determinants of host range in *X. fastidiosa*, the drivers of the unique relationships between *X. fastidiosa* and its hosts should be elucidated. Our objective with this study was to predict the ancestral plant hosts of this pathogen using phylogenetic and genomic methods based on a large dataset of pathogen whole genome data from agricultural hosts. We used genomic data to construct maximum likelihood (ML) phylogenetic trees of subsets of the core and pan-genomes. With those trees, we ran ML ancestral state reconstructions of plant host at two taxonomic scales (genus and multi-order clade). Both the core and pan-genomes were informative in terms of predicting ancestral host state, giving new insight into the history of the plant hosts of *X. fastidiosa*. Subsequently, gene gain and loss in the pan-genome was found to be significantly correlated with plant host through genes that had statistically significant associations with particular hosts.

Importance
Xylella fastidiosa is a globally important bacterial plant pathogen with many hosts, however, the underpinnings of host specificity are not known. This paper contains important findings about the usage of phylogenetics to understand the history of host specificity in this bacterial species, as well as convergent evolution in the pan-genome. There are strong signals of historical host range that give us insights into the history of this pathogen, and its various invasions. The data from this paper are relevant in making decisions for quarantine and eradication, as they show the historical trends of host switching, which can help us predict likely future host shifts. We also demonstrate that using MLST genes in this system, which is still a commonly used process for policymaking, does not reconstruct the same phylogenetic topology as whole genome data.

Glossary of terms

Core-genome: only the genes present in **all** genomes that are included in a study
Pan-genome: all genes present in any genomes that are included in a study
Accessory genome: all genes from the pan-genome that are not part of the core-genome
Ancestral state reconstruction: the historical inference of character traits based on modern distribution
Characters: heritable traits that can be compared across organisms in a matrix that can be used to make a hypothesis of phylogenetic relationships
Phylogeny: a hypothesis of the evolutionary relatedness of a set of organisms
Horizontal Gene Transfer/Bacterial Recombination: DNA transfer between organisms not via reproduction, in this paper only referring to bacteria. Typically mediated by transformation, transduction, and conjugation.
Pathogenicity: The ability of a pathogen to cause disease on a particular host

Keywords: *Xylella fastidiosa*, Host specificity, Quarantine, Trade, Policymaking, Ancestral state reconstruction, Genomic diversity, pan-genome, Phylogenomics

Introduction

Modern plant trade disturbs historical ecological relationships and creates opportunities for the development of novel pathogenic interactions (Lockwood *et al.*, 2005; Hulme, 2009), often with correlated genetic changes (Levy *et al.*, 2018). However, pathogens must be adapted to the environment of the novel host before they meet, or they will not be able to survive and reproduce (Woolhouse *et al.*, 2005). That does not mean pathogens necessarily pre-adapted to the exact same host, but either could have adapted to a similar host earlier and retained that adaptation until encountering a novel host. Convergent evolution in diverse pathogen populations can allow for divergent strains to have the ability to infect the same hosts. Three potential mechanisms of genetic change that can accompany host shifts are nucleotide changes leading to different alleles in the core genome of a pathogen (defined as the genes shared by all strains in a set of samples), whole gene gain and loss in the pan-genome, leading to unique sets of genes in individual strains, or regulatory/epigenetic changes. Due to the recent increase in whole genome sequencing of plant pathogens, we can now more effectively use phylogenetic analyses to investigate their genetic associations to both novel and historical host plants (Huang *et al.*, 2020). Understanding the phylogenetic relationships between specific host and pathogens should allow the development of preemptive plans to protect natural ecosystems as well as agriculture from the emergence of novel pathogens.

Xylella fastidiosa is an insect-transmitted, xylem-limited bacterial plant pathogen found across the Americas, and as of recently, globally. *X. fastidiosa* is considered to be a generalist pathogen, because, as a species; it reportedly infects at least 563 species belonging to 82 botanical families (Food & Authority, 2018). The lack of host specificity that *X. fastidiosa* exhibits as a species contrasts with increased plant host specificity in smaller clades and strains (Purcell *et al.*, 2003; Almeida *et al.*, 2008; Barrett *et al.*, 2009; Sanderlin, 2017; Food & Authority, 2018; Sicard *et al.*, 2018). It is still debated whether a pathogen like *X. fastidiosa* should be considered a generalist species that "leaps" between phylogenetically distant hosts or, alternatively, a crawler at shallower clades (Park *et al.*, 2018; Sicard *et al.*, 2018). The difference is biological as there are unique implications for either evolutionary path. *X. fastidiosa* could be repeatedly evolving specialization or it could have biological and genetic traits as a species that make particular hosts of disparate plant taxa suitable.

From an applied perspective, there have been recent calls from government agencies for increased focus on understanding the host range of *X. fastidiosa.* This is because the pathogen has been deemed likely to spread and to be of extremely high risk to crops of agricultural value (EFSA Panel on Plant Health (PLH), 2015). *Xylella fastidiosa* causes disease in a range of high-value crops, including Pierce's disease of grapevines, citrus variegated chlorosis disease in sweet oranges, almond leaf scorch, leaf scorch of coffee, olive quick decline syndrome (OQDS), spanning North and South America, Europe, the Middle East, and Taiwan (Sicard *et al.*, 2018; Amanifar *et al.*, 2019; Coletta-Filho *et al.*, 2020)**.** While there are three distinctive subspecies of *X. fastidiosa,* and it would be desirable to be able to use those subspecies for management decisions, so far the subspecies have not been found to have sufficient resolution to define host range or to infer risk (Sicard *et al.*, 2018). Understanding the molecular basis of plant host specificity in *X. fastidiosa* is vital for predicting and acting upon host shifts, but these are processes yet to be described (Sicard *et al.*, 2018).

Xylella fastidiosa is a member of the group Xanthomonadaceae, and phylogenetically clusters sister to *Xanthomonas albilineans*, technically within the paraphyletic genus *Xanthomonas*, although *Xylella* is considered a separate genus (Szurek *et al.*, 2009; Rodriguez-R

et al., 2012). *Xylella spp.*, and *Xanthomonas albilineans*, are the only xylem limited Xanthomonadaceae and have convergently reduced genomes compared to the rest of the genus (Szurek *et al.*, 2009). *Xylella* also lacks a Type III Secretion system (T3SS), a loss compared to its higher order taxonomic group. As the purpose of the T3SS in phytopathogens is to deliver effectors into living plant cells (Galán & Collmer, 1999) the loss has been hypothesized to be due to *X. fastidiosa* primarily interacting with non-living tissue; insect cuticle and mature xylem vessels (Chatterjee *et al.*, 2008).

While the molecular basis of host range is not understood, there are consistent patterns in the ability of particular *X. fastidiosa* isolates to infect specific plant hosts regardless of their environmental condition (Almeida *et al.*, 2008; Nunney *et al.*, 2019). This implies that genetics, as opposed to only environmental conditions, underlie the relationship between isolates and plant hosts that allow for colonization. Recurring pathogen specificity to a particular host can be either explained through phylogenetic signal, where members of a clade have shared traits that allow for pathogenesis in that host, or by pathological convergence, where more distantly related strains have separately acquired mechanisms for virulence. Both processes have underlying genetics but each shows different phylogenetic patterns (Hajri *et al.*, 2009). Lastly, we have seen that deletion of *rpfF*, which controls cell-cell signaling via a diffusible signal factor (DSF), can expand the host range of *X. fastidiosa* (Killiny & Almeida, 2011). Other insights into host range have been made in terms of plant immunological studies. For example, removing the O-antigen from the exterior of *X. fastidiosa* cells allows the plant to quickly recognize *X. fastidiosa* and initiate immune responses, thus decreasing its likelihood of colonization of the plant (Rapicavoli *et al.*, 2018). O-antigens are highly variable and evolve rapidly, and often are shown to have co-evolutionary histories between symbiotic organisms as they are the first exposed part of any bacterium (Zipfel & Felix, 2005). In terms of phylogenetic methods, cophylogenies have shown no cospeciation between plant hosts and *X. fastidiosa* or any other congruence between the evolutionary histories of *X. fastidiosa* and its plant hosts (Sicard *et al.*, 2018). Based on the current data, it is not generally possible to tell if *X. fastidiosa* is undergoing host jumps or range expansions, however the data available so far suggests that both are occurring given that in certain situations we see strains able to infect multiple hosts (Nunney *et al.*, 2019) while in other situations we see multiple strains co-existing in nature but no cross infections of hosts (Almeida *et al.*, 2008).

Using the influx of whole genome data generated in the past several years, we searched the genomes of *X. fastidiosa* for correlations with plant host species. The first method we pursued was conducting ancestral state reconstructions. Ancestral state reconstructions use genetic data (phylogenies), with a known phenotype for each taxon, to characterize the most likely state that each ancestral node of the tree would have possessed for the phenotype of interest. This tool has been used to understand host-pathogen interactions via ancestral state reconstructions in fungi and trematodes parasite systems (Razo-Mendivil & Pérez-Ponce de León, 2011; Navaud *et al.*, 2018). Ideally, we would be able to ask: what was the most likely ancestral host of the ancestor of all *X. fastidiosa*? If we can understand patterns in the past, it can help us better build models to predict future hosts based on the genomic changes associated with historical host shifts. Following the ancestral state reconstructions, we looked further into the pan-genome by calculating correlations between plant host types and the presence/absence of each gene.

This study aimed to compare the commonly used genetic datasets available for phylogenetic analyses of *X. fastidiosa* both to compare phylogenetic topologies as well as ancestral host states from each dataset. We hypothesized that the pathogen phylogeny would be correlated with host

history and that we could observe this trend through ancestral state reconstruction. If there is no relationship between host and the phylogeny, there should not be conclusive ancestral state reconstruction results. We hypothesize that by using either the core genome of *X. fastidiosa*, pan-genome phylogenetic tree, or both, it would be possible to estimate the likelihood of hypothetical plant hosts for ancestral nodes of interest (a node represents a common ancestor of the tips). This would show that the host is largely dependent and predictable based on the phylogeny of bacterial relationships and would lead to further pursuing allelic differences in core-genome and/or gene gain/loss in the pan-genome and estimate how either or both are correlated with plant host identity. While not biologically meaningful, since MLST data is still frequently used in *X. fastidiosa* management, we included that datatype in our analysis for comparison as well.

Results

Phylogenetic reconstruction of disparate regions and sizes are topologically similar

The pan-genome of all sequences and the outgroup *X. taiwanensis* - Wufong1_PLS229 (n = 349) contained 17,024 genes (14,564 of which come from the ingroup *X. fastidiosa*). The alignment of MLST genes totaled 4,146 bp in length, while the core genome comprised 1,411 concatenated regions in a total of 354,816 bp. Non-recombinant regions identified with ClonalFrameML (Didelot & Wilson, 2015) comprised only 32% of the core genome (68% of the alignment showed evidence of recombination), leaving an alignment consisting of only 112,819 base pairs (See Table 1). The alignment contained 130 pairs of sequences that were completely identical to each other, highly reducing the amount of within subspecies differentiation that is possible with this dataset and creating large polytomies of indistinguishable sequences within subspecies *fastidiosa* (mostly California *Vitis* samples) as well as within subsp. *pauca* (mostly Italian *Olea* samples). Due to this lack of within-subspecies resolution, the phylogeny with recombinant regions removed is only suitable for between subspecies comparisons due to the extensive data loss in removing recombinant regions. The strains and locations in the alignment with recombination can be visualized in the supplemental materials (Figure S3).

While between subspecies topologies are similar among the four trees generated, they are not identical. The core genome tree shows consensus of taxonomic division into three subspecies, however, subsp. *sandyi* and *morus* could be either part of subsp. *fastidiosa* or each their own small subspecies without affecting the monophyly of subsp. *fastidiosa*. (See Fig.1 and Fig. 2 for phylogenetics, and Table S1 for strain information). The non-recombinant tree is similar except that subsp. *morus* is clustered within subsp. *fastidiosa.* The pan-genome splits the most basal of the three subspecies, subsp. *pauca,* into a paraphyletic cluster, however places *multiplex, fastidiosa, morus, and sandyi* similarly to the core phylogeny (see Fig. 1). The MLST tree shows subsp. *morus* as the outgroup to subsp. *fastidiosa* while subsp. *sandyi* falls within subsp. *fastidiosa.* The other difference among the four topologically similar trees is variation in branch length. The phylogenetic diversities were calculated as summed length of each tree calculated from nodes to root were Core = 6.65, Non-recombinant = 3.32, MLST = 8.65 (in substitutions per site), and pan-genome = 59.15 (in gene gains and losses per site). Since the pan-genome tree was built with gene presence/absence data, it was calculated in gene changes per site. Phylogeny and alignment information is summarized in Table 1. A 16S rDNA phylogeny was also built as a comparison (See Figure S5), but the phylogeny provided very poor differentiation among strains (only 40 unique sequences out of the 349 strains).

Within subspecies *fastidiosa*, the core genome rearticulates the three PD clades that were found in Castillo et al. (Castillo *et al.*, 2021b) (See Supplemental Fig. 4). Within the clade defined as PD-III, the sequence similarity in the core has led to extensive polytomies, with many sequences indistinguishable in the core (Fig. 2). The three PD clades are also articulated in the non-recombinant phylogeny, and the pan-genome phylogeny, however the MLST tree does not differentiate these clades from one another. Not poorly resolved, the MLST does have high bootstrap support for clades that conflict with trees constructed with core and pan genome trees, suggesting that using MLST genes has the potential to subvert the analysis of relationships between taxa, while showing strong bootstrap support.

Within subspecies *multiplex*, there have typically been considered two groups, the non-recombining "non-IHR", *multiplex*, as well as the recombining outgroup "IHR" *multiplex* (Landa *et al.*, 2019). The core genome tree as well as the MLST tree both articulate these two groups, the clade "non-IHR", as well as the non-monophyletic recombining group, "IHR". The non-recombinant tree and the pan-genome tree do not re-create these groupings (See Fig. S4).

All phylogenies but the pan genome show a consistent split in subsp. *pauca* between the strains isolated from the Italian OQDS outbreak and the mixed host strains from Brazil. Within the OQDS strains, as well as several very closely related strains from Costa Rica, there is no clear resolution at this genomic scale. Within the Brazilian clade, strain Hib4 is the outgroup in all phylogenies except the MLST. subsp.

The reconstructed ancestral likelihoods suggest ancestral hosts of X. fastidiosa

Interrogating the results of the ancestral state reconstruction to the genus level of the core-genome phylogeny shows undetermined hosts at the deepest nodes (See Fig. 3). However, the ancestral node of the subspecies *fastidiosa* has a significant association with the plant genus *Coffea*, which persists throughout subspecies *fastidiosa* as the most likely ancestral host for all strains isolated from South and Central America. This changes for the Pierce's disease of grapevines clade, where the ancestral host of all nodes except one is *Vitis*, the one exception being an ancestral *Prunus* node. Subspecies *sandyi* and *morus* are undetermined in ancestral hosts. Subspecies *multiplex* has a more dynamic history, with *Vaccinium* shown to be the most likely ancestral host for the subspecies, and then within the clade, a switch to a large group of nodes whose most likely host in *Prunus*, as well as two nodes depicting *Platanus* and *Olea*. Subspecies *pauca* does not have a determined ancestral host of the whole subspecies, and internal nodes switch several times between *Citrus* and *Coffea*, and once to *Olea*.

In terms of the genera across the reconstructions, while the deep nodes (ancestors of a subspecies), are often undetermined, there is more resolution within subspecies (See Fig. 4). The node that is consistent across the four reconstructions is that there is a high likelihood of the genus *Coffea* being the ancestral host of the node representing the introduction of subsp. *fastidiosa* from Central to North America. The genus *Vaccinium* was predicted as the most likely ancestral host of subsp. *multiplex* in the core genome phylogeny, whereas in the non-recombinant phylogeny, the ancestor of all but one strain of subsp. *multiplex* is the genus *Prunus*. All four trees agree upon the ancestor of the internal "non-IHR" multiplex clade being *Prunus*. In terms of the transition models chosen for each reconstruction, most trees had lower AIC scores when using the equal rates model with fewer parameters than the symmetrical rates model, the exception being for the pan-genome super order reconstruction having a lower AIC score with the symmetric model than the equal rates (See Table S2).

At the node representing the ancestor of the species *X. fastidiosa*, both the non-recombinant core and pan-genome phylogenies predict that the clade Rosid is the most likely ancestral host (See Fig. 5). The core and MLST phylogenies predict Asterid to be the ancestral host, but at lower likelihoods of 87 and 78%, respectively, which are visualized, along with all likelihoods under 95%, as undetermined (See Fig. 5 & Table S2). There is enough discordance between reconstructions a consistent pattern at this host depth is unlikely.

Four plant clades correlated with gene presence and/or absence

Bacteria isolated from the genera *Coffea* and *Vitis* as well as the super-orders Asterid and Rosid have *X. fastidiosa* genes with which they are significantly correlated; totaling 30 genes (See Table 2). Ten of these 30 genes are significantly correlated with both Asterids and Rosids, with paired, opposite relationships (i.e., the same gene is significantly absent for one host, while present for another) (See Table 2). Some correlations are of significance due to elevated presence of the gene among strains found in a particular host, while most are significant due to an absence of particular genes in the host of interest. Since lineage-specific interdependencies are accounted for with the phylogeny, the correlated genes are representative of convergent processes, either evolutionarily or via lateral gene transfer, not shared ancestry by descent. Genes that are significant mark repeated non-vertical descent changes in the pan-genome of strains in convergent patterns specific to the hosts of interest. While most identified genes are hypothetical proteins, genes shown to be correlated with host were *fitB_1* (part of Toxin-Antitoxin (TA) system, involved in in-host migration), *vbhT* (Part of TA system, interbacterial effector protein), *socA* (antitoxin to SocB, which inhibits DNA replication), and a HTH-type transcriptional regulator (others known in *X. fastidiosa* to modulate biofilm formation) (see Table 2)(Barbosa & Benedetti, 2007; Aakre *et al.*, 2013; Harms *et al.*, 2017).

Discussion

In this paper, we show that there is a genetic basis to the host range of *X. fastidiosa.* We demonstrate that both the phylogeny and gene gain and loss in the pan-genome are connected to plant host of the diverse species *X. fastidiosa*, and that an Asterid of undetermined genus was the most likely ancestral plant host of *X. fastidiosa*. Our results indicate that the evolutionary trajectories of both the core and the pan-genomes allow for a bacterial species with an extensive host range to specialize many times over a broad array of plant hosts. We see this system as an example of one that "leaps", with host genera seemingly changing not via phylogenetic signal to related plant hosts, but switching across large regions of plant host phylogenies (Park *et al.*, 2018). Prior to this study, we have not been able to trace a pattern of underlying genetic origins of host specificity in *X. fastidiosa*. In this way, our study shows that the phylogeny and gene gain/loss are connected to the adaptations that diversify host specificity in *X. fastidiosa.*

Phylogenies for MLST, pan-genome, core-genome, and non-recombinant core-genome data were topologically similar, but not identical. While the subspecies relationships are not important to predicting host range, they are frequently used in management decisions and our ability to converse about outbreaks, so we are including our findings alongside our data on host use. In terms of taxonomic subspecies, there are differences between the four trees in whether the two debated subspecies, *X. fastidiosa* subsp. *morus* and subsp. *sandyi*, are contained within subsp. *fastidiosa* or subsp. *multiplex*, or if they should be considered their own subspecies. While there are pairs of strains that are consistently close to each other like the *morus* strains MulMD and Mul0034, the uncertainty in their position from phylogeny to phylogeny likely reflects large

gaps in diversity that we have not yet sequenced or horizontal gene transfer more intensely affecting the pan-genome and particular genes used for MLST than the core genes, leading to issues recreating the vertical descent we aim for in a phylogeny (See Fig. 1). The subspecies *morus* has been documented to have up to 15.30% of its core genome undergoing inter-subspecies homologous recombination, which could account for its uncertain placement in the four phylogenies (Vanhove *et al.*, 2019). The two strains that have been described as *sandyi*-like, CO33 and CFBP8356, both clustered within subspecies *fastidiosa*, not with the other potential *sandyi* strains Ann1 and RAAR8_XF70, supporting previous work showing that there is not a strong distinction between subspecies *sandyi* and *fastidiosa* (see Fig. 2) (Denancé *et al.*, 2019). The core genome tree also has very low bootstrap support for subsp. *pauca*, which is the most diverse and oldest of the three main subspecies that could be potentially due to conflicting histories between horizontal and vertical descent, or alternatively reflect that this group is simply not well supported as one subspecies (Felsenstein, 1985). In terms of the poor resolution in the OQDS clade, an analysis has recently been conducted to increase resolution within these strains (Sicard *et al.*, 2021). Given the diversity of subspecies *pauca*, the Hib4 strain, the outgroup of the subspecies, could be a potentially interesting strain in terms of both function and evolutionary history (Vanhove *et al.*, 2019).

It is difficult to know which phylogenies are more accurate than others, however we assume that the core genome is the most accurate at depicting the descent of this bacterial species and the topology should be robust to even high levels of recombination (Hedge & Wilson, 2014). While the non-recombinant core-genome might reduce some issues with horizontal gene transfer, the lack of resolution because of too many identical sequences makes it difficult to use. While more data are not intrinsically better, there are known issues with the MLST genes used for *X. fastidiosa* phylogenetics and having a larger set of unbiased homologous regions should be able to lend data to support nodes that are difficult to differentiate using the smaller MLST dataset (Landa *et al.*, 2019).

Using the core genome phylogeny, the most likely ancestral host was inferred from the phylogeny. These results show us that the phylogenetic history of *X. fastidiosa* is significantly correlated with the agricultural plant host that the strains were isolated from. While the core-genome phylogeny depicts mainly vertical descent within this bacterial species, the pan-genome phylogeny likely combines vertical descent with horizontal gene transfer. This is due to the pan-genome's inclusion of the accessory genome, which are genes not shared by all members of the group (Soucy *et al.*, 2015). Based on this, we speculate that there is both adaptation and convergence depicted in these results. Potentially, both convergent horizontal descent via gene gain and loss as well as vertical descent in the core leads to our modern distribution of traits. While the ancestral state reconstruction did not show a classic host-parasite story of cospeciating or phylogenetically conserved host specificity, the phylogeny and gene presence/absence are predictive of the hosts from which the strains were isolated from, and thus hypothetically, host specificity as well.

While the four ancestral state reconstructions do not show identical histories, they all infer high likelihood of ancestral hosts at many key branch points of the three subspecies. The pan and core-genome reconstructions predict the genus *Vaccinium* (based on isolates from blueberry) as the most likely ancestral host of the subspecies *multiplex,* which supports the overall reliability of the reconstruction as blueberry, like subsp. *multiplex*, is native to the eastern North America (Wood, 2004). Subsp. *pauca*, subsp. *multiplex*, and subsp. *fastidiosa* all exhibit host shifts from another genus to *Prunus*, suggesting potential for increased vulnerability in this genus to

infection from varied alternative hosts. All four reconstructions also support the genus *Coffea* as the most likely ancestor of the introduced subsp. *fastidiosa* strains from Central American to California. This supports a previous hypothesis made by Nunney et al. (Nunney *et al.*, 2010) wherein coffee plants that were imported from Central America to southern California in the mid 1800s might have brought *X. fastidiosa* subsp. *fastidiosa* along with them. Given the potential role of imported *Coffea* in devastating global outbreaks of disease caused by *X. fastidiosa* (California and Italy) (Marcelletti & Scortichini, 2016), it should be much more carefully monitored or restricted in global trade. Given the current policy emphasis on eradication, trade restrictions, it is vital to identify genera such as *Coffea* that are especially relevant to global outbreaks and that should be monitored carefully. The relationship between *X. fastidiosa* and *Coffea* should be further explored as a model host to aid our understanding of the molecular mechanisms of this complex interaction. A potential alternative hypothesis for these nodes could also be that *Coffea* and *Vaccinium* are permissive hosts. From a parsimony perspective, they could be akin to 'universal hosts' so that it takes very little change for *X. fastidiosa* strains to switch to *Coffea* or *Vaccinium* from other infected plants. This could be investigated by further interrogating the genes shown to be uniquely absent in *Coffea*-infecting strains. Phylogenetically, this would reflect deep homology in which the underlying genetic framework of the pathogens make it easy to shift from other plant hosts to *Coffea* or *Vaccinium* (Shubin *et al.*, 2009).

The two plant genera with genes significantly correlated with them, *Vitis* and *Coffea,* had 179 and 20 whole genome sequences from diverse sampling regions. The larger clades of Proteales, Asterid, and Rosid were also used to look for convergent gene presence absence and again the two groups with the majority of samples, Asterid (n = 126) and Rosid (n = 194) had genes correlated with them, while Proteales (n = 2) did not. The genes found to be correlated with these host groupings had varied functions. Unfortunately, out of these 23 genes, 20 are hypothetical proteins, the ones with known functions could have very interesting implications for host range. *fitB_1* has been known to be involved in in-host migration and metal binding, similar genes are also frequently gained and lost in other *Xanthomondaceae* and are hypothesized to affect both gene regulation and resistance mechanisms (Martins *et al.*, 2016). *vhbT* is an interbacterial effector protein, facilitating bacterial conjugation, another process with potential for large genomic and functional changes (Harms *et al.*, 2017). Another significant gene (group_2780) contains a helix-turn-helix region, a DNA binding-domain that has been found to control metal resistance bacteria generally and biofilm growth in *X. fastidiosa* specifically (Barbosa & Benedetti, 2007). These genes should be explored further through fitness tests with the presence and absence of these non-essential accessory genes in multiple host environments to further evaluate if their presence and absence is adaptive or due to drift.

Future research pertaining to host range should focus on both convergent gene gain and loss as well as the adaptive vertically descended genetics underlying host range. As both genomic assays have identified the pan-genome to be linked to host association, it would be beneficial to our understanding of host specificity to pursue this further. This study has identified a group of candidate genes associated with particular hosts, and they can be tested in the lab to determine if they are significantly linked to fitness in their particular hosts. The study has also identified *Coffea* as an especially relevant host in global plant trade in terms of spreading infection across borders and oceans. Using these data, we can start identifying patterns of likely host shifts that can help make decisions on when eradication and quarantine is necessary based on the historical likelihood of host shifts. However, we should also carry out further whole genome sequencing of strains outside of the classic agricultural settings. To truly understand a

biological system, we not only need to understand the relevant biological components but also how they interact both inside and outside of agricultural landscapes.

Materials and Methods

Whole genome sequence dataset

A total of 349 *Xylella* spp. genome sequences were used in this study, either downloaded from GenBank (Benson *et al.*, 2010) or assembled *de novo* in house from published FASTQ reads (see Table S1). *De novo* sequences were aligned as described in Castillo et al. (2020), and contigs were mapped to complete genomes of each subspecies using Mauve's contig mover function (Rissman *et al.*, 2009). No novel genomes are being presented in this study. *X. fastidiosa* subsp. *fastidiosa* scaffolds were reordered using the Temecula1 assembly (GCA_000007245.1), *X. fastidiosa* subsp. *pauca* scaffolds were mapped using the 9a5c assembly (ASM672v1), and *X. fastidiosa* subsp. *multiplex* scaffolds were reordered using the sequence of strain M23. Draft genomes as well as downloaded sequences were then annotated using Prokka (Seemann, 2014).

Phylogenetics

The first step in creating all phylogenies was building a nucleotide or gene alignment of the genomic regions of interest. Four alignments were created all using the same set of taxa (see Table S1): a core genome alignment, non-recombinant core alignment, a multilocus sequence type (MLST) alignment, and a pan-genome alignment.

The core genome was built with Roary (Page *et al.*, 2015) to identify nucleotide regions (genes or hypothetical proteins) shared by at least 99% of all taxa. We ran Roary with the parameters -s -ap to cluster paralogs and allow them in the core genome. The non-recombinant core alignment was based on the core genome, but recombinant sites identified with ClonalFrameML were removed from the alignment using an in-house R script (Didelot & Wilson, 2015). The MLST alignment was based on a nucleotide alignment of the 7 MLST housekeeping genes commonly used for *X. fastidiosa* (*petC, nuoL, malF, leuA, holC, gltT,* and *cysG*) with reference sequences acquired from the *X. fastidiosa* MLST database (Jolley *et al.*, 2004; Scally *et al.*, 2005). We then searched each MLST reference sequence against all whole genomes (see S1) using the Basic Local Alignment Search Tool (BLAST) at an E value of 10^{-3} in BLAST +, with a database created for each whole genome (Camacho *et al.*, 2009). We concatenated all MLST gene sequences for individual taxa and aligned them to all other taxa using MAFFT v. 7 (Katoh *et al.*, 2005). The pan-genome alignment was made using Roary's gene presence-absence output by constructing a matrix of all genes as characters with binary presence or absence of that gene in a strain as the character state. As each character represented a known genetic region and there were no gaps in this matrix, no additional alignment algorithm was used. In total, this alignment contained 17,024 characters, representing the 17,024 total genes that make up the pan-genome (every gene present in any strain) of *Xylella* spp. sequences. The outgroup used for all trees was *Xylella taiwanensis* strain Wufong1 isolated in Taiwan in 2014 from *Pyrus pyrifolia* (Su *et al.*, 2014).

We constructed four maximum likelihood phylogenies using RAxML v8.2.11 (Stamatakis, 2014) under a generalized time reversible model. Node support was measured with 1,000 bootstrap replicates (Felsenstein, 1985). Trees were visualized in FigTree v1.4.4(http://tree.bio.ed.ac.uk/software/figtree/) and the Interactive Tree of Life (Letunic & Bork, 2019). Phylogenetic diversity (PD) was calculated as the summation of total branch lengths for

each phylogeny (not including the outgroup) using the R package *adephylo* (Faith, 1992; Jombart & Dray, 2010; R Core Team, 2023).

Ancestral state reconstructions

To conduct ancestral state reconstructions, we used an extant distribution of characters (heritable traits of interest), in this case, the genera of plants from which we isolated the bacteria. Using that distribution, we constructed the most likely history of hosts across the phylogeny at all internal (ancestral) nodes. We are in essence seeking parameter values that maximize the probability of the data (the observed character states) given the hypothesis (a model of character evolution and a phylogeny relating the observed sequences or taxa). Based on available data, the identity of the host plant from which each strain was isolated in the field is identifiable to at least the genus level. This value is used as a point proxy for the true state of interest – potential host range. Since host range must be experimentally determined, in this study we use the host each strain was isolated from as a point representative of an unknown range of susceptibility. Due to this, any subsequent results cannot infer specificity to a given host but imply the ability to infect said host. Because sampling is heavily biased towards symptomatic agricultural crops in the case of *X. fastidiosa*, we interpret each ancestral state as the most likely agriculturally relevant host that the pathogen would have been isolated from.

All taxa were coded based on plant host genus and super order/order (the deepest clade grouping that combined our genera into more than one group). This included two super orders (Asterid and Rosid), one order (Proteales) and 26 genera that were potential hosts for *X. fastidiosa's* hypothetical ancestors at each internal node of the phylogenies. The marginal ancestral state likelihood estimates of each host for all internal nodes of the ML phylogenetic trees were calculated using the re-rooting method of Yang et al. (Yang *et al.*, 1995) in the R package *phytools* (Revell, 2016), and mapped using the package *APE* (Paradis & Schliep, 2019). This method uses the phylogeny of extant taxa to reconstruct ancestral traits of extinct ancestors by analyzing phylogenetic parameters (topology and branch length), along with a model of nucleotide substitution, to build posterior probabilities of character states at each interior node by randomly re-rooting the tree at each internal node and calculating the probability of observing the extant distribution of traits over all possibilities of that internal node character identity. The ML estimates at each internal node were calculated based on both the equal rates transition model (i.e., fixed rate of change between any two hosts) and the symmetrical rates transition model (i.e., fixed rates of host change symmetrically pairwise between hosts, but not between all hosts). The fit of the two models to the data was compared using the Akaike information criterion and can be seen in supplemental table (S2)(Akaike, 1974).

Correlation between host and gene presence/absence

Information on plant host taxonomy was gathered on NCBI's Taxonomy Browser (Schoch *et al.*, 2020). Scoary (Brynildsrud *et al.*, 2016) was used to test if the pan-genome was correlated with hosts at either the super-order scales or the genus scale by conducting a Fisher's Exact Test (FET)(Lury & Fisher, 1972). FET measures the association of each gene in the pan-genome to a trait of interest, which in this case is plant host. While FET requires no association between datapoints, Scoary uses a phylogeny in order to remove lineage specific interdependencies and corrects the p-value based on those interdependencies. Significance was evaluated by the "worst pairwise comparison P", for the phylogenetic corrections, not the naïve

p-values from FET. Individual analyses were conducted to test for correlation of gene presence and absence with each of the 29 coded host groups (26 genera, Asterid, Rosid, and Proteales).

Table 1. Summary of alignments and phylogenetic diversity for each of the four alignments and corresponding phylogenetic trees.

Phylogeny Data Source	**Total Alignment Length** (in base pairs (bp) or genes)	**Phylogenetic Diversity** (summed substitutions per site)
Non-Recombinant Genome	112,819 bp	3.32
Core Genome	354,816 bp	6.65
MLST Genes	4,146 bp	8.65
Pan-genome	17,024 genes	59.15

Table 2: Significant genes whose presence/absence were correlated with host, once phylogenetic history has been corrected for based on common descent

Host Taxonomy	Gene	Annotation	Number host present in	Number non-host present in	Number host not present in	Number non-host not present in	Naïve p-value	Tree corrected p-value
Asterid	group_943	hypothetical protein	92	1	36	220	5.19E-53	3.42E-03
Asterid	group_949	IS200/IS605 family transposase IS609	13	2	115	219	7.09E-05	7.81E-03
Asterid	group_9702	hypothetical protein	9	0	119	221	9.99E-05	1.56E-02
Asterid	group_1963	hypothetical protein	9	0	119	221	9.99E-05	1.56E-02
Asterid	group_1865	hypothetical protein	0	154	128	67	1.94E-45	3.13E-02
Asterid	group_2944	hypothetical protein	0	41	128	180	2.88E-09	3.13E-02
Asterid	fitB_1	Toxin FitB	10	0	118	221	3.50E-05	3.13E-02
Asterid	group_2361	hypothetical protein	10	0	118	221	3.50E-05	3.13E-02
Asterid	socA	Antitoxin SocA	30	6	98	215	1.77E-09	3.86E-02
Asterid	group_3382	hypothetical protein	2	23	126	198	1.90E-03	3.91E-02
Rosid	group_943	hypothetical protein	1	92	218	38	6.13E-52	3.42E-03
Rosid	group_949	IS200/IS605 family transposase IS609	2	13	217	117	8.62E-05	7.81E-03
Rosid	group_9702	hypothetical protein	0	9	219	121	1.15E-04	1.56E-02
Rosid	group_1963	hypothetical protein	0	9	219	121	1.15E-04	1.56E-02
Rosid	group_1865	hypothetical protein	154	0	65	130	1.23E-46	3.13E-02
Rosid	group_2944	hypothetical protein	41	0	178	130	1.29E-09	3.13E-02
Rosid	fitB_1	Toxin FitB	0	10	219	120	4.11E-05	3.13E-02
Rosid	group_2361	hypothetical protein	0	10	219	120	4.11E-05	3.13E-02
Rosid	socA	Antitoxin SocA	6	30	213	100	2.84E-09	3.86E-02
Rosid	group_3382	hypothetical protein	23	2	196	128	1.03E-03	3.91E-02
Vitis	group_3360	hypothetical protein	15	33	160	141	5.15E-03	1.56E-02

Vitis	group_4565	hypothetical protein	174	3	1	171	3.22E-96	3.13E-02
Vitis	group_1893	hypothetical protein	63	0	112	174	2.30E-22	3.13E-02
Vitis	group_4682	hypothetical protein	58	0	117	174	2.35E-20	3.13E-02
Vitis	group_2780	putative HTH_type transcriptional regulator	0	36	175	138	1.73E-12	3.13E-02
Vitis	group_4923	hypothetical protein	22	0	153	174	2.36E-07	3.13E-02
Vitis	group_3389	hypothetical protein	11	0	164	174	8.30E-04	3.13E-02
Vitis	group_3387	hypothetical protein	9	0	166	174	3.52E-03	3.13E-02
Vitis	group_4544	hypothetical protein	104	50	71	124	8.70E-09	3.91E-02
Coffea	group_84	hypothetical protein	19	203	1	126	1.49E-03	3.13E-02
Coffea	group_4414	hypothetical protein	5	208	15	121	1.39E-03	3.91E-02
Coffea	vbhT	Adenosine monophosphate_protein transferase VbhT	5	207	15	122	1.43E-03	3.91E-02
Coffea	group_1250	hypothetical protein	5	207	15	122	1.43E-03	3.91E-02

Figure Captions:

Figure 1. Maximum Likelihood phylogenies for all four genomic subsets with n = 349 strains, collapsed to the potential subspecies level. MulMD and Mul0034 are considered examples of the debated subspecies *morus,* while Ann1 and RAAR8 are examples of the second debated subspecies, *sandyi.* Subspecies *fastidiosa, pauca,* and *multiplex,* have only been collapsed to clades which do not include those four strains. Each phylogeny has a separate scale of substitutions across its branches. Node support is shown as bootstrap values, with values under 50 not displayed. The outgroup used for all trees is the strain Wufong1, a member of the species *Xylella taiwanensis*, which has been trimmed from these trees for visualization.

Figure 2. Maximum Likelihood phylogeny of the core genome, without any clades collapsed. The branch lengths are quantified by substitution rate, and the bootstrap support is depicted by branch color. Each strain name includes an abbreviated reference to its origin, see S1 for more details on the history of each strain. The branch length leading to the outgroup, Wufong1, has been removed to clarify relationships within the species *X. fastidiosa.* Dashed lines are used to connect tips of the phylogeny to the taxa names.

Figure 3. Cladogram of the core genome with the most likely genus of each node mapped onto the branches. Nodes with likelihood of less than 95% for one genus are colored in gray and marked as Undetermined. The ancestral state reconstruction was conducted with an equal rates model.

Figure 4. Cladograms of all four genomic regions with the most likely genus of each node mapped onto the branches. Nodes with likelihood of less than 95% for one genus are colored in gray and marked as Undetermined. (A) is the non-recombinant core cladogram, (B) is the core cladogram, (C) is the MLST cladogram, and (D) is the pan-genome cladogram. All four ancestral state reconstructions were done with an equal rates transition model between hosts.

Figure 5.

Cladograms of all four genomic regions with the most likely super order or order of each node mapped onto the branches. Nodes with likelihood of less than 95% for one super order are colored in gray and marked as Undetermined. (A) is the non-recombinant core cladogram, (B) is the core cladogram, (C) is the MLST cladogram, and (D) is the pan-genome cladogram. All ancestral state reconstructions were done with an equal rates transition model between hosts except the pan-genome, which performed better with additional parameters of the symmetrical rates model.

Figure 1.

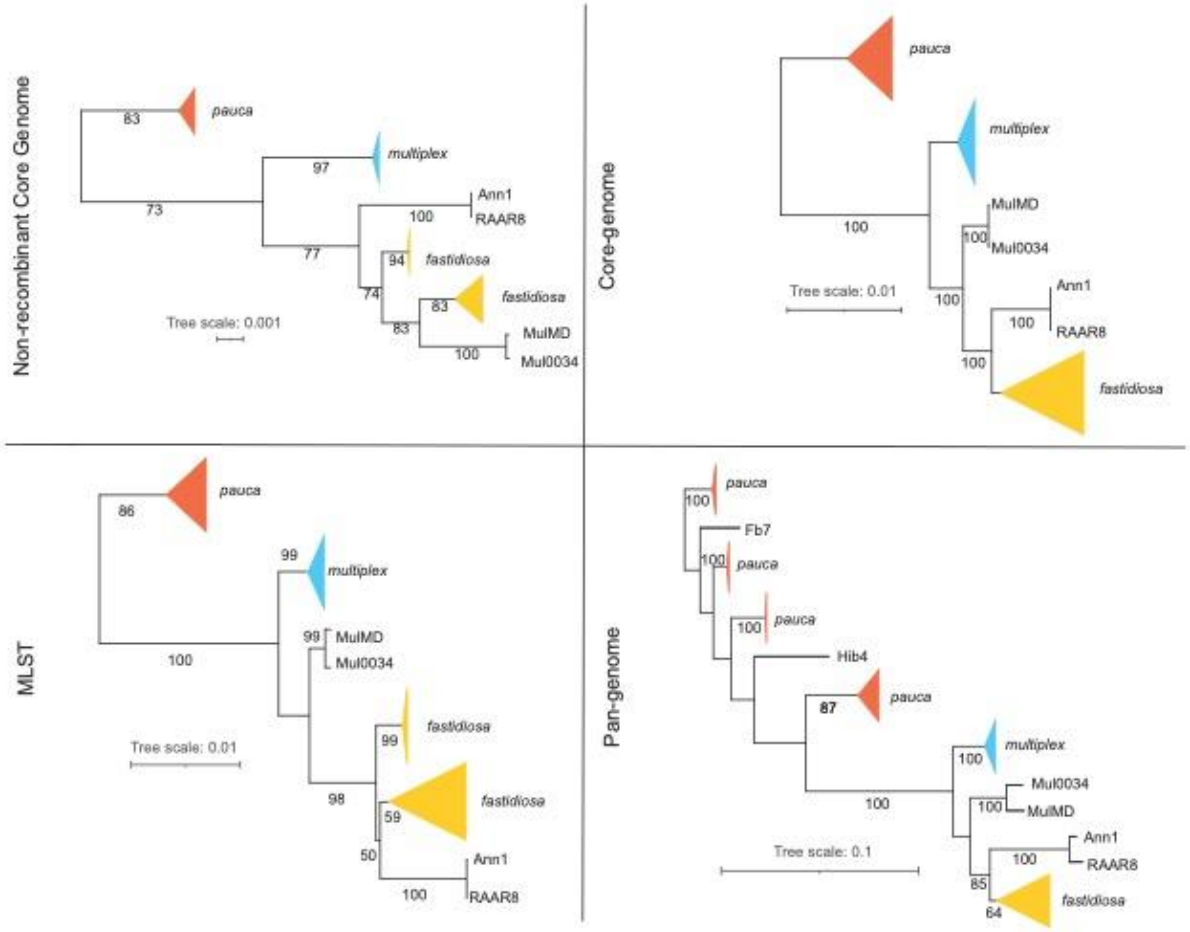

Non-recombinant Core Genome
pauca
multiplex
Ann1
RAAR8
fastidiosa
fastidiosa
MulMD
Mul0034
Tree scale: 0.001
Core-genome
pauca
multiplex
MulMD
Mul0034
Ann1
RAAR8
fastidiosa
Tree scale: 0.01
MLST
pauca
multiplex
MulMD
Mul0034
fastidiosa
fastidiosa
Ann1
RAAR8
Tree scale: 0.01
Pan-genome
pauca
Fb7
pauca
pauca
Hib4
pauca
multiplex
Mul0034
MulMD
Ann1
RAAR8
fastidiosa
Tree scale: 0.1

Figure 2.

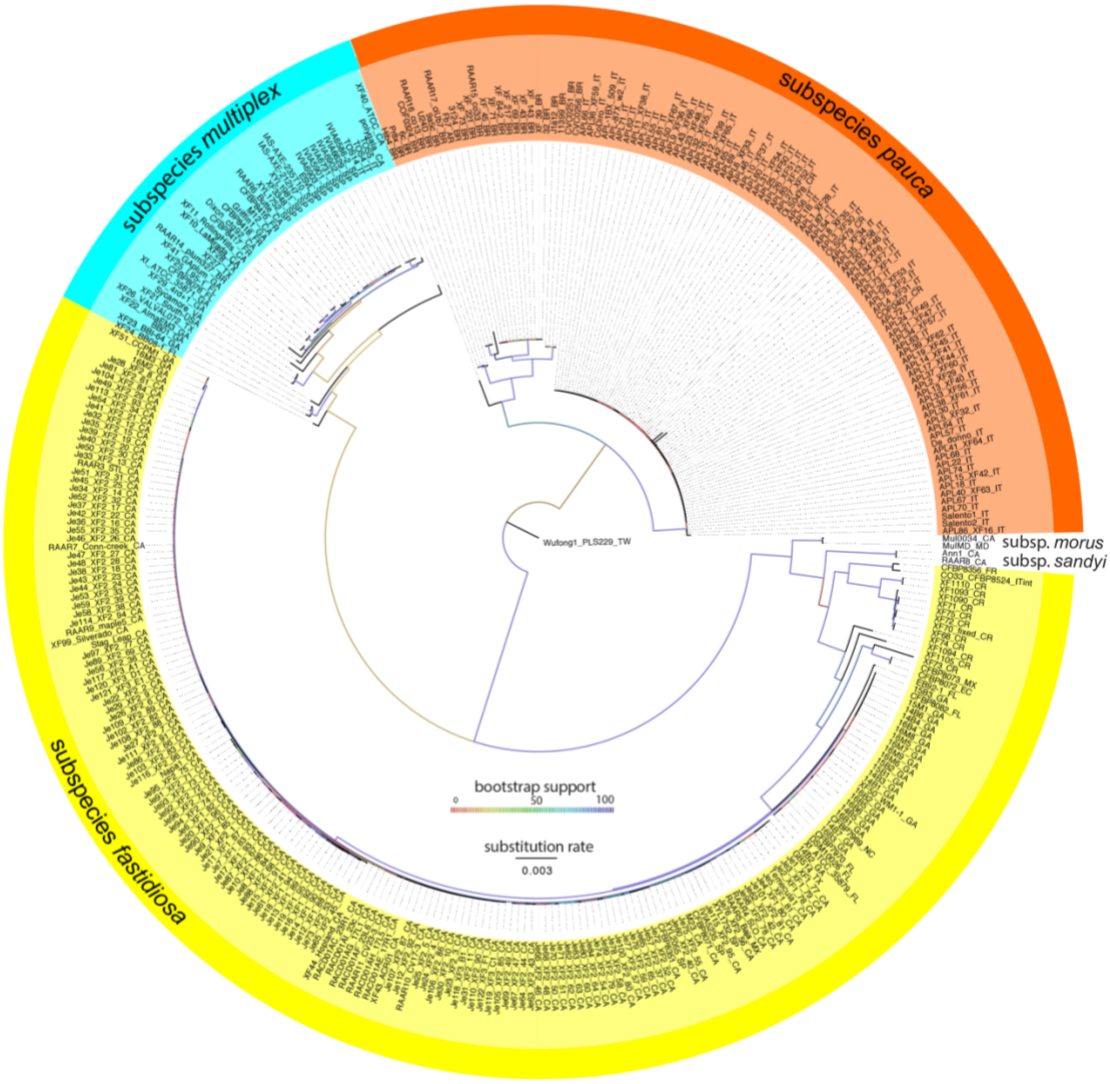

subspecies multiplex
subspecies pauca
subspecies fastidiosa
subsp. morus
subsp. sandyi
Wufong1_PLS229_TW
bootstrap support
0
50
100
substitution rate
0.003

Figure 3.

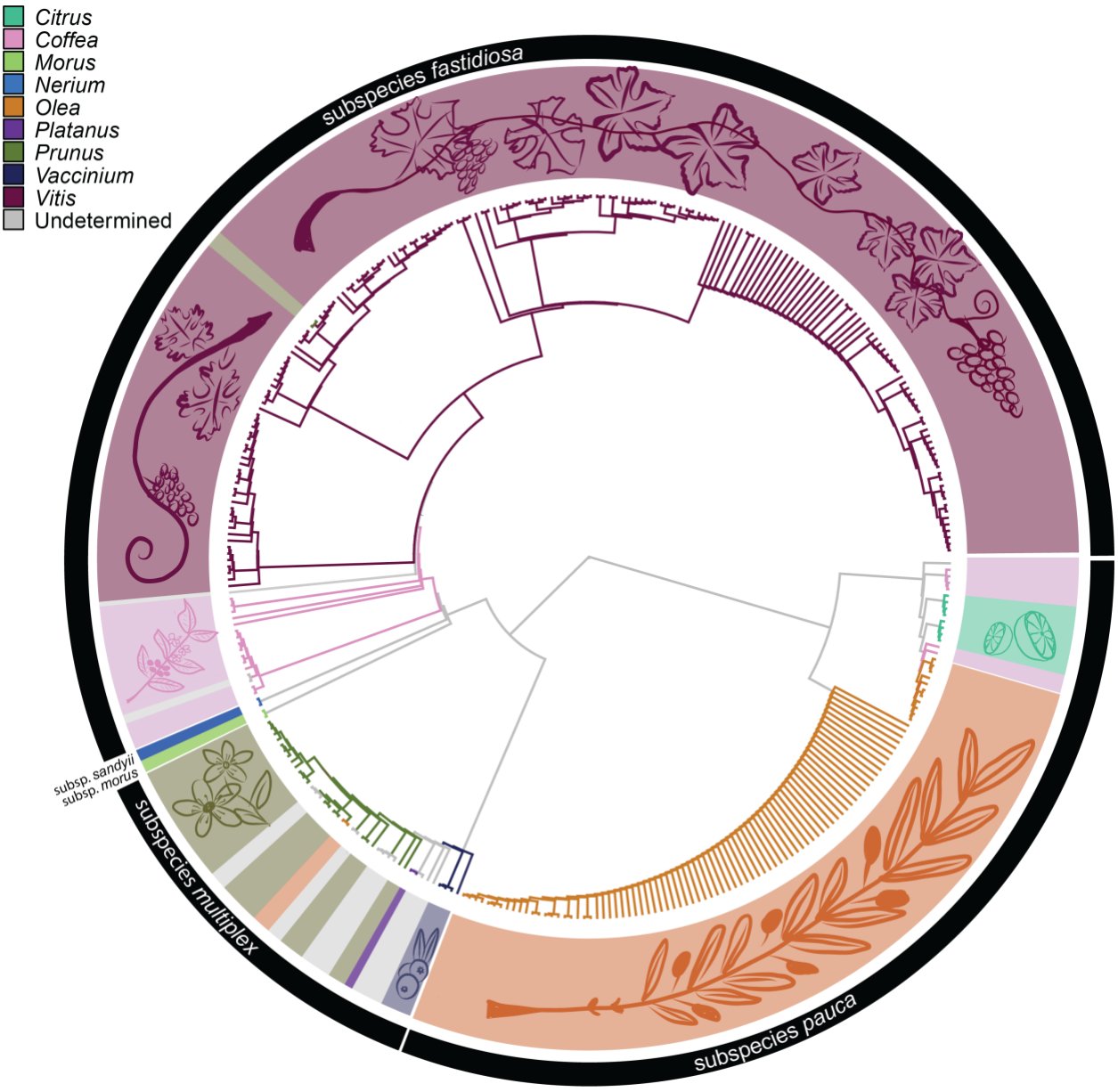

Citrus
Coffea
Morus
Nerium
Olea
Platanus
Prunus
Vaccinium
Vitis
Undetermined
subspecies fastidiosa
subsp. sandyii
subsp. morus
subspecies multiplex
subspecies pauca

Figure 4.

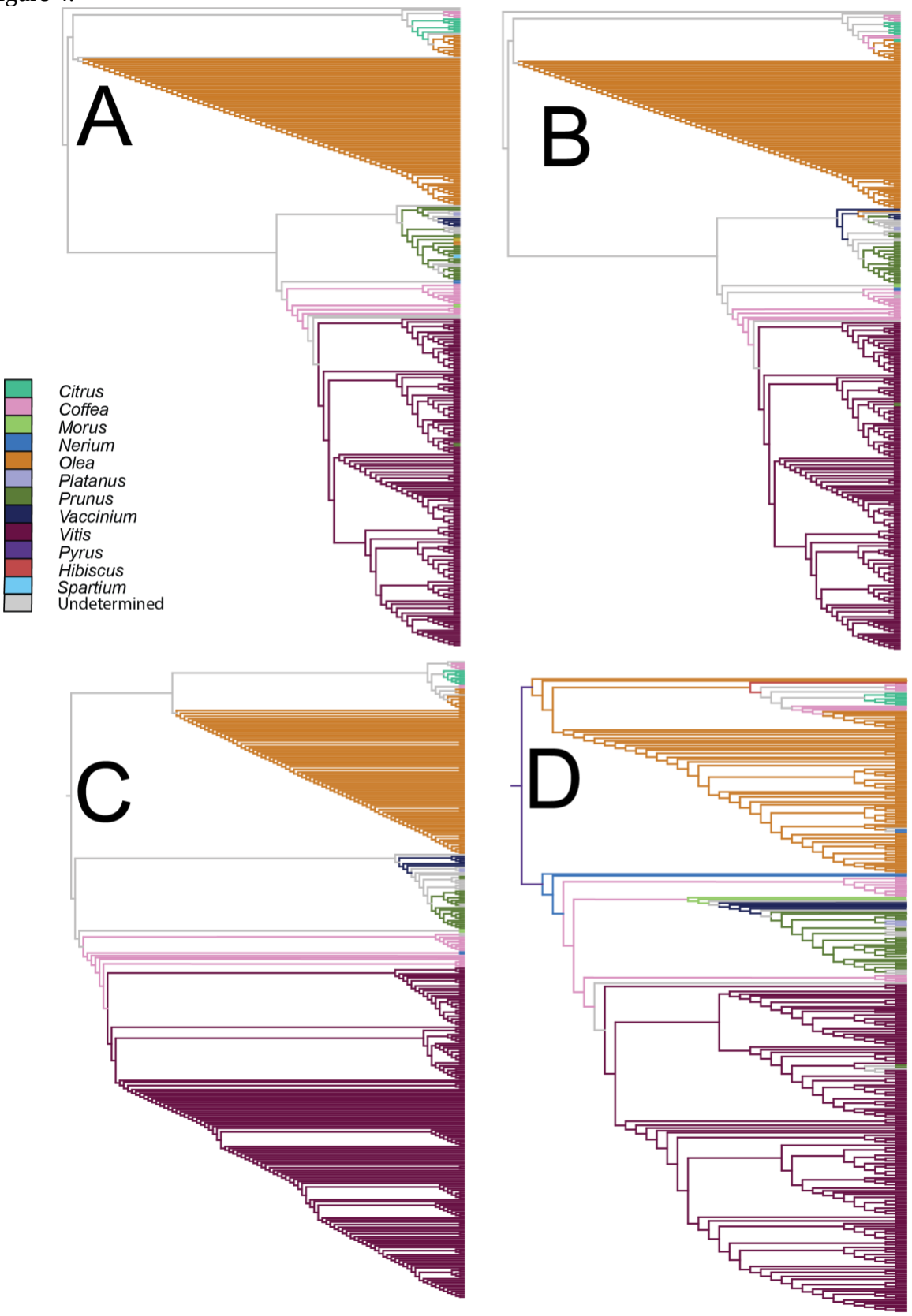
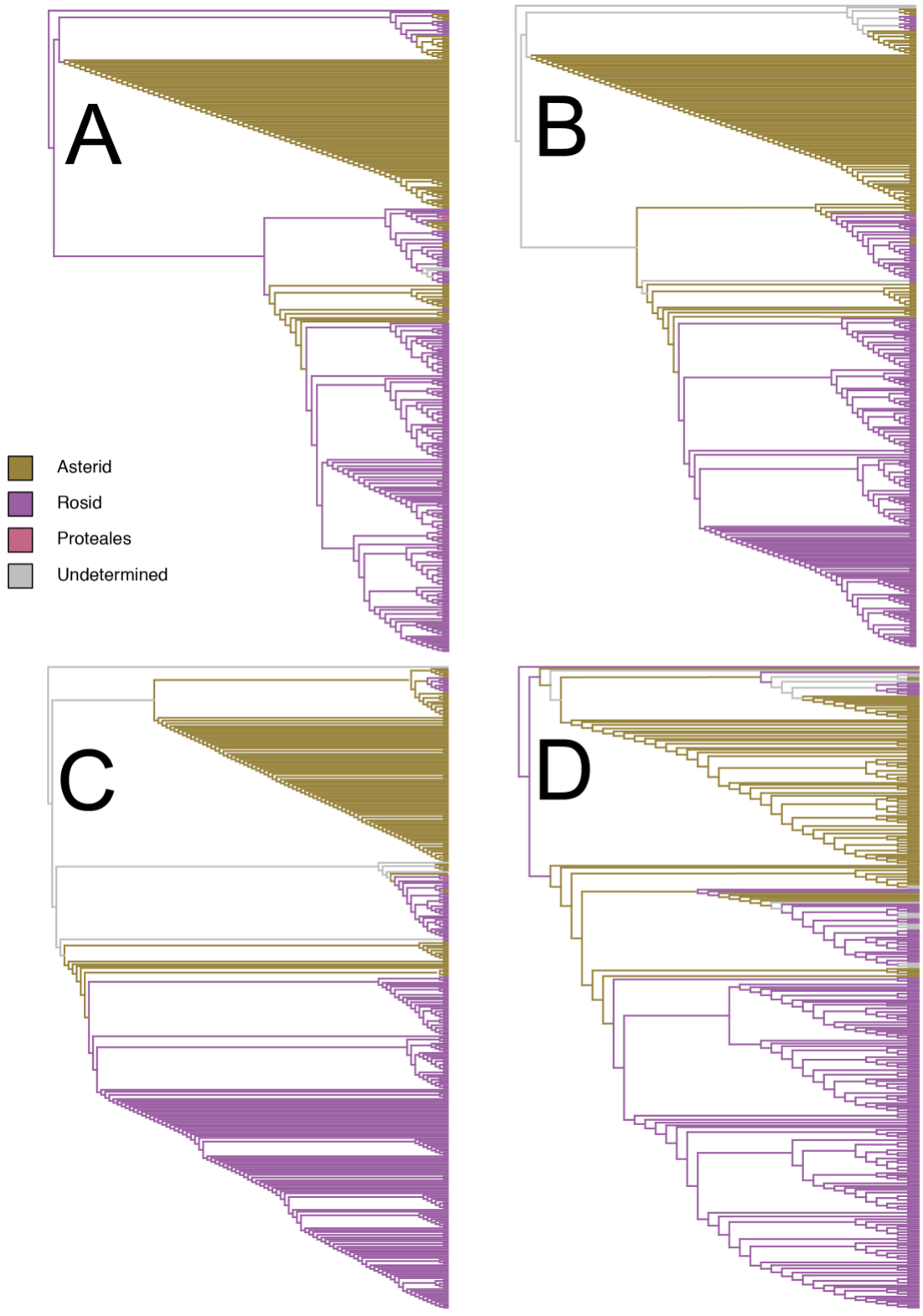

A
B
C
D
Citrus
Coffea
Morus
Nerium
Olea
Platanus
Prunus
Vaccinium
Vitis
Pyrus
Hibiscus
Spartium
Undetermined

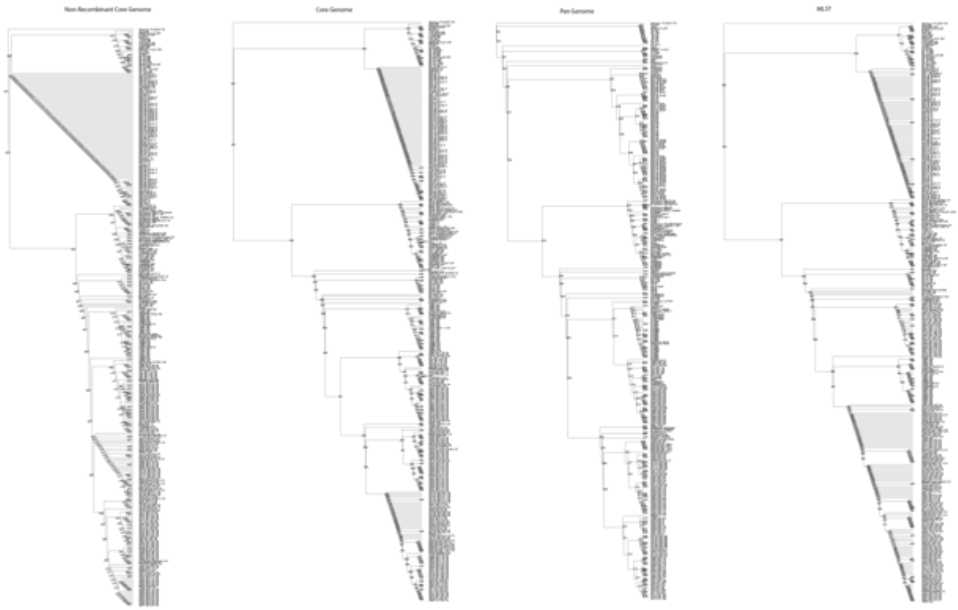
Figure 5.



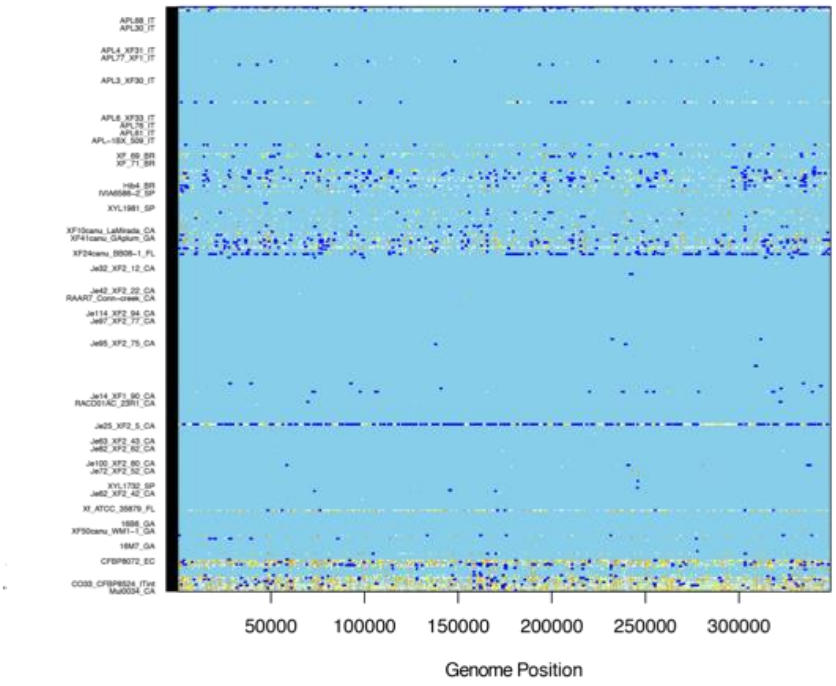
Supplementals – Chapter 1

S1 Metadata with sequence information: The metadata sheet includes information on all sequences used in this study. It includes the sequence name, host information, subspecies information, isolation year, geographic origin, and biosample information or a citation for sequences which that was not available. (S1_metadata.xlsx is linked)

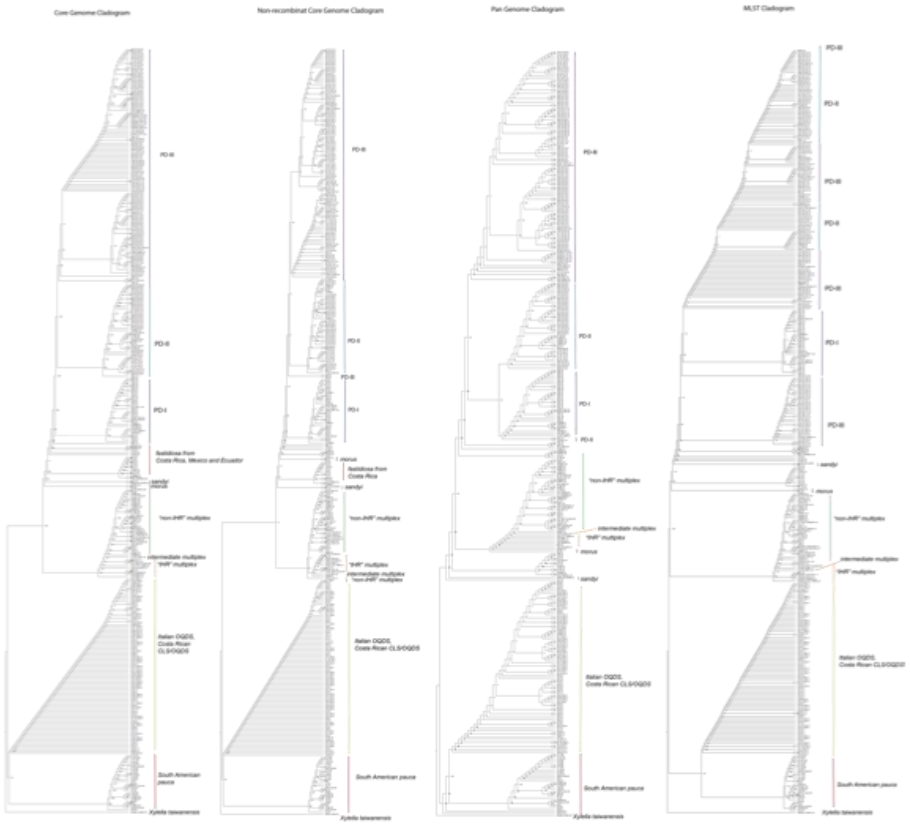
S2 Ancestral State Reconstruction Results: The S2 datasheet includes all of the ancestral state reconstruction results used in this study. The first tab shows the AIC scores comparing the symmetrical and equal rates models for transitions between hosts for all four phylogenies, for both genus and super order level reconstructions. In an equal rates model, all hosts are equally likely to switch to any other host. In the more parameterized symmetrical rates model, each pair of hosts has a unique transition rate, but they are parallel between each pair (i.e. *Prunus* > *Vitis* is the same likelihood as *Vitis* > *Prunus*). The following 8 tabs include the results for the model with the lowest AIC score for each reconstruction. In these tabs, the columns represent the node number from the phylogeny, and the columns represent the potential hosts. The data in each cell is the likelihood that a given node would have been able to infect a particular host. The S2 node numbers file is paired with this datasheet, as the node numbers are needed to interpret which part of the phylogeny the reconstruction is referring to. (**S2_AncStateResults.xlsx** is linked and S2_nodenumbers included here)



S3 ClonalFrameML results: Genomic map of recombination events based on the core genome visualized along the length of the core-genome alignment. Blue regions represent loci that have undergone recombination events. The core genome alignment is 354,816 bp in length and the identified regions with recombination comprise 112,819 bp, or 31.8% of the genome. (S3_ClonalFrameML included)



S4 Topological Comparisons between Phylogenies: Figure that has four cladograms, with all sequence names included as well as bootstrap values. The general regions of within-subspecies clades that are discussed in the paper are marked alongside each cladogram. (Included)



S5 16S rDNA Phylogeny: Maximum Likelihood phylogeny of the 16S rDNA, without any clades collapsed. The branch lengths are quantified by substitution rate. Each strain name includes an abbreviated reference to its origin, see S1 for more details on the history of each strain. Out of the 349 strains used in the study, there were 40 unique sequences resulting in large unresolved portions of the tree. 5 APL strains were removed due to poor coverage. Methods used to build the 16S rDNA phylogeny were the same as those for

the MLST phylogeny in the main text, with the exception of using the 16S rDNA gene as a query instead of the MLST genes.

CHAPTER TWO

A pathogen of good taste: a bacterial host shift from coffee to wine grapes

Alexandra Katz Kahn[1], Jasslin Cervantes[1], Rodrigo P. P. Almeida[1]

[1]Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA 94720

Corresponding Author: rodrigoalmeida@berkeley.edu

Abstract

Xylella fastidiosa is a recurring nuisance, this bacterial pathogen has recently emerged in novel geographic locations infecting a breadth of host plants: 679 species and counting (Delbianco *et al.*, 2023). An introduction of *X. fastidiosa* ssp. *fastidiosa* from Central America to the United States several hundred years ago has since been the source of outbreaks across the globe. In the U.S., particularly in California, the introduced bacterium is frequently found in grapevines (*Vitis spp.*), *X. fastidiosa's* most negatively affected post-introduction host. We test the ability of the introduced strains to infect an ancestral host, coffee (*Coffea arabica*), to elucidate the process of a documented and economically important host jump. Using *X. fastidiosa* ssp. *fastidiosa* whole genome sequences, 15 from Costa Rica and 292 from the introduced clade, we test for traces of adaptation to grapevines. We also demonstrate that the introduced strains do not, for the most part, persistently infect *C. arabica*. Furthermore, we do not observe an overall increase in the virulence of the introduced strains towards coffee, indicating a lack of hypervirulence. We also found both genes and single nucleotide polymorphisms (SNPs) that are associated with the host shift to grapevines. These results support the hypothesis that genetic adaptation to a novel host occurred following the introduction of the pathogen.

Introduction

California is the center of American fruit, vegetable and nut production and is a globally important exporter of plants (Coe *et al.*, 2020). As a hub of international plant trade, California has been both a source and a sink for the introduction of novel phytopathogens across the globe. *Xylella fastidiosa* ssp. *fastidiosa* is a bacterial pathogen that impacts grapevines and almonds in California. *X. fastidiosa* subspecies *fastidiosa*, was likely of Central American origin and possibly introduced via coffee (*Coffea arabica*) imports to the United States in the 1800s (Nunney *et al.*, 2010; Castillo *et al.*, 2020; Vanhove *et al.*, 2020; Kahn & Almeida, 2022). After its introduction into the U.S., the pathogen has continued to spread, including recent introductions into Spain (estimated 1993) and Taiwan (estimated 2002) (Moralejo *et al.*, 2020; Castillo *et al.*, 2021a).

In California, outbreaks of *X. fastidiosa* ssp. *fastidiosa* have been devastating to the grape (*Vitis spp.*) industry. In contrast, infections have not caused notable issues in the agriculturally important regions of Central America, where the ssp. has likely existed in the ecosystem for at least thousands of years (Nunney *et al.*, 2010). This is not a universal trait of *X. fastidiosa* in *C. arabica* as opposed to *Vitis*. In Brazil, *X. fastidiosa* ssp. *pauca* infections reduce *C. arabica* yields, despite the pathogen's long history in the region (Rocha *et al.*, 2010). *X. fastidiosa* is complex in its strain specificity to various host plants, with many conflicting examples of host susceptibility in geographic areas. Nonetheless, documenting host susceptibility to different subspecies and strains of the pathogen is vital. Clarifying the process of host jumps in *X. fastidiosa* is urgent for global agricultural security, as new outbreaks continue to put plant health at risk.

While it is likely that *X. fastidiosa* ssp. *fastidiosa* is broadly present in Central America, currently only Costa Rica has reported this pathogen. Here we refer to "Costa Rican" strains as our representative Central American strains, although there is certainly much larger diversity of *X. fastidiosa* in the region. *X. fastidiosa ssp. fastidiosa* frequently infects *C. arabica* in Costa Rica without substantially impacting production, although subtle leaf curling symptoms correlate with *X. fastidiosa* infections (Rodríguez *et al.*, 2001; Montero-Astúa *et al.*, 2008). In Costa Rica, *X. fastidiosa* ssp. *fastidiosa* also infects periwinkle (*Vinca sp.*), guava (*Psydium guava*), and avocado (*Persea americana*)(Castillo *et al.*, 2020). While symptoms have been reported in these hosts, disease severity and progression have not been thoroughly evaluated.

Some strains of *X. fastidiosa* ssp. *fastidiosa* have been detected in *Vitis* in Costa Rica, where *V. vinifera* is not commonly grown (Goheen *et al.*, 1979; Aguilar *et al.*, 2008; Nunney *et al.*, 2010) but only sequence-type data was collected for these strains. However, a recent study found that related strains from Costa Rica did not cause infections in *Vitis* (Castillo *et al.*, 2020). Given these results, infectivity and virulence by the same strain may differ between *Vitis* and *C. arabica*. A clear phylogenetic delineation of pathogenicity, which requires extensive cross-inoculation experiments, has not yet been determined. This study begins to test the adaptation of the U.S.-introduced clade to *Vitis* cross-inoculation and computational methods.

Within the genus *Vitis*, there is wide genetic diversity conferring variation in both tolerance and resistance to Pierce's disease (PD), the disease caused by *X. fastidiosa* ssp. *fastidiosa* infections in grapevines. Differences are seen between species such as the naturally tolerant *V. arizonica* and *Muscadinia rotundifolia*, and the susceptible *V. vinifera* (Ruel and Walker, 2006). This variation has recently been used to hypothesize that there might have been PD present in the U.S. for longer than has been estimated by all evolutionary data (Morales-cruz *et al.*, 2023). But disease susceptibility also varies largely within *V. vinifera* (Kahn et al., 2023;

Raju and Goheen, 1981), an intensely bred species that includes a wide and diverse range of cultivars. In contrast to crespera (the disease caused by *X. fastidiosa* infections in coffee) in Costa Rica, PD in the United States is quite virulent, persistent, and economically damaging. Infection with the pathogen causes severe symptoms such as stunting, leaf scorch, shriveled fruit, and eventual plant death in many cases (Kahn *et al.*, 2023). There are several additional adversely impacted hosts of species within the introduced clade. Of economic importance is almond (*Prunus dulcis*) suffering Almond Leaf Scorch (ALS). Other hosts of strains in the introduced clade include maple (*Acer sp.*), plum (*Prunus domestica*), ragweed (*Amrosia artemisiifolia)*, sweet cherry (*Prunus avium*), and western redbud (*Cerceris occidentalis*).

C. arabica is a unique host for *X. fastidiosa*. It has widespread documented infectivity coupled with non-deadly symptoms, and it was likely the source of two economically devastating introduction events: to the United States in grape and to Italy in olive (Nunney *et al.*, 2010; Sicard *et al.*, 2021). Nonetheless, *C. arabica* is generally understudied as a host plant. Data from Brazil show that *C. arabica* and *Citrus* strains of *X. fastidiosa* ssp. *pauca* do not cross-infect (Almeida *et al.*, 2008; Francisco *et al.*, 2016), showing variation in host specificity. However, *C. arabica* is also a host of at least two of the three major subspecies of *X. fastidiosa,* showing broad susceptibility. It has been hypothesized that strains of *X. fastidiosa* do not need to undergo many genetic changes in order to infect *C. arabica*, potentially making it a host susceptible to low-cost host jumps (Thines, 2019; Kahn & Almeida, 2022). While ssp. *pauca* and ssp. *fastidiosa* have both been documented to infect *C. arabica*, slow progression and low virulence have led *C. arabica* to be described as a "latent carrier" (Sicard *et al.*, 2021). So far, research has shown that the strains able to infect *C. arabica* are quite diverse (Jacques *et al.*, 2016). The infectivity of *X. fastidiosa* ssp. *multiplex* has never been tested in *C. arabica*. Subspecies *multiplex* is native to the south-eastern United States and is present throughout the U.S., South America, and Europe (Landa *et al.*, 2019). However, the advent of coffee production in California coupled with historic documentation of ssp. *multiplex* in Argentina, Brazil, and Paraguay, call for investigation of whether *C. arabica* can serve as a host to ssp. *multiplex* (French & Kitajima, 1978).

Upon introduction to the U.S., *X. fastidiosa* was exposed to new hosts, climate, vectors, and agricultural practices alongside a rapidly growing agricultural industry. While it is likely that the introduced strains were exposed to selective pressures that increased specificity to the conditions in the U.S., mainly climate and host, that is not the only possibility. It is also plausible that instead of adapting to the specific pressures in the new environment, the introduced population of strains became generally hypervirulent. This could explain the high level of virulence towards *V. vinifera* in the U.S. compared to the lower virulence observed in Costa Rica. We investigated both neutral and adaptive changes during the process of naturalization for this pathogen using whole genome sequences.

Our major hypotheses are as follows:

1. During the host shift and adaptation to *Vitis*, *X. fastidiosa* strains from California may have lost the ability to infect to *C. arabica* plants. This loss may have been due to lack of selective pressure to infect *C. arabica* and/or selective pressure to infect *Vitis*. This case would be described as a host jump that leads to genetic differentiation from the parent population as defined by Thines (2019). We would then expect less persistent infections of the introduced strains in inoculated *C. arabica* plants.

2. If the introduced strains have generally increased in virulence over the past few centuries compared to the ancestral strains, strains from California may be hypervirulent in *C. arabica* which we can test through documenting disease symptoms during inoculation studies.
3. Given that experimentally inoculated strains from Costa Rica have not been infectious to *Vitis* (Castillo *et al.,* 2020), we hypothesize that the strains present in California have undergone genetic changes that are associated with their ability to both infect and be highly virulent to *V. vinifera* in the U.S. We hypothesize that there will be genes under positive selection associated with this host shift, as well as genes that have been gained and lost and an accumulation of single nucleotide polymorphisms (SNPs) that are associated with this change.
4. Because a wide diversity of strains are able to infect *C. arabica* (Jacques *et al.*, 2016), we hypothesize that strains of ssp. *multiplex* circulating in California will be able to infect *C. arabica* plants.

MATERIALS & METHODS

Coffee plant infections

In March 2022, 120 coffee plants (*C. arabica*) were divided into equal groups and inoculated with either a buffer control or one of four CA strains of *X. fastidiosa,* ssp. *fastidiosa*: Napa1 (from *V. vinifera*, Napa), ALS17T5 (from *P. dulcis*, San Joaquin), Je115 (from *V. vinifera*, Bakersfield) or subspecies *multiplex* ALS15T2 (from *P. dulcis*, Claribel), or the sterile succinate-citrate-phosphate buffer (SCP) control (Hopkins, 1984). While Napa1 and Je115 are both isolated from *V. vinifera*, they are from different climates and clades of *X. fastidiosa,* and might have experienced distinctive selective regimes (Vanhove *et al.*, 2020). ALS17T5 was isolated from a symptomatic almond plant and is in one of several clades of strains infecting only almond trees, however this clade is nested within other clades that infect *Vitis* (see Figure 3).

Disease-free *C. arabica* plants (60 cv. Geisha, 60 cv. Red Catuaí; 12 replicates per treatment) were donated from Frinj coffee company for this experiment. As positive controls, 20 *V. vinifera* cv. Chardonnay (4 per strain) were inoculated as well as 50 *Helianthus annus. V. vinifera* is known to be highly susceptible to subspecies *fastidiosa* but not subspecies *multiplex*, while *Helianthus annus* is susceptible to both subspecies (Wistrom & Purcell, 2005; Delbianco *et al.,* 2023). Cell suspensions were prepared by suspending week-old cells grown on solid medium in SCP buffer just prior to inoculation. Each suspension was made by scraping 10 streaks of 20µL into 1 mL of buffer. Inoculations were conducted using 2 10µL beads of inoculum and a 00-size entomological pin used to pin prick through the bead of cell suspension several times until the inoculum absorbed into the plant xylem. Inoculations were conducted on small plants with typically 3 full-leaf pairs, and two inoculum beads were placed just above and below the center leaf pair. Plants were not watered the morning of inoculation to optimize absorption of the inoculum into the xylem vessels.

Symptom measurements took place in August, September, October, December, 2022, and January and April 2023, following the March 2022 inoculation. In August and September 2022, all internode lengths along the main stem of the plant were measured along with the heights of each plant. In October, December, January, and April, only the total plant heights were measured to detect stunting. Plants were also visually assessed for foliar scorching.

Control and experimental plants were tested for the presence of *X. fastidiosa* via qPCR or culturing. DNA was extracted using a DNeasy plant mini kit (Qiagen) and then quantified using qPCR with a primer pair targeting the gene encoding *RecF*, RecF1_F+R; for qPCR protocol see Sicard et al. (2020). qPCR was run in duplicate with positive and negative controls on each plate; all samples with C_t values of 37 or higher were considered "undetected" and were considered negative. Culturing was conducted using approximately 0.1 grams of petiole tissue, or the entire petiole and midrib of a *C. arabica* leaf. Samples were surface sterilized, chopped, ground in a Polytron, and plated on PWG media using the method from Hill and Purcell (1995).

On April 3rd, 2022, petioles that were directly above the inoculations site were sampled from all 50 sunflower plants and *X. fastidiosa* populations were measured via both qPCR and cell culturing. On June 16th, 2022, *V. vinifera* plants were sampled from (for culturing & qPCR) and also visually assessed for symptoms. On July 14th, one leaf from the 2nd leaf pair above the inoculation point was used for detection via qPCR. On August 1st, ,15th, and 18th, one leaf from the 2nd leaf pair above the inoculation point was cultured from each *C. arabica* plant. On October 4th, 2022, the opposite leaf was taken for qPCR. While *Vitis* is not as susceptible infections from ssp. *multiplex*, typically there is some detectable infection, however those infections have 10-100 fold lower population sizes than infections by ssp. *fastidiosa* (Almeida & Purcell, 2003).

In February – April 2023, we cultured from the leaves using a different method due to loss of leaves from the ~ 8 cm above the inoculation point. The lowest 8 leaves were collected, and the petioles were pooled for sterilization, tissue grinding, and plating.

Statistical Analysis of Inoculation Data

All analyses were performed using R statistical software version 4.2.3 (R Core Team, 2023). Linear mixed models were built using the package lme4 to add a random effect to account for repeated sampling of the plants over the course of the experiment (Bates *et al.*, 2015). (a) As analyses of virulence, we tested the effects of variety and the interaction between treatment and sampling date on both internode length, and (b,c) separately on plant height (split into two models by variety), each with a random effect of plant ID using the lmer function of lme4. (d) As an analysis of infection persistence via detection assays, we tested the effects of variety and the interaction between treatment and sampling date on the count of plants that tested positive using a generalized linear mixed model with binomial error using the package glmmTMB and the function glmmTMB (Brooks *et al.*, 2017).

Whole Genome Dataset

Sequences used for this analysis were assembled in house for prior publications from our group or downloaded from NCBI and annotated in-house using Prokka (Seemann, 2014). All publicly available whole genome sequences from *X. fastidiosa* ssp. *fastidiosa* were used, as well as one sequence from ssp. *multiplex* which was used as an outgroup for phylogenetic analyses. Some included sequences have not yet been made publicly available. A list of the metadata for all strains and sequence information can be found in Table S1.

Alignment and Phylogenetics

The annotations for each sequence were run using the alignment program Panaroo (Tonkin-Hill *et al.*, 2020) which was run in strict clean mode and used to create the pan and core genomes as well as the core gene alignments. Core genome alignments (with and without the ssp. *multiplex* outgroup CFBP8173) were run through Clonal Frame ML (Didelot & Wilson,

2015) to identify recombinant regions, which were removed using an in-house Python script. A SNP alignment was extracted from the non-recombinant alignment using snp-sites (Page *et al.*, 2016). The non-recombinant SNP alignment was then used to generate a phylogenetic tree in RAxML (Stamatakis, 2014) using the GTR cat parameters and 100 iterations.

Genome-wide association studies

Scoary (Brynildsrud *et al.*, 2016) was run on both the core genome SNPs (without recombination removed) as well as the gene presence absence file (generated using Panaroo and snp-sites) to correlate genome content among strains isolated from plants in the genus *Vitis.* Most strains were isolated from *Vitis vinifera* aside from two isolates from *Vitis rotundifolia* and three unidentified *Vitis sp.* Scoary uses a phylogeny in order to remove lineage-specific interdependencies and offers both a multiple hypothesis corrected P-value (Bonferroni's correction) as well as a phylogenetically corrected P-value.

Detection of regions under positive selection

To detect whether there was evidence of either positive or negative selection via dN/dS ratio calculations, the SLAC (Single-Likelihood Ancestor Count) test from HyPhy (Kosakovsky Pond & Frost, 2005) was used. SLAC is a site-specific tool that calculates substitution rates at individual codons given the alignment and phylogeny. Gene alignments were built using Panaroo and subsequently modified for input using Macse (Ranwez *et al.*, 2011). Individual gene trees were constructed using RAxML. These inputs were collectively used to run SLAC.

RESULTS

Coffee is a poor host of the U.S. adapted ssp. *fastidiosa* strains and ssp. *multiplex* strain

Although infections did not persist in all *C. arabica* plants, we detected bacteria of all four strains by cell culture at 2 leaf nodes above the initial inoculation point 5 months post inoculation (Figure 2). Four months after inoculation, in July, most inoculated plants were positive, however by the final detection 1 year after inoculation, there were only a few persistent infections. There was no detectable difference between the detection rates of between the two cultivars Geisha and Red Catuaí (Variety effect $p = 0.76$, $df = 1$, $chi^2 = 0.095$) however there was a significant difference between strains used for inoculation (discounting the control in which no samples tested positive) which varied based on climate, host, and subspecies according to a generalized mixed model (Strain effect $p = 0.019$, $df = 3$, $chi^2 = 9.91$). Estimates were lowest for the strain from ssp. *multiplex*, showing fewer overall positive samples than from plants infected by ssp. *fastidiosa.* Both *P. dulcis* and *V. vinifera* derived strains, as well as ssp. *fastidiosa* and ssp. *multiplex* were able to infect the *C. arabica*, but all infections reduced substantially throughout the year. Positive controls (sunflower and grape inoculations) showed that all four bacterial strains were viable and able to infect to known hosts (Figure S1) .

No evidence of hypervirulence in coffee plants

During the summer of 2022, internode lengths along the main stem were measured for each plant to check for the classic crespera symptom of internode shortening. When tested with a linear mixed model there was no difference found between the treatment groups for internode length. Heights were also measured from August 2022 – April 2023 in order to detect stunting. Of the two *C. arabica* cultivars, only Red Catuaí showed evidence of symptoms associated with the infection – minor stunting of the infected plants compared to the controls. The interaction

between height and sampling time was significant (p=<2e-16, chi^2 = 509.06, df = 20) in Red Catuaí but treatment alone was not significant (p=0.15, chi^2 =6.72, df=4). The cultivar Geisha, which exhibited more overall variation in height, showed no detectable stunting of the infected varieties, the control treatment intercept for plant height was the lowest estimate of all 5 treatments.

Genome wide association study shows SNPs & genes associated with host

A total of 6,236 SNPs were detected as being significantly associated with the host *Vitis* using the Bonferroni correction for multiple hypothesis testing with p-values <0.05. However, this still does not exclude the possibility of significance via phylogenetic proximity across the tree. Given the few clades of *Vitis*-associated bacteria, this dataset did not offer the power to use the most conservative method used by Scoary, the worst pairwise comparison p-value, which would identify only SNPs that have arisen independently across the phylogeny. However, data presented here (Table 2), show genes that have a significant corrected P-value, and also have a best pairwise comparison P <=.125, showing some indication of independent emergence. 22 SNPs fit those criteria. Out of those 22, there were 9 genes in which several SNPs within them emerged as significant, likely due to linkage disequilibrium. Amongst the then total of 13 genes identified with significant SNPs, 5 are only identified to the level of hypothetical protein. Most genes that had multiple SNPs detected as significant also had identical frequencies of those SNPs across the populations, showing high linkage of those sites. Due to this, only one SNP is shown per gene and included in the enrichment tests. The 8 non-hypothetical significant genes are *azu* (Azurin), *carA*, (Carbamoyl-phosphate synthase small chain), *clpP* (ATP-dependent Clp protease proteolytic subunit), *nadD* (putative nicotinate-nucleotide adenylyltransferase), *mutY* (Adenine DNA glycosylase), *ubiJ* (Ubiquinone biosynthesis accessory factor UbiJ), *exbD_1/2* (Biopolymer transport protein ExbD), *IpxD3/4* (UDP-3-O [3-hydroxymyristoyl] glucosamine N-acyltransferase). KEGG orthologies show that 5 of the 8 identified and named genes are classified in the metabolism family in terms of their molecular functions (Table 2) (Kanehise & Goto, 2000). All identified SNPs were present at lower frequencies in strains from *Vitis* than other hosts, with a mean frequency of 0.047 in *Vitis* derived strains and a mean frequency of 0.49 in non-*Vitis* derived strains.

Gene gain and loss was also tested using Scoary for host associations and the same p-value corrections to go from 473 genes with Bonferroni p <0.05 (data not shown) to 37 genes that also had a best pairwise comparison p <= 0.125 (Table 3). Out of the 37 identified genes, 33 are only identified as hypothetical proteins. The four named genes are *xerC_1/2* (Tyrosine recombinase XerC), *hcaB* (3-phenylpropionate-dihydrodiol/cinnamic acid-dihydrodiol dehydrogenase), *mdtA_1/2/3* (Multidrug resistance protein MdtA), and a cluster identified using Panaroo that includes the three genes *CnrA* (Nickel and cobalt resistance protein CnrA), *swrC_2* (Swarming motility protein SwrC), and *acrF_2* (Multidrug export protein AcrF).These genes each belong to a different KEGG orthology (genetic information processing, metabolism, environmental information processing, and signaling and cellular processes, respectively).

Detection of selection in the introduced population

Thirty-eight annotated genes were found to have sites under positive selection, identified via HyPhy's SLAC program. SLAC identifies positive selection at a per codon scale across all samples included, for future analyses with this dataset I intend to also test for directional selection and lineage specific positive selection, which should have more computational power

to identify relevant genes. Out of these 38 genes identified, 30 also contained SNPs that were found to be significantly associated with the host *Vitis* using the program Scoary (Table 1), with a Bonferroni corrected p value <0.05. In terms of molecular functions, the largest KEGG orthology groupings were for Metabolism (15) and Genetic Information Processing (10) (Kanehise & Goto, 2000).

DISCUSSION

Information about *X. fastidiosa* ssp. *fastidiosa* infections in coffee in Central America is still limited – little symptomatology is associated with the disease; however, there are persistent infections in the field (Rodríguez *et al.*, 2001). The results of this work suggest that the host jump included adaptation of the *X. fastidiosa* strains to *Vitis* and lessened the ability of the U.S. strains to infect *C. arabica*. However, the ability for the pathogen to move throughout the *C. arabica* plants' xylem vessels demonstrates a higher infectivity than expected in a fully resistant plant. Movement of the strains away from the inoculation point was observed, which shows the ability of the bacteria to successfully degrade the pit membrane, which is often a barrier to colonization. However, in contrast to that finding, there was a reduction of positive-testing *C. arabica* plants over the course of the year after the inoculation. This demonstrates a lower infectivity in *C. arabica* than the strains have to *V. vinifera* where infections in a greenhouse are sustained post-inoculation. This could demonstrate a reduction in the ability to create chronic infections in *C. arabica* by the *Vitis*-adapted strains. It is possible that the chronic infections of *X. fastidiosa* in *C. arabica* plants observed *in situ* may be caused by high rates of re-inoculation by insect vectors rather than strain (or subspecies) level variation in infectivity. That is not the case for ssp. *fastidiosa* in *V. vinifera*, just one inoculation is sufficient for high virulence, and only cold winter temperatures have been known to cure infections that are otherwise chronic (Purcell, 1977; Hopkins & Adlerz, 1988).

The virulence of the U.S. strains to *C. arabica* is not as high as to *V. vinifera*, as shown by the lack of severe symptoms in *C. arabica*. Given that, this study does not offer evidence that the California ssp. *fastidiosa* strains are generally more virulent than the ancestral strains in Central America. While not a likely scenario, there is a possibility that instead of experiencing adaptation to a specific host, the introduced strains became generally more virulent, which has been hypothesized about the globally spreading ssp. *pauca* strain infecting olive in Italy (Giampetruzzi *et al.*, 2017; Sicard *et al.*, 2021). In this paper, we present hypervirulence as a possible scenario, but as it is not supported by the data, we were able to rule it out. Symptom development in infected *C. arabica* plants consists of minor stunted growth, compared to the severe leaf scorch, matchstick petioles, shriveled fruit, and often plant death that occurs in *V. vinifera*. There are records of virulence of *X. fastidiosa* in *C. arabica*, such as evaluation of ssp. *pauca* in Brazil. In inoculation experiments with those strains, *C. arabica* may still develop symptoms slowly, and the proportion of positive plants does also reduce, however after 8 months the percentage positive was around 30% (Prado *et al.*, 2008), not entirely dissimilar to the results from this project, which extends some uncertainty about this system.

All methods used detected genetic signatures of adaptation, and many genetic candidates were identified by multiple methods. We were able to identify genes and SNPs associated with the host *Vitis* as well as genes under positive selection in strains isolated from *Vitis*. These genes included many hypothetical proteins, however genes with known functions pertaining to infections by *X. fastidiosa* were also identified.

The genes identified by multiple methods included those whose functions in *X. fastidiosa* have been previously investigated, while many still have unknown functions (Cicero *et al.*, 2007). *ClpX*, the gene for the ATP-dependent protease ATP binding subunit, was previously identified as being upregulated 4x during induction of biofilm condition (De Souza *et al.*, 2004), a physiological state that is vital for virulence and vector colonization (Killiny *et al.*, 2013). Mutations in the copper related gene *copA* have been found to drastically alter copper tolerance in *X. fastidiosa*, which is vital for agricultural survival given the frequency of copper in treating fungal infections in the vineyard, a fungicide that has been in use since the 18th century (Lafforgue, 1928; Brun *et al.*, 1998; Ge *et al.*, 2021). It is conceivable that after a host jump into a vineyard, it would be necessary for pathogens to survive higher levels of copper exposure. *DegP* has been found to be upregulated upon heat shock in *X. fastidiosa* (Koide *et al.*, 2006). *TolB* encodes for a translocation protein involved in membrane integrity and has been shown to be important for biofilm development, an important aspect of pathogenicity (Santos *et al.*, 2015). These are among a suite of other genes that are both hypothetical proteins, or just understudied in *X. fastidiosa* but show evidence of being involved in the process of this climate and host shift. Now that these genes have been identified, they are prime candidates for targeting in future experiments to determine their effects on host range and climate adaptation.

This study also includes one hypothesis that lies outside the general narrative of the introduction event, namely the evaluation of infectivity of ssp. *multiplex* towards *C. arabica*, which has not been tested before. While in California, ssp. *multiplex* has never been found infecting grapevine in the field, it has been shown to generate non persistent infections, similar to what we observed in *C. arabica*, in the greenhouse (Almeida & Purcell, 2003; Vanhove *et al.*, 2020). Recently, infections of grapevine by ssp. *fastidiosa* have been detected in the field in Virginia (Abdelrazek *et al.*, 2023). While the ssp. *multiplex* infections were not highly virulent, they were just as persistent as the ssp. *fastidiosa* strains. All three main subspecies of *X. fastidiosa* are able to infect *C. arabica* to some degree (See Figure 2).

In conclusion, we have identified a suite of genes that are related to a host switch to *Vitis* with a corresponding reduction in the ability to infect an ancestral host. These data support the hypothesis that the shift was not a host range expansion of the subspecies, but a reduction of ability to infect a former host while optimizing the ability to infect a new host species.

Tables

Table 1: SLAC results on grape only samples, list of genes under positive selection. Only named genes (not hypothetical proteins) are included in this analysis.

Gene	Annotation	Kegg	*Vitis* assoc. Bonferonni corrected p <0.05
acnM	Aconitate hydratase A	Metabolism	Yes
apaG	Protein ApaG	Unclassified: signaling and cellular processes	Yes
atpF	ATP synthase subunit b	Metabolism	Yes
clpX	ATP-dependent Clp protease ATP-binding subunit ClpX	genetic information processing	Yes
copA~ftsP	Copper resistance protein A;hypothetical protein;Cell division protein FtsP	Translocases	Yes
dapE	Succinyl-diaminopimelate desuccinylase	Metabolism	Yes
dcd	dCTP deaminase	Metabolism	Yes
ddlB_1~ddlB_2~ddl	D-alanine--D-alanine ligase B	Metabolism	Yes
degP	Periplasmic serine endoprotease DegP	Environmental Information Processing	Yes
dgkA	Diacylglycerol kinase	Metabolism	Yes
fmt	Methionyl-tRNA formyltransferase	Metabolism	No
gpsA	Glycerol-3-phosphate dehydrogenase [NAD(P)+]	Metabolism	Yes
gyrB	DNA gyrase subunit B	Protein families: genetic information processing	Yes
kdsD	Arabinose 5-phosphate isomerase KdsD	Metabolism	Yes
lacF	Lactose transport system permease protein LacF	Metabolism	Yes
leuB	3-isopropylmalate dehydrogenase	Metabolism	No
lpoA	Penicillin-binding protein activator LpoA	Poorly characterized	No
lpxC	UDP-3-O-acyl-N-acetylglucosamine deacetylase	Metabolism	Yes

mdtA_1	Multidrug resistance protein MdtA	Environmental Information Processing	Yes (mdtA_2)
ohrB	Organic hydroperoxide resistance protein OhrB	Unclassified: metabolism	Yes
oleC_2~oleC_1~ppsB	Olefin beta-lactone synthetase;Plipastatin synthase subunit B	Ligases	Yes
pilA_1	Fimbrial protein	Environmental Information Processing	No
pilA_2	Fimbrial protein	Environmental Information Processing	No
pilY1_5~pilY1_6~pilY1_1 ~ pilY1_3	Type IV pilus biogenesis factor PilY1	Protein families: signaling and cellular processes	Yes
ppsA_1~ppsA_2~ppsA	Phosphoenolpyruvate synthase	Metabolism	Yes
recG	ATP-dependent DNA helicase RecG	Genetic Information Processing	Yes
recQ	ATP-dependent DNA helicase RecQ	Genetic Information Processing	No
rpoA	DNA-directed RNA polymerase subunit alpha	Genetic Information Processing	Yes
secD_1~secD_2~secD	Protein translocase subunit SecD	Genetic Information Processing	Yes
sgcG	2-amino-4-deoxychorismate dehydrogenase	No KO assigned	No
spoT	Guanosine-3'5'-bis(diphosphate) 3'-pyrophosphohydrolase	Metabolism	No
sucC_1~sucC_2~sucC	Adenylosuccinate lyase	Poorly characterized	Yes
tatA	Sec-independent protein translocase protein TatA	Genetic Information Processing	Yes
tatC	Sec-independent protein translocase protein TatC	Genetic Information Processing	Yes
thrS	Threonine--tRNA ligase	Genetic Information Processing	Yes
tolB_2~tolB_1	Tol-Pal system protein TolB;Protein TolB	Protein families: signaling and cellular processes	Yes
ybaL	Putative cation/proton antiporter YbaL	Unclassified: signaling and cellular processes	Yes

yheS_2~yheS_1~yheS	putative ABC transporter ATP-binding protein YheS;hypothetical protein	genetic information processing	Yes

Table 2: Single nucleotide polymorphism GWAS results, including loci with best pairwise comparison P ≤ 0.125 and Bonferroni P < 0.05. Annotations and Kegg orthologies are also shown for available genes. For genes with multiple SNPs, frequencies are shown with a / between values when they vary between sites.

Locus	Bonferroni p	Best Pairwise Comp P	Frequency in strains from *Vitis*	Frequency in non-*Vitis* strains	Annotation	Kegg
azu	3.48E-27	1.25E-01	4.31E-03	5.97E-01	Azurin	No KO assigned
carA	3.48E-27	1.25E-01	4.31E-03	4.44E-01/5.97E-01	Carbamoyl-phosphate synthase small chain	Metabolism
group_1754	2.03E-18	1.25E-01	4.31E-03	6.25E-01/ 4.58E-01/ 4.44E-01	-	-
group_1454	1.34E-17	1.25E-01	4.31E-03	4.44E-01	-	-
clpP	1.34E-17	1.25E-01	4.31E-03	4.44E-01	ATP-dependent Clp protease proteolytic subunit	Metabolism
nadD	1.34E-17	1.25E-01	4.31E-03	4.44E-01	putative nicotinate-nucleotide adenylyltransferase	Metabolism
group_1675	1.34E-17	1.25E-01	4.31E-03	4.44E-01	-	-
mutY	1.34E-17	1.25E-01	4.31E-03	4.44E-01	Adenine DNA glycosylase	Genetic Information Processing
ubiJ	1.34E-17	1.25E-01	4.31E-03	4.44E-01	Ubiquinone biosynthesis accessory factor UbiJ	Metabolism
exbD_2~~~exbD_1	1.34E-17	1.25E-01	4.31E-03	4.44E-01	Biopolymer transport protein ExbD	Signaling and cellular processes
group_1746	1.80E-05	1.09E-01	1.59E-01	5.42E-01	-	-
group_418	8.62E-03	1.25E-01	7.07E-01	9.72E-01	-	-
lpxD_3~~~lpxD_4	4.54E-02	1.17E-02	8.62E-02	3.33E-01	UDP-3-O-(3-hydroxymyristoyl)glucosamine N-acyltransferase	Metabolism

Table 3: Whole gene gain and loss GWAS results, including loci with best pairwise comparison P ≤ 0.125 and Bonferroni P < 0.05. Annotations and Kegg orthologies are also shown for available genes.

Position	Gain or Loss	Bonferroni p	Best pairwise Comp p	Annotation	Kegg
group_1796	Gain	3.09E-05	7.81E-03	hypothetical protein	
group_1080	Gain	5.12E-02	1.56E-02	hypothetical protein	
group_1801	Gain	1.22E-15	3.13E-02	hypothetical protein	
group_1194	Gain	1.74E-12	3.13E-02	hypothetical protein	
group_1060	Gain	2.07E-03	3.13E-02	hypothetical protein	
xerC_1~~~xerC_2	Gain	3.50E-02	3.13E-02	Tyrosine recombinase XerC	genetic information processing
group_605	Loss	5.69E-05	3.86E-02	hypothetical protein	
group_843	Loss	1.93E-09	3.91E-02	hypothetical protein	
group_787	Gain	2.18E-06	3.91E-02	hypothetical protein	
group_1349	Gain	1.89E-12	6.25E-02	hypothetical protein	
group_1320	Gain	2.20E-07	6.25E-02	hypothetical protein	
hcaB	Gain	1.56E-05	6.25E-02	3-phenylpropionate-dihydrodiol/cinnamic acid-dihydrodiol dehydrogenase	Metabolism
group_1046	Gain	1.75E-05	6.25E-02	hypothetical protein	
group_971	Gain	5.88E-05	6.25E-02	hypothetical protein	
group_1008	Gain	8.46E-03	6.25E-02	hypothetical protein	
group_661	Gain	1.32E-02	6.25E-02	hypothetical protein	
group_925	Gain	1.32E-02	6.25E-02	hypothetical protein	
group_451	Loss	1.86E-09	7.03E-02	hypothetical protein	
group_744	Loss	1.47E-03	7.03E-02	hypothetical protein	
group_817	Gain	8.80E-12	1.09E-01	hypothetical protein	
group_299	Loss	3.10E-09	1.09E-01	hypothetical protein	
group_1656	Gain	1.29E-06	1.09E-01	hypothetical protein	

group_846	Gain	1.21E-03	1.09E-01	hypothetical protein	
group_830	Gain	3.45E-03	1.09E-01	hypothetical protein	
group_504	Gain	2.29E-02	1.09E-01	hypothetical protein	
group_509	Gain	2.21E-03	1.18E-01	hypothetical protein	
group_1001	Loss	1.84E-11	1.25E-01	hypothetical protein	
group_809	Loss	1.15E-06	1.25E-01	hypothetical protein	
group_731	Gain	1.15E-06	1.25E-01	hypothetical protein	
group_1582	Gain	5.17E-06	1.25E-01	hypothetical protein	
mdtA_3~~~mdtA_2~~~mdtA_1	Gain	3.11E-05	1.25E-01	Multidrug resistance protein MdtA	Environmental Information Processing
cnrA~~~swrC_2~~~acrF_2	Gain	3.11E-05	1.25E-01	Nickel and cobalt resistance protein CnrA;Swarming motility protein SwrC;Multidrug export protein AcrF	acrF: signaling and cellular processes
group_181	Gain	3.11E-05	1.25E-01	hypothetical protein	
group_1750	Gain	6.92E-05	1.25E-01	hypothetical protein	
group_710	Loss	9.77E-05	1.25E-01	hypothetical protein	
group_201	Loss	9.77E-05	1.25E-01	hypothetical protein	
group_1188	Loss	9.77E-05	1.25E-01	hypothetical protein	

Figures

Figure 1: Symptom assessment of *C. arabica* plants, evaluating virulence by way of measuring disease severity. The data are divided by the two *C. arabica* cultivars: Geisha and Red Catuaí as well as the symptoms evaluated: Height and Internode Length. Height was measured at 5 timepoints for Geisha and 6 timepoints for Red Catuaí. Internode Lengths were measured 3 times. Data are all clustered and colored by inoculation treatment.

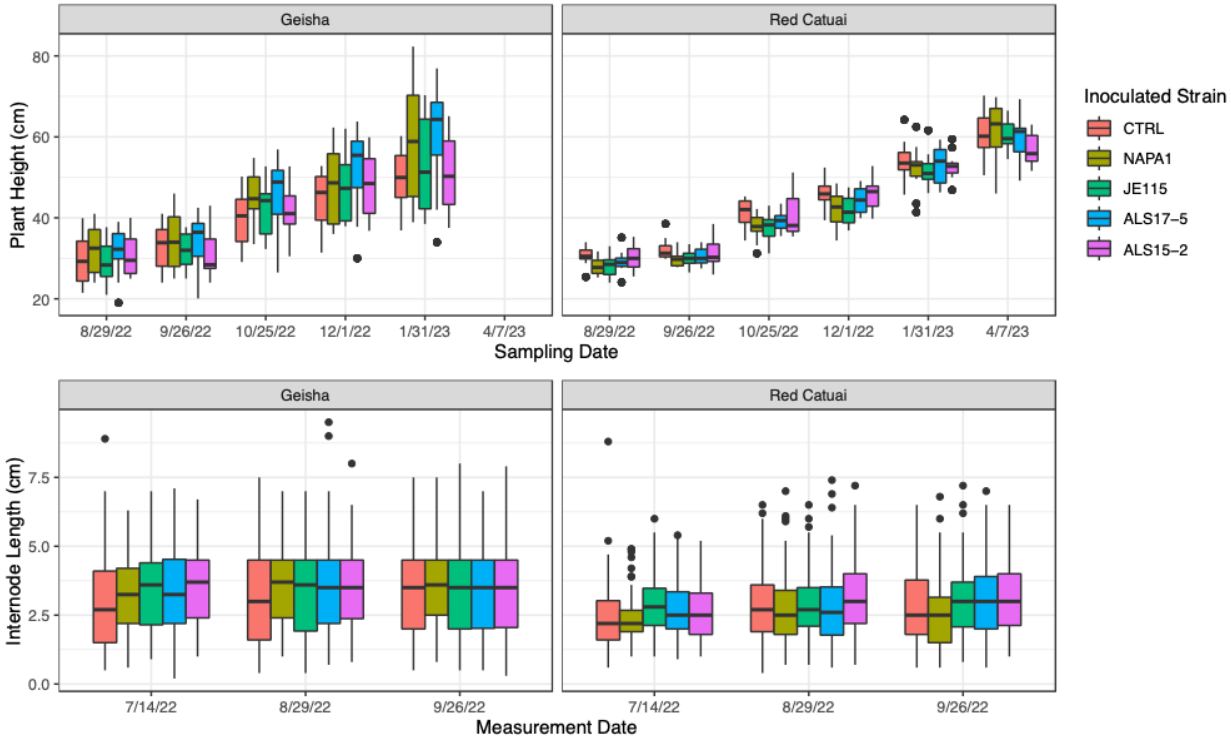


Figure 2: Detection of the strains in *C. arabica* plants based on proportion of plants testing positive at each detection point. Of the four detection points, two were conducted via qPCR (July 2022 & October 2022) and two were conducted via culturing (August 2022 & March 2023). Data are split between the two *C. arabica* cultivars, Geisha and Red Catuaí and are clustered and colored by inoculation treatment.

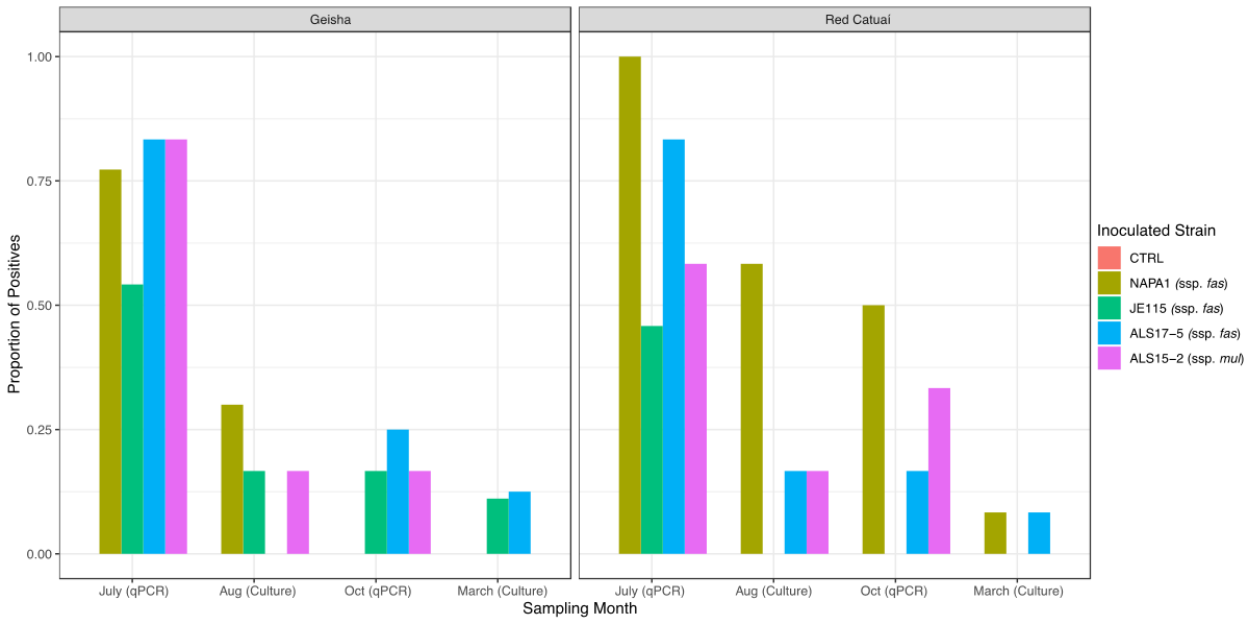
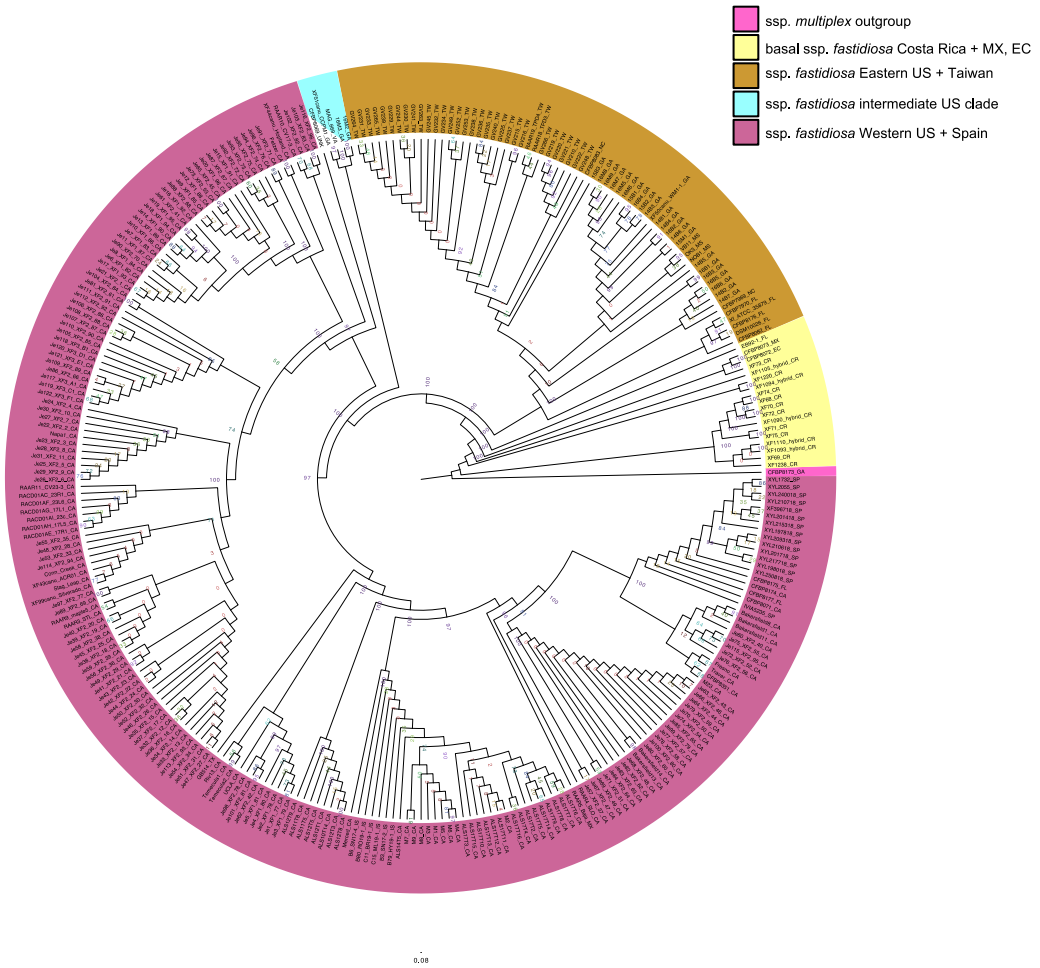


Figure 3: Phylogeny of subspecies fastidiosa with one multiplex sample, CFBP8173_GA as the outgroup (pink). At the base of the tree are samples from Costa Rica (yellow). One sample each from Mexico, Ecuador, CR, and Florida, are the closest strains to the introduction into the rest of the United States (yellow). The introduced clade includes a clade from the Eastern U.S. and an introduction into Taiwain (mustard), an intermediate clade between the east and west coasts (cyan) and the clade on the west coast which includes an introduction into Spain (purple). Bootstrap support is shown numerically at each node. Relevant geographic groupings are colored in an exterior ring around the phylogeny.



Supplementals – Chapter 2

Table S1 Metadata with sequence information: The metadata sheet includes information on all sequences used in this study. It includes the sequence name, host information, isolation year, geographic origin, and biosample information.

seq_name	Host	Isolation year	Geographic Origin (Code)	Accession
CFBP8071	P. dulcis	1987	USA, California	SAMN07998881
M23	P. dulcis	2003	USA, California	CP001011.1
B3 (SN17-2)	P. dulcis	2017	Israel	To be published
B6 (SN17-4)	P. dulcis	2017	Israel	To be published
XYL197818	P. dulcis	2018	Spain	11018302
XYL198018	P. dulcis	2018	Spain	14052381
XYL201418	P. dulcis	2018	Spain	14052501
XYL201718	P. dulcis	2018	Spain	14052502
XYL209318	P. dulcis	2018	Spain	14052504
XYL210618	P. dulcis	2018	Spain	14052529
XYL210718	P. dulcis	2018	Spain	14052539
B79 (HY19-1)	P. dulcis	2019	Israel	To be published
B80 (RO19-1)	P. dulcis	2019	Israel	To be published
C11 (BR19-1)	P. dulcis	2019	Israel	To be published
C12 (BL19-1)	P. dulcis	2019	Israel	To be published
C15 (ML19-1)	P. dulcis	2019	Israel	To be published
M1	P. dulcis	2019	USA, California	To be published
M2	P. dulcis	2019	USA, California	To be published
M3	P. dulcis	2019	USA, California	To be published
M4	P. dulcis	2019	USA, California	To be published
M5	P. dulcis	2019	USA, California	To be published
M6	P. dulcis	2019	USA, California	To be published
M7	P. dulcis	2019	USA, California	To be published
M8	P. dulcis	2019	USA, California	To be published
M9	P. dulcis	2019	USA, California	To be published
ALS10T14	P. dulcis	2020	USA, California	To be published
ALS11T5	P. dulcis	2020	USA, California	To be published
ALS11T8	P. dulcis	2020	USA, California	To be published
ALS12T1	P. dulcis	2020	USA, California	To be published
ALS12T3	P. dulcis	2020	USA, California	To be published
ALS12T5	P. dulcis	2020	USA, California	To be published
ALS12T8	P. dulcis	2020	USA, California	To be published
ALS12T9	P. dulcis	2020	USA, California	To be published
ALS14T5	P. dulcis	2020	USA, California	To be published
ALS17T10	P. dulcis	2020	USA, California	To be published
ALS17T11	P. dulcis	2020	USA, California	To be published
ALS17T12	P. dulcis	2020	USA, California	To be published
ALS17T13	P. dulcis	2020	USA, California	To be published
ALS17T14	P. dulcis	2020	USA, California	To be published
ALS17T15	P. dulcis	2020	USA, California	To be published
ALS17T16	P. dulcis	2020	USA, California	To be published
ALS17T2	P. dulcis	2020	USA, California	To be published
ALS17T3	P. dulcis	2020	USA, California	To be published
ALS17T4	P. dulcis	2020	USA, California	To be published
ALS17T5	P. dulcis	2020	USA, California	To be published
ALS17T6	P. dulcis	2020	USA, California	To be published
ALS17T7	P. dulcis	2020	USA, California	To be published
ALS17T8	P. dulcis	2020	USA, California	To be published
ALS17T9	P. dulcis	2020	USA, California	To be published
Fresno	P. dulcis	1995	USA, California	27988137
CFBP8072	Coffea arabica	2012	Ecuador	LKDK00000000.1
CFBP8073	Coffea canephora	2012	Mexico	LKES00000000.1
XF69	Coffea sp.	2016	Costa Rica	To be published
XF70	Coffea sp.	2016	Costa Rica	SAMN12994818
XF71	Coffea sp.	2016	Costa Rica	SAMN12994819

XF72	Coffea sp.	2016	Costa Rica	SAMN12994820
XF73	Coffea sp.	2016	Costa Rica	SAMN12994821
XF74	Coffea sp.	2016	Costa Rica	SAMN12994822
XF75	Coffea sp.	2016	Costa Rica	SAMN12994823
XF1090	Coffea sp.	2017	Costa Rica	SAMN12994824
XF1093	Coffea sp.	2017	Costa Rica	SAMN12994825
XF1105	Coffea sp.	2017	Costa Rica	SAMN12994827
XF1220	Coffea sp.	2018	Costa Rica	To be published
XF1238	Coffea sp.	2018	Costa Rica	To be published
EB92-1	Sambucus sp.	1992	USA, Florida	AFDJ00000000.1
NOB1	Vitis rotundifolia	2019	USA, Mississippi	JABCJG000000000
CFBP7969	Vitis rotundifolia cv Carlos	1985	USA, North Carolina	SAMN07998892
CFBP7970	Vitis sp.	1987	USA, Florida	SAMN07998897
TPD3	Vitis sp.	2012	Taiwan	SAMN12097273
TPD4	Vitis sp.	2012	Taiwan	SAMN12097274
CFBP8083	V. vinifera	1985	USA, North Carolina	27988126
DSM10026	V. vinifera	1987	USA, Florida	FQWN00000000.1
Xf_ATCC_35879	V. vinifera	1987	USA, Florida	SAMN02997312
CFBP8351	V. vinifera	1993	USA, California	SAMN07999362
SLO	V. vinifera	1994	USA, California	SAMN09941414
STL	V. vinifera	1998	USA, California	SAMN09941416
Temecula1	V. vinifera	1998	USA, California	AE009442.1
RAAR5_Baja	V. vinifera	2001	Mexico	SAMN09941403
Stags_Leap	V. vinifera	2005	USA, California	LSMJ00000000.1
GB514	V. vinifera	2006	USA, Texas	CP002165.1
Xf50 (WM1-1)	V. vinifera	2009	USA, Georgia	SAMN09941434
Xf51 (CCPM1)	V. vinifera	2010	USA, Georgia	SAMN09941435
RAAR10_CV17-3	V. vinifera	2011	USA, California	To be published
RAAR11_CV23 (exp. inoculation in field of ACR01)	V. vinifera	2011	USA, California	SAMN09941404
Xf43_ACR-01	V. vinifera	2011	USA, California	SAMN09941432
RACD01AC (23R1)	V. vinifera	2012	USA, California	SAMN09941408
RACD01AE (17R1)	V. vinifera	2012	USA, California	SAMN09941409
RACD01AF (23L6)	V. vinifera	2012	USA, California	SAMN09941410
RACD01AG (17L1)	V. vinifera	2012	USA, California	SAMN09941411
RACD01AH (17L5)	V. vinifera	2012	USA, California	SAMN09941412
RACD01AI (23c)	V. vinifera	2012	USA, California	SAMN09941413
14B1	V. vinifera	2014	USA, Georgia	SAMN15732826
14B2	V. vinifera	2014	USA, Georgia	SAMN15732827
14B3	V. vinifera	2014	USA, Georgia	SAMN15732828
14B4	V. vinifera	2014	USA, Georgia	SAMN15732829
14B5	V. vinifera	2014	USA, Georgia	SAMN15732830
14B6	V. vinifera	2014	USA, Georgia	SAMN15732831
14B7	V. vinifera	2014	USA, Georgia	SAMN15732832
GV210	V. vinifera	2015	Taiwan	SAMN18344333
Napa1	V. vinifera	2015	USA, California	SAMN09941406
15B1	V. vinifera	2015	USA, Georgia	SAMN15732833
15B2	V. vinifera	2015	USA, Georgia	SAMN15732834
15B3	V. vinifera	2015	USA, Georgia	SAMN15732835
15M1	V. vinifera	2015	USA, Georgia	SAMN15732837
Bakersfield-1	V. vinifera	2016	USA, California	SAMN11914553
Bakersfield-8	V. vinifera	2016	USA, California	SAMN16582176
Je1	V. vinifera	2016	USA, California	10358662
Je10	V. vinifera	2016	USA, California	SAMN10358663
Je100	V. vinifera	2016	USA, California	SAMN10358664
Je101	V. vinifera	2016	USA, California	SAMN10358665
Je102	V. vinifera	2016	USA, California	SAMN10358666
Je103	V. vinifera	2016	USA, California	SAMN10358667
Je104	V. vinifera	2016	USA, California	SAMN10358668
Je105	V. vinifera	2016	USA, California	SAMN10358669
Je106	V. vinifera	2016	USA, California	SAMN10358670
Je107	V. vinifera	2016	USA, California	SAMN10358671
Je108	V. vinifera	2016	USA, California	SAMN10358672
Je109	V. vinifera	2016	USA, California	SAMN10358673

Je11	V. vinifera	2016	USA, California	SAMN10358674
Je110	V. vinifera	2016	USA, California	SAMN10358675
Je111	V. vinifera	2016	USA, California	SAMN10358676
Je112	V. vinifera	2016	USA, California	SAMN10358677
Je113	V. vinifera	2016	USA, California	SAMN10358678
Je114	V. vinifera	2016	USA, California	SAMN10358679
Je115	V. vinifera	2016	USA, California	SAMN10358680
Je116	V. vinifera	2016	USA, California	SAMN10358681
Je117	V. vinifera	2016	USA, California	SAMN10358682
Je118	V. vinifera	2016	USA, California	SAMN10358683
Je119	V. vinifera	2016	USA, California	SAMN10358684
Je12	V. vinifera	2016	USA, California	SAMN10358685
Je120	V. vinifera	2016	USA, California	SAMN10358686
Je121	V. vinifera	2016	USA, California	SAMN10358687
Je122	V. vinifera	2016	USA, California	SAMN10358688
Je13	V. vinifera	2016	USA, California	SAMN10358689
Je14	V. vinifera	2016	USA, California	SAMN10358690
Je15	V. vinifera	2016	USA, California	SAMN10358691
Je16	V. vinifera	2016	USA, California	SAMN10358692
Je17	V. vinifera	2016	USA, California	SAMN10358693
Je18	V. vinifera	2016	USA, California	SAMN10358694
Je19	V. vinifera	2016	USA, California	SAMN10358695
Je2	V. vinifera	2016	USA, California	SAMN10358696
Je20	V. vinifera	2016	USA, California	SAMN10358697
Je21	V. vinifera	2016	USA, California	SAMN10358698
Je22	V. vinifera	2016	USA, California	SAMN10358699
Je23	V. vinifera	2016	USA, California	SAMN10358700
Je24	V. vinifera	2016	USA, California	SAMN10358701
Je25	V. vinifera	2016	USA, California	SAMN10358702
Je26	V. vinifera	2016	USA, California	SAMN10358703
Je27	V. vinifera	2016	USA, California	SAMN10358704
Je28	V. vinifera	2016	USA, California	SAMN10358705
Je29	V. vinifera	2016	USA, California	SAMN10358706
Je3	V. vinifera	2016	USA, California	SAMN10358707
Je30	V. vinifera	2016	USA, California	SAMN10358708
Je31	V. vinifera	2016	USA, California	SAMN10358709
Je32	V. vinifera	2016	USA, California	SAMN10358710
Je33	V. vinifera	2016	USA, California	SAMN10358711
Je34	V. vinifera	2016	USA, California	SAMN10358712
Je35	V. vinifera	2016	USA, California	SAMN10358713
Je36	V. vinifera	2016	USA, California	SAMN10358714
Je37	V. vinifera	2016	USA, California	SAMN10358715
Je38	V. vinifera	2016	USA, California	SAMN10358716
Je39	V. vinifera	2016	USA, California	SAMN10358717
Je4	V. vinifera	2016	USA, California	SAMN10358718
Je40	V. vinifera	2016	USA, California	SAMN10358719
Je41	V. vinifera	2016	USA, California	SAMN10358720
Je42	V. vinifera	2016	USA, California	SAMN10358721
Je43	V. vinifera	2016	USA, California	SAMN10358722
Je44	V. vinifera	2016	USA, California	SAMN10358723
Je45	V. vinifera	2016	USA, California	SAMN10358724
Je46	V. vinifera	2016	USA, California	SAMN10358725
Je47	V. vinifera	2016	USA, California	SAMN10358726
Je48	V. vinifera	2016	USA, California	SAMN10358727
Je49	V. vinifera	2016	USA, California	SAMN10358728
Je5	V. vinifera	2016	USA, California	SAMN10358729
Je50	V. vinifera	2016	USA, California	SAMN10358730
Je51	V. vinifera	2016	USA, California	SAMN10358731
Je52	V. vinifera	2016	USA, California	SAMN10358732
Je53	V. vinifera	2016	USA, California	SAMN10358733
Je54	V. vinifera	2016	USA, California	SAMN10358734
Je55	V. vinifera	2016	USA, California	SAMN10358735
Je56	V. vinifera	2016	USA, California	SAMN10358736
Je57	V. vinifera	2016	USA, California	SAMN10358737
Je58	V. vinifera	2016	USA, California	SAMN10358738
Je59	V. vinifera	2016	USA, California	SAMN10358739

Je6	V. vinifera	2016	USA, California	SAMN10358740
Je60	V. vinifera	2016	USA, California	SAMN10358741
Je61	V. vinifera	2016	USA, California	SAMN10358742
Je62	V. vinifera	2016	USA, California	SAMN10358743
Je63	V. vinifera	2016	USA, California	SAMN10358744
Je64	V. vinifera	2016	USA, California	SAMN10358745
Je65	V. vinifera	2016	USA, California	SAMN10358746
Je66	V. vinifera	2016	USA, California	SAMN10358747
Je67	V. vinifera	2016	USA, California	SAMN10358748
Je68	V. vinifera	2016	USA, California	SAMN10358749
Je69	V. vinifera	2016	USA, California	SAMN10358750
Je7	V. vinifera	2016	USA, California	SAMN10358751
Je70	V. vinifera	2016	USA, California	SAMN10358752
Je71	V. vinifera	2016	USA, California	SAMN10358753
Je72	V. vinifera	2016	USA, California	SAMN10358754
Je73	V. vinifera	2016	USA, California	SAMN10358755
Je74	V. vinifera	2016	USA, California	SAMN10358756
Je75	V. vinifera	2016	USA, California	SAMN10358757
Je76	V. vinifera	2016	USA, California	SAMN10358758
Je77	V. vinifera	2016	USA, California	SAMN10358759
Je78	V. vinifera	2016	USA, California	SAMN10358760
Je79	V. vinifera	2016	USA, California	SAMN10358761
Je8	V. vinifera	2016	USA, California	SAMN10358762
Je80	V. vinifera	2016	USA, California	SAMN10358763
Je81	V. vinifera	2016	USA, California	SAMN10358764
Je82	V. vinifera	2016	USA, California	SAMN10358765
Je83	V. vinifera	2016	USA, California	SAMN10358766
Je84	V. vinifera	2016	USA, California	SAMN10358767
Je85	V. vinifera	2016	USA, California	SAMN10358768
Je86	V. vinifera	2016	USA, California	SAMN10358769
Je87	V. vinifera	2016	USA, California	SAMN10358770
Je88	V. vinifera	2016	USA, California	10358771
Je89	V. vinifera	2016	USA, California	SAMN10358772
Je9	V. vinifera	2016	USA, California	SAMN10358773
Je90	V. vinifera	2016	USA, California	SAMN10358774
Je91	V. vinifera	2016	USA, California	SAMN10358775
Je92	V. vinifera	2016	USA, California	SAMN10358776
Je93	V. vinifera	2016	USA, California	SAMN10358777
Je94	V. vinifera	2016	USA, California	SAMN10358778
Je95	V. vinifera	2016	USA, California	SAMN10358779
Je96	V. vinifera	2016	USA, California	SAMN10358780
Je97	V. vinifera	2016	USA, California	SAMN10358781
Je98	V. vinifera	2016	USA, California	SAMN10358782
Je99	V. vinifera	2016	USA, California	SAMN10358783
16B1	V. vinifera	2016	USA, Georgia	SAMN15732838
16B2	V. vinifera	2016	USA, Georgia	SAMN15732839
16B3	V. vinifera	2016	USA, Georgia	SAMN15732840
16B4	V. vinifera	2016	USA, Georgia	SAMN15732841
16B5	V. vinifera	2016	USA, Georgia	SAMN15732842
16B6	V. vinifera	2016	USA, Georgia	SAMN15732836
16M2	V. vinifera	2016	USA, Georgia	SAMN15732843
16M3	V. vinifera	2016	USA, Georgia	SAMN15732844
16M5	V. vinifera	2016	USA, Georgia	SAMN15732845
16M6	V. vinifera	2016	USA, Georgia	SAMN15732846
16M7	V. vinifera	2016	USA, Georgia	SAMN15732847
16M8	V. vinifera	2016	USA, Georgia	SAMN15732848
16M9	V. vinifera	2016	USA, Georgia	SAMN15732849
XYL1732	V. vinifera	2017	Spain	SAMN09767243
XYL2055	V. vinifera	2017	Spain	SAMN09767242
GV215	V. vinifera	2017	Taiwan	SAMN18344334
GV216	V. vinifera	2017	Taiwan	SAMN18344335
GV219	V. vinifera	2017	Taiwan	SAMN18344336
GV220	V. vinifera	2017	Taiwan	SAMN18344337
GV221	V. vinifera	2017	Taiwan	SAMN18344338
GV222	V. vinifera	2017	Taiwan	SAMN18344339
GV225	V. vinifera	2017	Taiwan	SAMN18344340

GV229	V. vinifera	2017	Taiwan	SAMN18344341
Bakersfield-11	V. vinifera	2017	USA, California	SAMN16582177
Bakersfield-13	V. vinifera	2017	USA, California	SAMN16582178
Bakersfield-14	V. vinifera	2017	USA, California	SAMN16582179
XF396718	V. vinifera	2018	Spain	14052556
XYL215318	V. vinifera	2018	Spain	14052541
XYL217718	V. vinifera	2018	Spain	14052551
XYL240018	V. vinifera	2018	Spain	14052552
XYL250818	V. vinifera	2018	Spain	14052553
GV230	V. vinifera	2018	Taiwan	SAMN15783671
GV231	V. vinifera	2018	Taiwan	SAMN18344342
GV232	V. vinifera	2018	Taiwan	SAMN18344343
GV233	V. vinifera	2018	Taiwan	SAMN18344344
GV234	V. vinifera	2018	Taiwan	SAMN18344345
GV235	V. vinifera	2018	Taiwan	SAMN18344346
GV236	V. vinifera	2018	Taiwan	SAMN18344347
GV237	V. vinifera	2018	Taiwan	SAMN18344348
GV238	V. vinifera	2018	Taiwan	SAMN18344349
GV239	V. vinifera	2018	Taiwan	SAMN18344350
GV240	V. vinifera	2018	Taiwan	SAMN18344351
GV241	V. vinifera	2018	Taiwan	SAMN18344352
GV244	V. vinifera	2018	Taiwan	SAMN18344353
GV245	V. vinifera	2018	Taiwan	SAMN18344354
GV248	V. vinifera	2019	Taiwan	SAMN18344355
GV249	V. vinifera	2019	Taiwan	SAMN18344356
GV252	V. vinifera	2019	Taiwan	SAMN18344357
GV253	V. vinifera	2019	Taiwan	SAMN18344358
GV263	V. vinifera	2019	Taiwan	SAMN18344359
GV264	V. vinifera	2019	Taiwan	SAMN18344360
GV265	V. vinifera	2019	Taiwan	SAMN18344361
GV266	V. vinifera	2019	Taiwan	SAMN18344362
OK3	V. vinifera	2019	USA, Mississippi	JABCJH000000000
VB11	V. vinifera	2019	USA, Mississippi	JABCJI000000000
MAG_669	V. vinifera	2020	USA, Virginia (VA)	22645440
Conn_Creek	V. vinifera	1995	USA, California	27988135
Fetzer	V. vinifera	1995	USA, California	SAMN27988136
Merced	V. vinifera	1995	USA, California	27988138
Temecula2	V. vinifera	1995	USA, California	27988139
Traver	V. vinifera	1995	USA, California	27988140
UCLA	V. vinifera	Pre 1982	USA, California	27988141
CFBP8069	V. vinifera	Pre 1989	Unknown	27988122
CFBP8174	V. vinifera	Pre 1981	USA, California	27988129
Xf44_Hopland	V. vinifera	1990-1999	USA, California	SAMN09941433
Xf99_Silverado	V. vinifera	1990-1999	USA, California	SAMN09941436
CFBP8175	V. vinifera	1985	USA, Florida	27988130
CFBP8176	V. vinifera	1985	USA, Florida	27988131
CFBP8177	V. vinifera	1985	USA, Florida	27988132
XF68	Psidium sp.	2012	Costa Rica	SAMN12994817
maple5	Acer sp.	2001	USA, California	SAMN09941405
XF1094	Vinca sp.	2017	Costa Rica	SAMN12994826
XF1110	Vinca sp.	2017	Costa Rica	SAMN12994828
CFBP8173	Prunus domestica	1983	USA, Georgia	27988128
CFBP8082	Ambrosia artemisiifolia	1983	USA, Florida	SAMN07999361
IVIA5235	Prunus avium	2017	Spain	CM010656.1
Riv13	Cercis occidentalis	2006	USA, California	24371049

Figure s1: Sunflower and grape positive controls. 4 Chardonnay 1 year old rooted cuttings were inoculated per treatment and 10 sunflower plants per treatment. Mean Ct values of positive samples (Ct < 37) are plotted and colors are scaled to show the highest bacterial population (lowest Ct) in dark green, up to the lowest populations (highest Ct) in pale green. ALS15-2 is the one subspecies *multiplex* strain used for the inoculations.

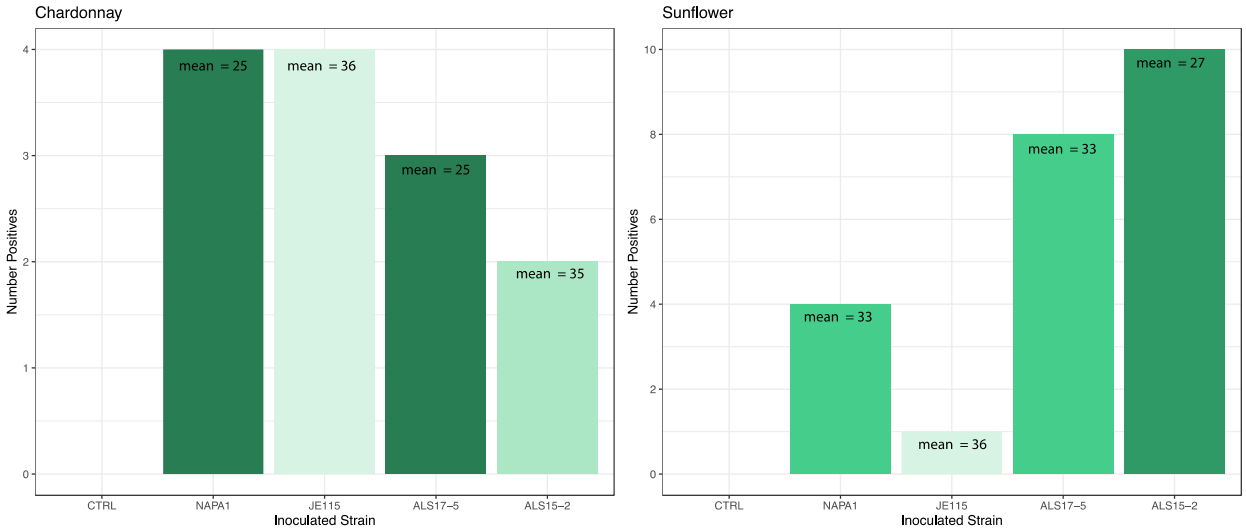
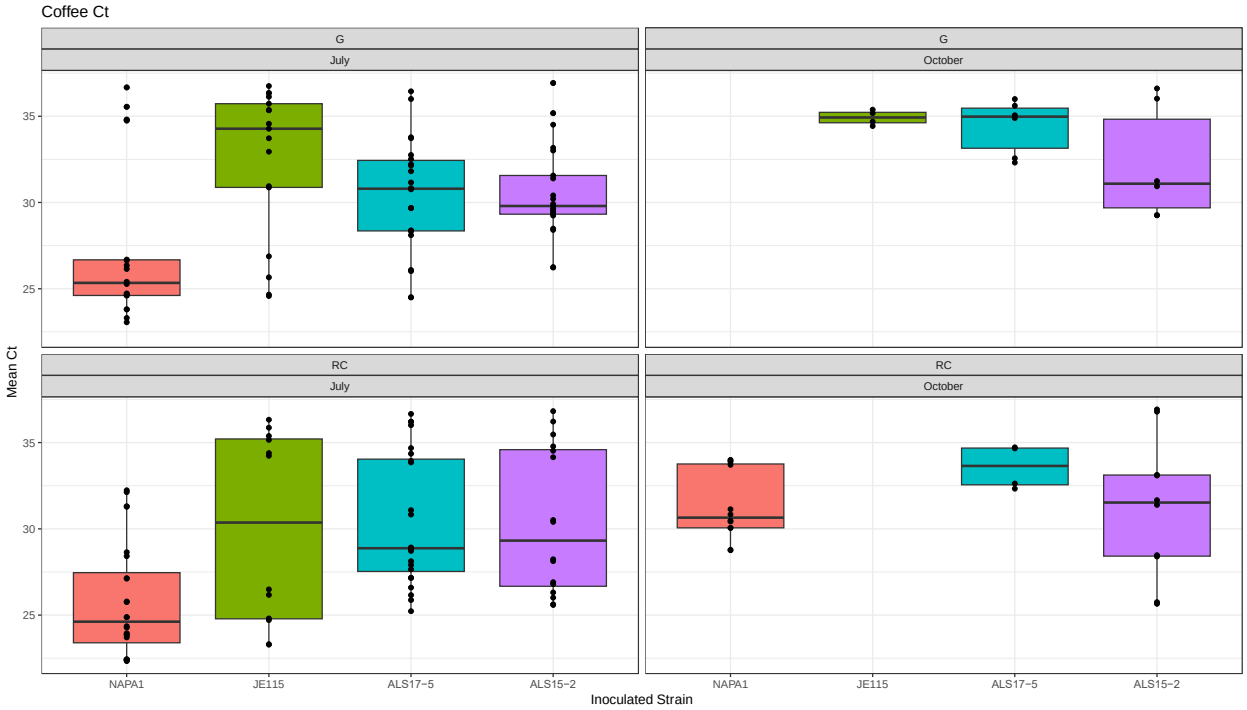


Figure S2: Mean Ct values from positive (Ct<37) *C. arabica* plants qPCR sampled in July and October, 2022. In July, one leaf from each plant was sampled 1 internode above the inoculation point, and in October one leaf from each plant was sampled 2 internodes above the inoculation point.



CHAPTER THREE

Progression of *Xylella fastidiosa* infection in grapevines under field conditions

Alexandra K. Kahn[1], Anne Sicard[1]*, Monica L. Cooper[2], Matthew P. Daugherty[3], Monica A. Donegan [1], and Rodrigo P.P. Almeida[1]

[1] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720
* INRAE, Université de Strasbourg, SVQV UMR-1131, Colmar, France
[2] Division of Agriculture and Natural Resources, University of California, Cooperative Extension, Napa, CA 94559
[3] Department of Entomology, Univ. of California, Riverside, 900 University Ave., CA 92501
Corresponding author: rodrigoalmeida@berkeley.edu

Abstract

The pathogen *Xylella fastidiosa* subspecies *fastidiosa* has circulated through California's vineyards since its introduction from Central America in the 1800s. This pathogen is responsible for a bacterial disease called Pierce's disease of grapevine (PD). With no known cure, PD has had devastating effects on some vineyards. Important factors that impact disease severity and persistence include: the presence of insect vectors, grapevine cultivar, management, ecology, and winter temperatures. Removal of infected vines is critical for reducing pathogen spread but relies on accurate and rapid pathogen detection. In this study, we foster a greater understanding of disease symptom emergence by way of a three-year field inoculation project in Napa Valley. Although PD emergence and symptom progression have been studied in greenhouse and experimental plots, there is a large knowledge gap in quantifying disease progression under commercial conditions. After inoculating 80 mature *Vitis vinifera* vines in April 2017, bacterial populations and six symptom types were measured at four locations within each plant, throughout the subsequent 3 growing seasons. The main foci of the project were understanding *X. fastidiosa* movement through the plants, infection, overwinter curing, and symptom development. We observed greater winter recovery than expected, and shriveled grape clusters proved to be a more reliable early indication of infection than other more commonly used symptoms. Although there were differences among wine grape cultivars, this work suggests that disease progression in the field may not fit the paradigm of predominant leaf scorch and low recovery rates as neatly as has been previously believed.

Introduction

The results of in-situ experiments can differ greatly from those conducted in controlled laboratory conditions, especially in applied biological systems. In pathology specifically, there are often ethical and regulatory hurdles that prevent the introduction of disease-causing pathogens into naïve systems, whether it is for human health, food security, or in natural systems. Researchers are therefore typically constrained to using controlled or model systems that are ethically, financially, and logistically feasible. For example, mice are often used to understand human disease, and quarantine greenhouses are used to study plant disease (Perlman, 2016). These proxy systems are imperfect, as we intuitively know that there are differences between a mouse and a human, so there are differences between the potted plant on your windowsill and the towering redwoods of California. These differences are intuitive, but also biological as the effects of interactions between organisms and their complex environments are impossible to recreate, and these environmental and physiological differences directly affect the outcomes of research in plant pathology (Velásquez *et al.*, 2018). The chances to experimentally infect plants in accurate conditions are few and far between, so rare opportunities provide critical research windows.

One such system that suffers from a lack of realistic experimental control is Pierce's disease of grapevines (PD), a persistent burden to the vineyards of California since the introduction of the etiological agent (*Xylella fastidiosa* subsp. *fastidiosa*) into North America in the late 1800s (Nunney *et al.*, 2010; Vanhove *et al.*, 2019). The pathogenic bacterial strains that colonized *Vitis vinifera* and cause PD have since spread to Taiwan, Europe, and the Eastern United States (Landa *et al.*, 2019; Castillo *et al.*, 2021a). Although the relationship between *Xylella fastidiosa* and PD has been understood since 1978 (Davis *et al.*, 1978), this study is the first to document the progression of symptoms in mature commercial grapevines under field conditions over the course of several years with a known inoculation time and location on the plant. Despite the challenges of conducting this work in-situ, predominantly concern from the proprietors about infection spreading, we have made insights that counter much of the classic understanding of this disease system and deepen our understanding of symptom development and bacterial multiplication and movement in PD.

Several studies have documented PD, which collectively create expectations for the progression of this disease. The pathogen is known to move through mature grapevines at least fast enough to enter the cordon from the shoot within the first year of infection (Feil *et al.*, 2003). However, given that removal of the cordon does not reduce disease severity in subsequent years (Daugherty *et al.*, 2018), the pathogen must be moving further into the plant, presumably into the trunk during the first year. It is also well established that there are differences in response to *X. fastidiosa* infection based on *Vitis vinifera* cultivar, with some cultivars more susceptible than others (Raju & Goheen, 1981; Krivanek *et al.*, 2005; Lieth *et al.*, 2011). Symptoms are presumed to first appear late in the season that infection occurred, except in especially susceptible cultivars, such as Chardonnay, in which they can arise sooner. Characteristic symptoms have been described as leaf scorch, uneven lignification of the shoots, shriveled berries, and "matchstick" petioles, a situation in which the leaf blade becomes detached from the petiole. In this system, disease symptoms and pathogen infection do not always persist through the winter, a phenomenon called overwinter curing. However, the mechanism for recovery has not yet been determined, and current explanations range from pathogen temperature susceptibility to plant defensive responses. Expectations for overwinter curing are both temperature and cultivar

dependent, and in Napa Valley, the expected recovery rate for early-season inoculations is around 30% (Feil *et al.*, 2003; Burbank, 2018).

Progression of PD symptoms has been monitored under controlled conditions (Stevenson *et al.*, 2004; Krivanek *et al.*, 2005). However, such data may not translate to plants grown in the vineyard for various biotic and abiotic reasons. Greenhouse conditions do not account for the complex nutritional, pest, and other biotic (such as the microbiome) (Deyett & Rolshausen, 2019) and abiotic pressures of the field (Poorter *et al.*, 2016), and greenhouse assays of grapevines are typically conducted with an excised shoot, rather than a mature plant. Observational studies are similarly limited because they often miss aspects such as recovery, asymptomatic infections, or unexpected symptoms by not relying on controls, but instead seeking out expected symptoms. These studies have given us a view on characteristics of PD and its progression that we were able to experimentally test. Based on the available observational and greenhouse data, we set out to test the following predictions:

1. Pathogen movement and multiplication post inoculation: By early fall of the inoculation year, symptoms were expected throughout the infected shoot, as well as one or two adjacent shoots, indicating likely spread of the pathogen to those areas (Feil *et al.*, 2003).
2. Symptom expression over time: Leaf scorch as early as 3 months post inoculation has been reported to be the first observable symptom (Krivanek *et al.*, 2005). Based on this we expected leaf scorch beginning mid-late July 2017 as the earliest detectable and most prevalent symptom in year 1.
3. Overwinter curing rates: The paradigm assumes variation in overwinter curing rates based on grapevine cultivar and winter temperatures, and necessitates early-season inoculation to avoid curing (Feil *et al.*, 2003; Burbank, 2018). According to these papers, in Napa County we should expect approximately a 30% recovery rate for Chardonnay, with variability between cultivars shown through cold-treatment greenhouse experiments.
4. Cultivar differences: *V. vinifera* cv. Cabernet Sauvignon is described as less susceptible than Chardonnay and Merlot, which are predicted to have lower recovery rates associated with their susceptibility (Purcell, 1981; Feil & Purcell, 2001; Burbank, 2018).

For a system that has been studied for so long, knowledge gaps persist regarding how the pathogen and host interact to cause disease. This study adds some clarity to gaps surrounding early symptom development and overwinter curing rates. Using a strain of *X. fastidiosa* isolated from Napa County, we tested the effect of this pathogen on symptom development, described symptom expression over time, and followed pathogen movement and multiplication within the vine.

Materials and Methods

Study sites and Vitis vinifera

Three commercial vineyard blocks were selected as trial sites at two locations in Napa Valley. Plots were set up with permission from vineyard managers; however, the identity and specific locations of sites must be kept confidential. Each block was planted with one of three wine grape cultivars, *Vitis vinifera* cv. Cabernet Sauvignon, Chardonnay or Merlot. The Merlot and Chardonnay blocks were in the Oak Knoll District and the Cabernet Sauvignon block in Calistoga. Given the arrangement of three cultivars at two sites, cultivar and site effects are confounded in this design. Due to the close location of the two sites, these effects are analyzed as cultivar effects with the assumption that the site effects are not large. The Chardonnay and Merlot blocks were planted in 2008, on 101-14 rootstock, and trained to a bilateral system. The Cabernet Sauvignon block was planted in 2006, also on 101-14 rootstock and trained to a quadrilateral system. All blocks were cane-pruned (See Table 1 for summary of study sites).

Inoculation of pathogen X. fastidiosa

Each site included equal numbers of treatment vines (inoculated with sterile succinate-citrate-phosphate buffer (SCP) (Hopkins, 1984) & *X. fastidios*a (Napa isolate strain "Napa1") and control vines (mock inoculated with SCP only) (See Table 1). Napa1 was isolated in 2015 in Napa Valley, and was sequenced and published as BioSample SAMN09941406 (Vanhove *et al.*, 2020).

Vines were inoculated on April 14, 2017, (Chardonnay and Merlot) and April 28, 2017 (Cabernet Sauvignon). Cells were grown on Periwinkle wilt-GelRite (PWG) solid medium (Hill and Purcell 1995) for one week and suspended in succinate-citrate-phosphate buffer (SCP) in the field just prior to inoculation. The inoculations were at the terminal shoot of the west cane, where 10 µl droplets were needle inoculated into 2 inoculation points, basal and middle of the shoot, by pin pricking through the bead of cell suspension until absorbed into the plant xylem. Timing of infection was selected to match natural spring infections by the local vector *Graphocephala atropunctata* (Hemiptera, Cicadellidae) (Purcell, 1975; Feil *et al.*, 2003).

Monitoring and sampling

Following inoculation, the vines were monitored at 14-day intervals for the remainder of the 2017 growing season. Subsequently, in the 2018 and 2019 growing seasons, the vines were monitored from bud break through leaf fall. On each monitoring date, the vines were visually assessed for common symptoms of PD: stunted growth, chlorotic leaves, shriveled berries, leaf scorch, incomplete lignification, and matchstick petioles. Each symptom was recorded as present or absent on each of 5 areas of the vine: (1) west and (2) east, corresponding to the terminal ends of each cane; (3) mid-west and (4) mid-east, corresponding to the middle of each cane; and (5) center corresponding to the head of the vine, where the canes originate. Shriveled berries could not be evaluated after the harvest date for each block. For Cabernet Sauvignon, the harvest date for 2017 was October 27th, 2017 and for 2018 – 2019 the harvest was after the completion of our data collection. For Merlot, the harvest dates were September 13th, 2017, September 25th, 2018, and October 8th, 2019. For Chardonnay, the harvest dates were September 13th, 2017, September 25th, 2018, and September 9, 2019.

Three times annually (May, July, and September), petioles were collected from all vines. The basal petiole from the terminal shoots (west and east) as well as from the shoots adjacent to the trunk (west-center and east-center) were collected at each time point (four samples per plant). An additional sampling point was also measured in only July 2017 for finer resolution on the first year's movement: shoot 2 was measured, the next shoot over from the inoculation site. For

the quadrilaterally trained vines, eight samples were taken per plant instead of four, with additional designation as originating from either the south or north side of the vine.

DNA Extraction and qPCR

Petiole samples were analyzed via DNA extraction and quantitative polymerase chain reaction (qPCR) to quantify *X. fastidiosa* population size (in colony forming units (CFU) per gram of petiole tissue) and population distribution across the vines. DNA was extracted from petioles using the DNeasy plant mini kit (Qiagen). All extractions were 5-fold diluted with RNAse free water before being quantified by qPCR wherein they were run using a primer pair targeting the gene encoding *RecF*, RecF1_F+R (Sicard et al., 2020); for qPCR protocol see Sicard et al. 2020. Following analyses with LinRegPCR (Ramakers *et al.*, 2003), bacterial populations were interpolated using a standard curve. To produce the standard curve, we grew Napa1 cells on PWG media in a series of 10x dilutions and plated D0-D8 while also keeping the same volumes for qPCR. CFU/g were counted per streak of 20 µl and compared to the C_t values from qPCR to generate a slope to calculate CFU/g based on C_t values. qPCR was run in duplicate with positive and negative controls on each plate; any plates that had duplicates of more than 1 cycle apart were re-run. All samples with a Ct value greater than 36 are marked as undetected.

Statistical analysis

All analyses were performed using R statistical software version 4.1.2 (R Core Team, 2021). We used generalized linear models (GLM) with binomial error to build several models. (1) First we tested grapevine cultivar as a factor predictor for inoculation success (measured as the count of positive plants from July, 2017, 1 for positive, 0 for negative). The only predictor value used was cultivar which included three levels (Merlot, Chardonnay, and Cabernet Sauvignon). (2) We also tested the effect of pathogen presence (1/0) on symptom detection (1/0) with a GLMM for each symptom with a random effect of plant identity. We then computed significance using an analysis of deviance with a chi-squared test statistic. (3) Additionally, we built GLMMs with binomial error to test if the CFU/g in a dataset of only positive plants was predictive of each symptom's presence or absence again with a random effect of plant identity. Most symptoms were tested across all three seasons except for stunting and leaf chlorosis, early-season symptoms for which the 2017 year was not included due to spring inoculation precluding year 1 symptoms. For both sets of symptom models (models #2 and #3), the response variable (presence / absence of symptoms) was clustered by symptom monitoring dates around the three annual qPCR detection dates. The May qPCRs were paired with symptoms from spring (from bud break in May – June 15th), July qPCRs were paired with summer symptoms (from June 16th – August 15th), and September qPCRs were paired with fall symptoms (from August 16th – the end of sampling in October). If symptoms were not detected during these windows, then the response variable was 0. If symptoms were detected at any point, the response variable was 1. (4) Winter recovery was analyzed as a binomial GLM, testing the effect of cultivar on plant recovery between 2017 – 2018. The response, plant recovery, was a binary variable in which plants that had been positive during the 2017 year were marked as either recovered or still infected. (5) We then built a generalized linear mixed effect model (GLMM) with binomial error to test the effects of cultivar and year on binary infection status, with plant identity included as a random effect. (6) Finally, we built a set of annual vine position models (separate models for 2017, 2018, and 2019) to test the effect of sampling location and cultivar on detection of the pathogen, with plant identity as a random effect. For all GLM models we used the glm function in the stats package in R, and for all GLMM models we used the glmmTMB package (v. 1.1.5)(Brooks *et al.*, 2017). A survival analysis for symptom onset in 2017 (the year of inoculation) was

conducted using the Surv function in the survival package in R (v. 3.4.0) wherein symptom onset times were compared (Therneau, 2022).

Plotting

All data were plotted in R using the package ggplot2 (Wickham, 2016) and color sets were picked using the package RColorBrewer (Neuwirth, 2022).

Results

Inoculations were detectable by qPCR across the three years

The inoculation success ranged between 73-100% per cultivar (Table 2), measured via positive qPCR in either July or September, 2017. At the first sampling post inoculation, in July 2017, between 74-90% of total vines were positive for *X. fastidiosa* presence, showing a successful inoculation (see Figure 1). That percentage dropped by autumn of year 1, then increased again in July and September of year 2. By the second year, most Chardonnay plants had been removed by the vineyard manager (rogued), so our proportion increased drastically for that cultivar, while Merlot and Cabernet Sauvignon were not as frequently infected as in either previous year due to most of the positive plants recovering from the *X. fastidiosa* infection. The effect of grape cultivar on the inoculation success in year 1 was not significant when analyzed by a GLM with binomial error ($Chi^2 = 3.5808$, $p = 0.1669$, $df = 2$). Using the 3-year dataset, we also tested the effect of cultivar and year, with plant identity as a random effect. Cultivar was overall not significant in the effect on infection status throughout the experiment ($Chi^2 = 4.624$, $p = 0.0991$, $df = 2$); however, year was significant ($Chi^2 = 65.317$, $p = 6.555e\text{-}15$, $df = 2$). One mock inoculated plant tested positive by qPCR in Fall, 2019 (see supplemental Figure 1).

Symptoms were detectable early, and in an unexpected order

In the Merlot and Chardonnay blocks, symptoms were first noted on July 21, 2017 (3 months post inoculation), and included uneven lignification, leaf scorch and shriveled berries (see Figures 2 and 3). Of these, the shriveled berries (Figure 4E) were the most consistent symptom across vines and within single vines. In the Cabernet Sauvignon block, the first symptoms were also noted on July 21, 2017. The only symptom noted on that date, and the following evaluation date, was shriveled berries. Leaf scorch and uneven lignification were not noted until August 28, 2017, for the Cabernet block. Stunted growth was noted on the first evaluation dates of 2018 and 2019 (Figure 4A), coinciding with bud break and early shoot growth on Apr 10, 2018, and Apr 23, 2019 (Merlot and Chardonnay) and Apr 25, 2018, and May 6, 2019 (Cabernet Sauvignon).

In 2017, shriveled berries were the most common symptom; however, all other symptoms were also observable in the first year. The only symptom that was detectable in the control inoculated vines was stunting, which, in 2018 and 2019 was still much less common in the SCP-controls compared to the pathogen-inoculated plants (see Figure 2, and Supplemental 2). These symptoms were analyzed with survival curves based on their times of emergence in 2017 and the cox proportional hazards models did not converge to compare symptoms to each other. A comparison of symptom emergence in controls and inoculated plants can be seen in the supplemental materials (See supplemental Figure 3).

Differences in symptom progression among cultivars were substantial

The three cultivars showed different trends of symptom development over the three years of the study. Overall, Chardonnay exhibited the highest counts of symptoms including leaf chlorosis, leaf scorch, matchstick petioles, stunting, and shriveled berries. In Cabernet Sauvignon year 1, the prominent symptoms were shriveled berries and uneven lignification in the fall. In years 2-3, the few continuously infected plants exhibited all measured symptoms starting in July or August through the end of the growing season (see Figure 3). The controls and experimental vines exhibited stunting in Cabernet Sauvignon in May and June (see Figure 2). In Chardonnay, year 1 began with shriveled berries in July that peaked in September (the last observation time

before berry harvest) and exhibited subsequent leaf scorch and matchstick petioles in the fall. The following years showed high stunting, scorch, and chlorosis, all beginning in May when symptoms were first measured. There was uneven lignification detected in the summer and fall. Matchstick petioles increased throughout the fall in both years. For Merlot, July of year 1 showed shriveled berries and leaf scorch/chlorosis, followed by fall matchstick petioles. The following years showed stunting beginning in May similar to Chardonnay along with some leaf scorch, matchstick petioles, uneven lignification, and shriveled berries in the fall (See Figure 3).

With the data from three cultivars combined, annual trends emerged. Shriveled berries were the most informative symptom during the late season of the inoculation year. Although less abundant than shriveled berries, other symptoms in the first year included leaf scorch, matchstick petioles, uneven lignification, and stunting. In the second year, there was widespread stunting immediately observable after bud break, as well as leaf scorch which increased in August and September. Although many plants in this dataset had recovered by the third year, during 2019 there was predominantly early-season stunting along with some leaf chlorosis, followed by late season leaf scorch and matchstick petioles.

Number of positive plants drops in all three cultivars due to recovery and roguing

Although between 73-100% of plants per cultivar tested positive for *X. fastidiosa* infections during the first year (See Table 2), the number of positive plants dropped in all three cultivars throughout the experiment. For Chardonnay, upon symptom development in year 1, 10 of the 15 experimental plants were rogued by the caretakers of the vineyard (see Figure 1). However, in Merlot and Cabernet Sauvignon, there was substantial recovery during the first winter, and between both seasons, respectively (see Figure 1). For Cabernet Sauvignon, the recovery is immediate, with all but one plant recovering during the first winter, and that one plant remaining both positive and highly symptomatic throughout the course of the experiment. In Merlot, there is a more gradual decline, with some plants recovering each winter. The differences between cultivars in recovery in the winter of 2017-2018 were not significant according to a binomial GLM ($Chi^2 = 3.5297$, $p = 0.1712$, $df = 2$).

Bacterial populations of positive samples remained relatively consistent

Bacterial populations for each positive sample were calculated; however, most positive samples were within several orders of magnitude in terms of their CFU/g tissue (see Figure 5). Bacterial populations ranged from 4.1×10^4 ($C_t = 35.8$) to 4×10^6 ($C_t = 19.6$). The mean CFU/g for all positive samples across the three years were respectively 2.3×10^5, 1.4×10^5, and 4.9×10^5. While the number of total plants testing positive decreased over the course of the three years (see Figure 1), the population sizes remained fairly stable (see Figure 5). The only year in which the expected seasonal change in population sizes was observed was in 2018, where no samples test positive in May, then population sizes for all three cultivars in July average 10^4 CFU/g, followed by the September sampling in which the average population sizes for the three cultivars increase to averages in the 10^5 CFU/g range.

Movement was faster than previously documented, and inoculated site only relevant in year 1

The inoculations were intentionally conducted on the terminal western shoot in order to standardize the inoculation point for tracking of visual symptoms and population sizes by vine position. During the first year, both symptoms and qPCR-positive locations were concentrated near the inoculation point in the west, mid-west, and west-center sampling positions. In year 1,

leaf scorch, matchstick petioles, uneven lignification, and shriveled berries were predominantly detected on the western part of the vine. By year 2, vine position was no longer relevant, and both symptom development and bacterial populations appeared evenly throughout the plant (see Figures 6, 7 and supplemental Figure 4). Vine location was significant in 2017, with vine locations other than the western shoot having negative coefficients. This indicates that there was the greatest detection on the western (inoculated) shoot ($Chi^2 = 13.971$, $p = 0.00184$, $df = 3$). In 2018, vine location was also significant; however, the highest coefficient was for the mid-west sampling location ($Chi^2 = 11.793$, $p = 0.00813$, $df = 3$). In 2019, the effect of vine position was no longer significant ($Chi^2 = 0.722$, $p = 0.868$, $df = 3$).

Pathogen positivity was predictive of symptom presence, but pathogen population size was not
Several GLMMs with binomial error were constructed to test if the presence or absence of each symptom was dependent on the CFU/g value in the plant at that point. This model only included positive plants (CFU/g > 0) to examine if increasing bacterial populations resulted in increased symptom prevalence. These relationships were not significant for any symptom. Next, the presence of symptoms over the course of the year was evaluated as a response to the plant being infected. Those models were set up with the symptoms as the response variable and with one predictor variable: infection status. Infection status is a significant factor influencing all symptoms aside from uneven lignification (See Table 3).

Discussion

This study emphasizes the necessity of in-situ experimentation with grapevine pathogens. The established paradigm for PD of low recovery and leaf scorch in Napa Valley was challenged with this study. Our results indicate shriveled grape berries as the predominant early symptom, rapid spread of the bacterial pathogen through the grapevine, and high overwinter curing.

Our findings suggest that previously overlooked initial symptoms are significantly associated with bacterial presence at the inoculation site. Leaf scorch is widely considered the most prominent and first visible symptom of PD. However, we showed that the most prominent symptoms in year 1 were shriveled berries and uneven lignification. While the symptom of shriveled berries has been previously observed (Feil *et al.*, 2003), it has been largely ignored since then. These symptoms became visible in the year the vines were inoculated, near the point of veraison (July 21st), as vines were starting to ripen fruit. In addition, shriveled berries were the only symptoms that occurred in all three cultivars at that time. Given the large amount of variability between cultivars, having a symptom that is consistent across cultivars may be essential for management. On the other hand, leaf scorch was observable in Chardonnay and Merlot, but was much less frequent in Cabernet Sauvignon. Uneven lignification was a prominent early symptom in Chardonnay, but not in the two other cultivars. Shriveled berries appeared more reliably and sooner after inoculation than foliar symptoms in all cultivars. Foliar symptoms tended to appear more consistently in grapevines that had been infected for over a year and may already be serving as sources of inoculum. Given the consistency of shriveled berries across the three wine grape cultivars, this appears to be the key symptom to visually detect early infections of *X. fastidiosa*, whereas other symptoms such as stunting, leaf scorch and uneven lignification arise later in the infection cycle.

Symptom and pathogen movement through the plants were tracked throughout the project. In the first year, neither symptoms nor detectable bacterial populations extended past the trunk to the non-inoculated side of the plant (these vines are trained horizontally similar to a candelabra). However, in years 2-3, symptoms and populations were evenly distributed on both sides of the trunk, even concentrated near the trunk, supporting that the infection favors the side of initial inoculation only in the first year. Although not explored in the current study, symptom distribution and pathogen movement may be affected by the manner of pruning. These plants were annually subjected to cane pruning, whereby the fruiting wood is retained as 8 to 15 node fruiting units (canes). Annual growth is from buds on these canes, and the permanent vine structure consists of trunk and arms. In contrast, spur pruning retains shorter fruiting units (1 to 3 nodes) distributed along the length of a cordon, another aboveground permanent structure. Differences in the way in which fruiting wood is selected and retained between cane and spur pruned vines, as well as the overall proportion of permanent to annual vine structure may impact recovery rates and should be investigated further.

Overwinter curing varied substantially between the three wine grape cultivars, but overall, it was greater than previously hypothesized, with two of the cultivars experiencing nearly 75% recovery of plants. Based on early-season inoculation and high bacterial population levels, these plants were expected to become chronically infected. The high recovery rates suggest that there might be substantially more plants exposed to *X. fastidiosa* infections than has been previously presumed. If many more plants are infected, but avoid chronic infection, we may be underestimating the amount of vector spread of this pathogen. These commercial vines were inoculated early in the growing season, which has been shown to be most effective inoculation timing to result in chronic infection in Napa Valley (Purcell, 1981). Therefore, the diseased

plants that we observe under commercial growing conditions may represent a fraction of plants that are exposed to *X. fastidiosa*. Overwinter curing is still poorly understood, but temperature adaptation in PD strains across California might contribute to the pathogen's tolerance of colder winters (Vanhove *et al.*, 2020).

A strong association was present between symptom detection and bacterial presence in the plant. Typical infections averaged 10^5 CFU/g of plant petiole tissue; however, population size was not found to impact symptoms. Only binary infection status significantly influenced symptoms. Plants that tested positive were above the threshold of population size needed for transmission by the local vector, *Graphocephala atropunctata*, measured as populations above 10^4 CFU/g of stem tissue (Hill & Purcell, 1997). In May, qPCR detection was rare, even for plants that tested positive later in the growing season. Given the complications with detection via qPCR that vary seasonally and throughout a plant, this study supports symptom identification as the best way to detect infected grapevines in Napa Valley early in the growing season. These data also indicate that even in plants that are chronically infected and those with severe PD symptoms, populations will still be undetectable if measured even a few months too early.

This project provides new insight into the characteristics of early PD infection, which should improve disease detection in vineyards. The long-term goal of elucidating the development of this infectious disease in a high-value agricultural system is vital to the California agricultural economy. As PD has recently expanded into global grape growing regions in Europe and Asia (Castillo et al., 2021), there has never been a more urgent time to understand the infection process of *X. fastidiosa* in *Vitis vinifera*.

Figures and Tables

Table 1. Description of study sites, vines, and inoculation dates. Vines are split at each site into experimental, which are inoculated with *Xylella fastidiosa* subsp. *fastidiosa* strain Napa1 (exp) and the negative control vines which are inoculated with SCP buffer (ctrl).

Vineyard	Cultivar	Number of Vines	Inoculation Date	Vine Training	Pruning
Oak Knoll	Chardonnay	15 Exp 15 Ctrl	April 14, 2017	bilateral canes	Cane pruned
Oak Knoll	Merlot	15 Exp 15 Ctrl	April 14, 2017	bilateral canes	Cane pruned
Calistoga	Cabernet Sauvignon	10 Exp 10 Ctrl	April 28, 2017	quadrilateral canes	Cane pruned

Table 2. Plants that were positive based on 2017 July or September samplings, the year inoculations occurred.

Cultivar	Positive Plants	Percentage
Overall	*34/40*	*85%*
Cabernet Sauvignon	8/10	80%
Chardonnay	15/15	100%
Merlot	11/15	73%

Table 3. Results from binomial GLMM showing relationships between all symptoms and infection status of the plant. P-values, Chi-square values, and degrees of freedom are shown for each test.

Symptom	Infection Status (df = 1)	
	p	Chi^2
Chlorosis	0.02367	5.1186
Matchstick petioles	1.4e-07	27.723
Leaf scorch	1.567e-09	36.449
Stunting	0.003282	8.6439
Uneven lignification	0.9987	0
Shriveled berries	4.164e-05	16.795

Figure 1. Count of plants positive by qPCR, divided by cultivar, year, and sampling month. In Chardonnay, 10 plants were removed over the winter of 2017-2018. In Merlot, only one plant was removed, between the sampling dates of July and September 2019. Plants that were removed are not plotted but subsequently might have been either positive or negative if they had not been rogued. In May of all years, all samples were negative so May is not included in the plot.

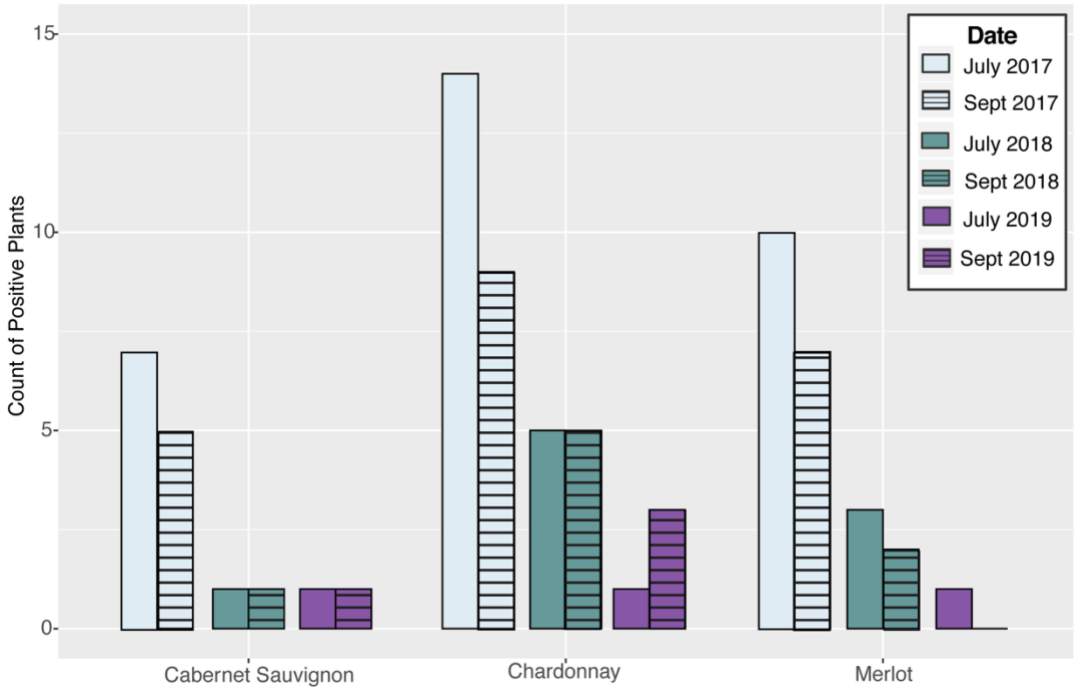


Figure 2. Total number of symptomatic analyzed shoots per sampling time measured from April 26, 2017 – October 23rd, 2019, at two-week intervals during the growing season. The plants were each measured at 5 locations for symptoms, with a binary presence or absence of each symptom, so the scale is the count of symptoms at an observation point. Axes vary by year as the magnitude of symptoms is highest in year 2. Symptom names are shortened to LC (Leaf Chlorosis), LS (Leaf Scorch), MP (Matchstick Petioles), SB (Shriveled Berries), St (Stunting), and UL (Uneven Lignification).

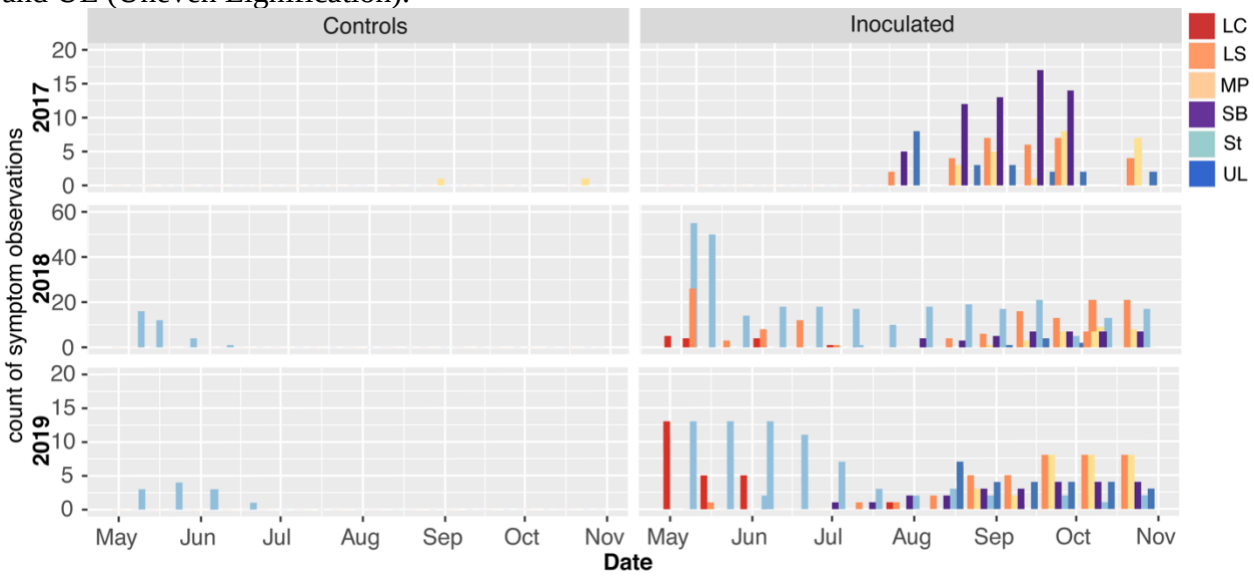


Figure 3. Total number of symptomatic shoots per sampling time divided by cultivar, only including inoculated plants. Axes vary by year and show total sites per sampling point with each symptom. Symptom names are shortened to LC (Leaf Chlorosis), LS (Leaf Scorch), MP (Matchstick Petioles), SB (Shriveled Berries), St (Stunting), and UL (Uneven Lignification). Asterisks indicate the harvest date for each block, after which, berry shrivel can no longer be evaluated. For Cabernet Sauvignon, the harvest date for 2017 was October 27th, 2017 and for 2018 – 2019 the harvest was after the completion of our data collection. For Merlot, the harvest dates were September 13th, 2017, September 25th, 2018, and October 8th, 2019. For Chardonnay, the harvest dates were September 13th, 2017, September 25th, 2018, and September 9, 2019.

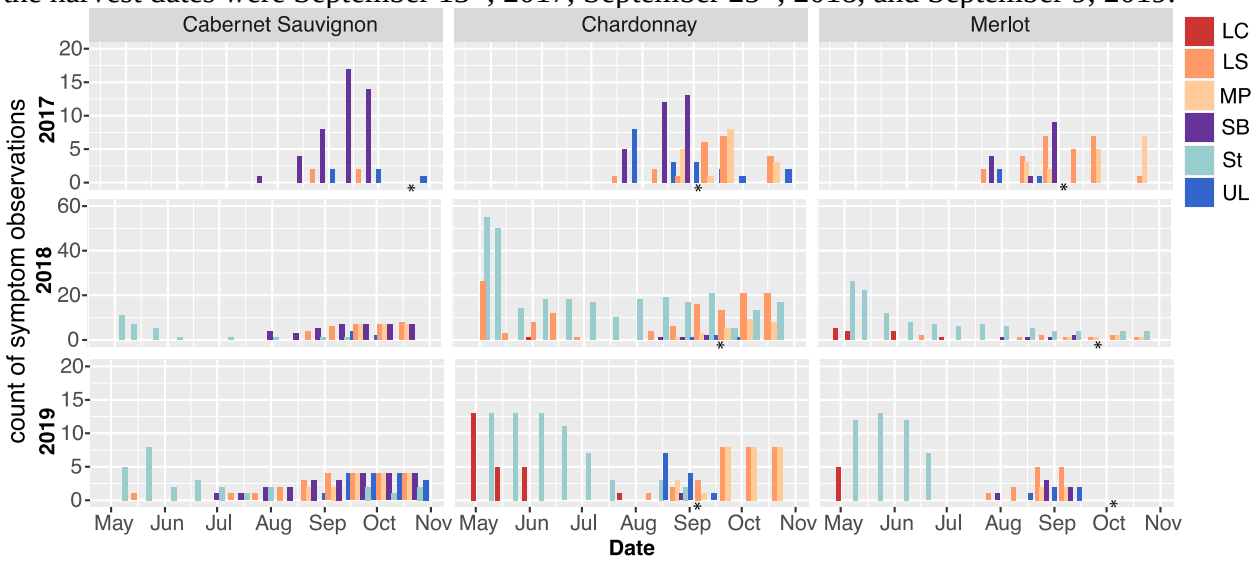


Figure 4. Visual symptoms of inoculated vines included A. stunting of one vine surrounded by unstunted growth (Chardonnay, June 19th, 2018), B. asymptomatic leaf (Merlot, June 19th, 2018), C. leaf chlorosis (Merlot, June 19th, 2018), D. matchstick petioles circled in red (Merlot, August 14th, 2017), E. shriveled berries adjacent to healthy berries (Cabernet Sauvignon, September 11th, 2018), F. uneven lignification (Merlot, July 21st 2017), G. leaf scorch (Chardonnay, August 12th, 2019), and H. leaf scorch (Merlot, August 12th, 2019).

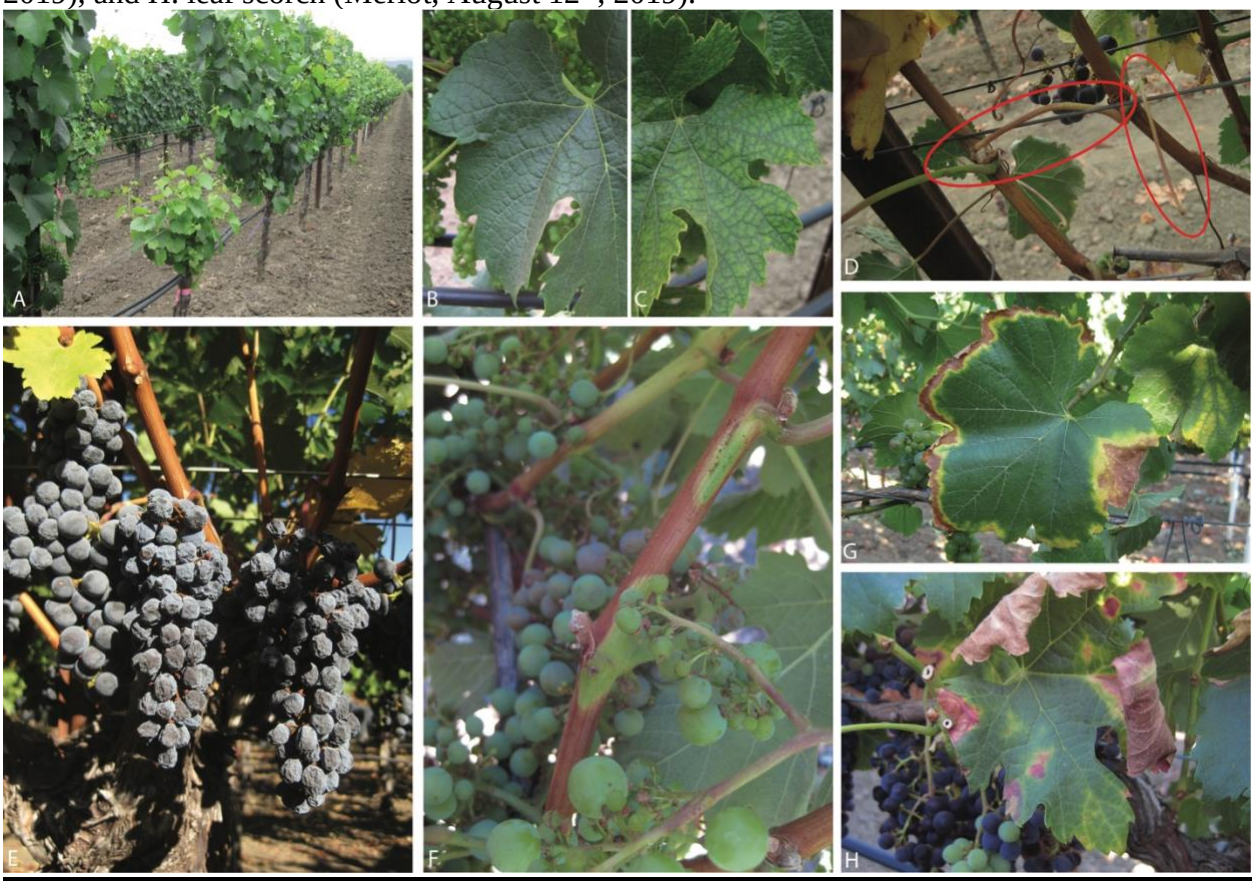


Figure 5. Boxplot of log transformed CFU per gram of positive samples only, samples with C_t > 36 are not included in these calculations. Sampling points were from July 2017 – September 2019. In September, 2019, no Merlot samples tested positive. In May of all years, all samples were negative so May is not included in the plot.

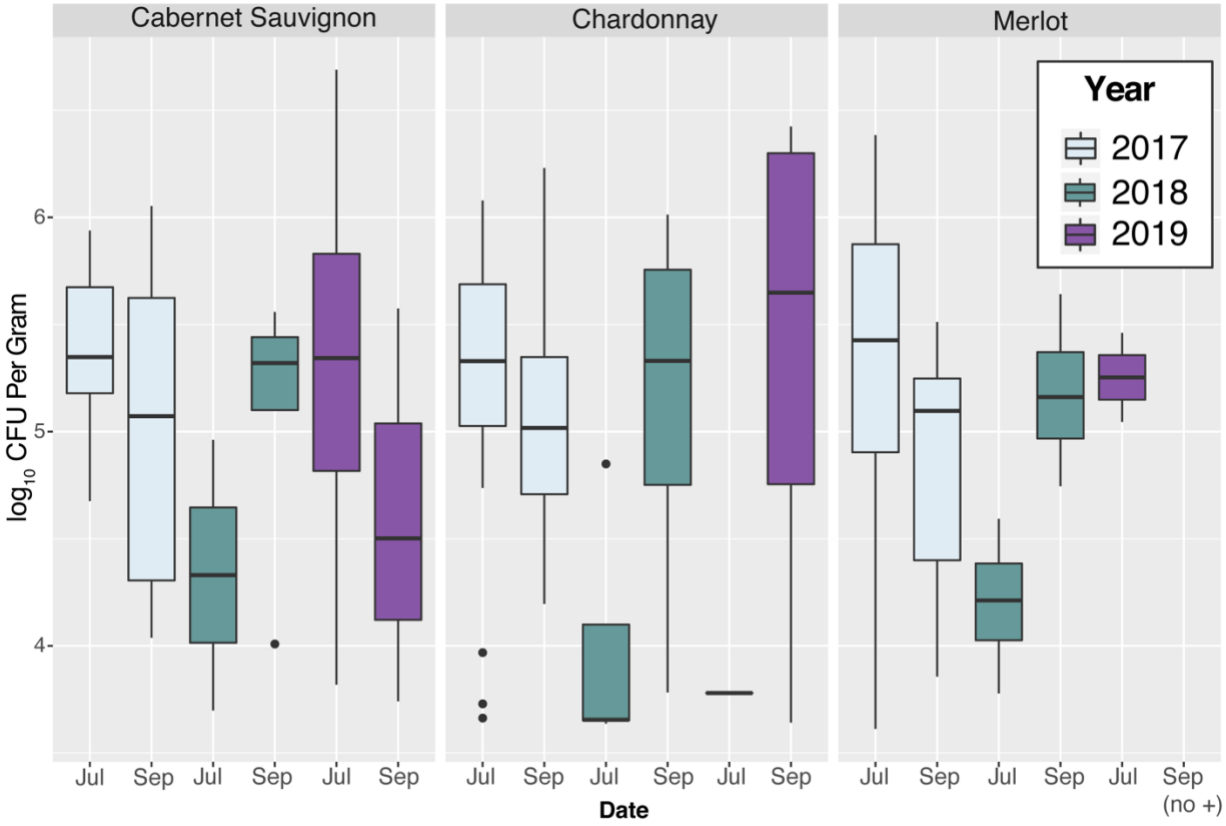


Figure 6. Number of positive plants divided by position on the vine. Y axes vary due to recovery and fewer plants in years 2-3 and show qPCR positive plants per location on the vine. Most sampling was done at the basal petiole from the terminal shoots (west and east) and from the shoots adjacent to the trunk (west-center and east-center), which were collected at each time point (four samples per plant). The only exception is shoot 2, a location only sampled in 2017 – representing the proximal shoot to the terminal shoot on the western cane. In 2017, the terminal shoot on the eastern cane (east) was not sampled. In 2018-2019, plants were tested in May, July, and September. However, in May, all samples were negative so they are not included in the plots. Grape cultivars are shortened to Ch (Chardonnay), M (Merlot), and CS (Cabernet Sauvignon).

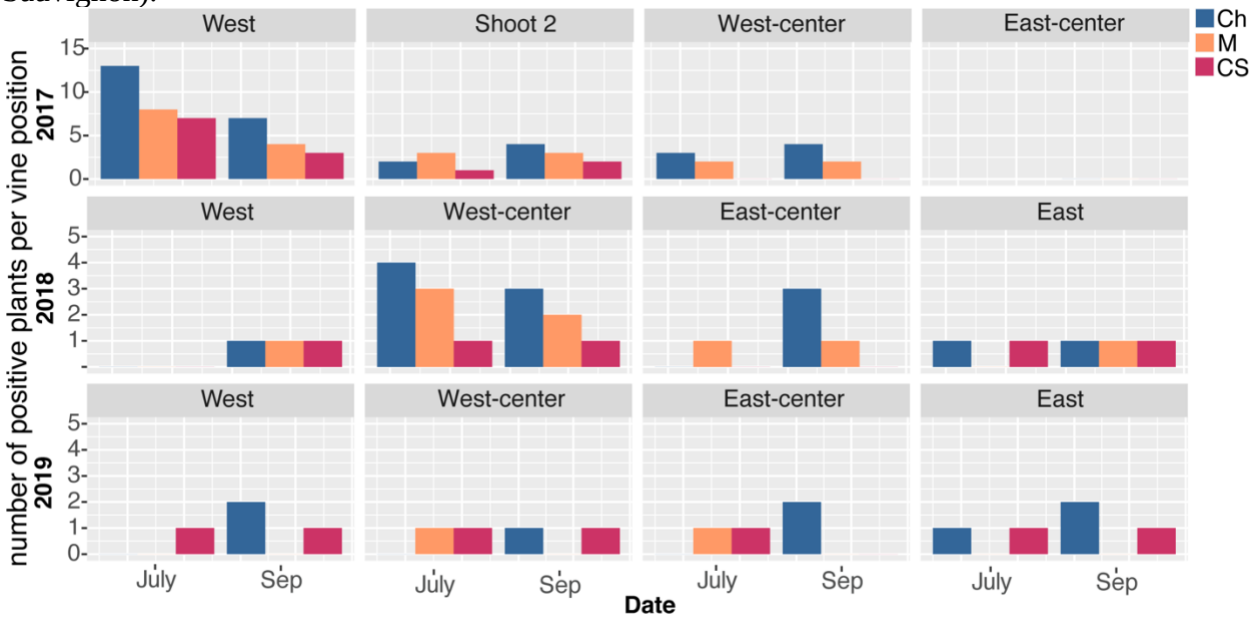
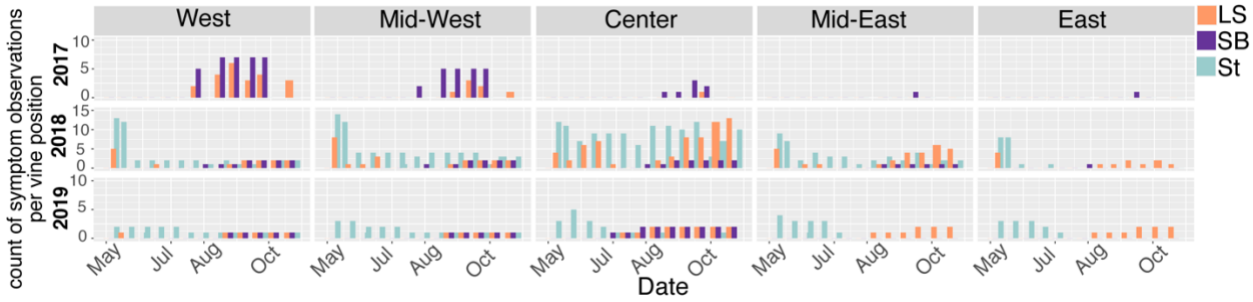
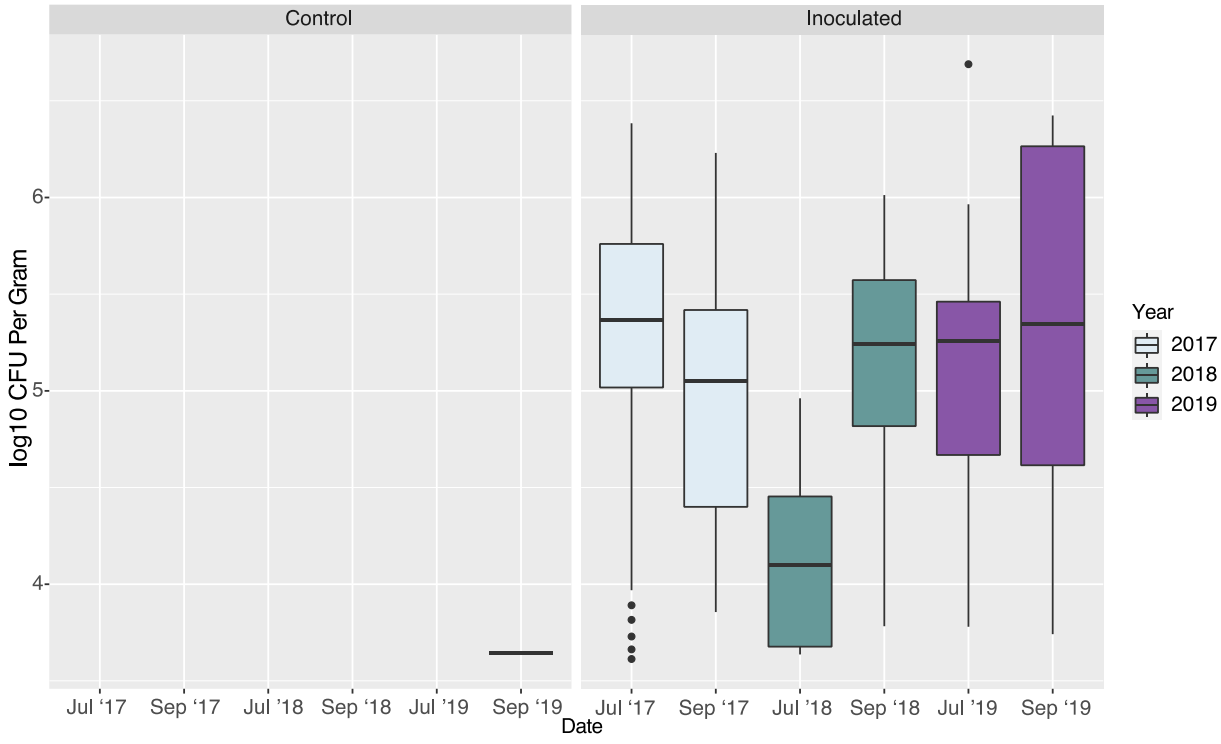


Figure 7. Symptom progression over the three years of the study, broken down by vine position, including 3 symptoms. Sampling was conducted every 2 weeks throughout the growing season. Y axes vary by year. Symptom names are shortened to LS (Leaf Scorch), SB (Shriveled Berries), and St (Stunting). Each symptom was recorded as present or absent on each of 5 areas of the vine: (1) west and (2) east, corresponding to the terminal ends of each shoot; (3) mid-west and (4) mid-east, corresponding to the middle of each cane; and (5) center corresponding to the head of the vine, where the canes originate. Shriveled berries could not be evaluated after the harvest date for each block. For Cabernet Sauvignon, the harvest date for 2017 was October 27th, 2017 and for 2018 – 2019 the harvest was after the completion of our data collection. For Merlot, the harvest dates were September 13th, 2017, September 25th, 2018, and October 8th, 2019. For Chardonnay, the harvest dates were September 13th, 2017, September 25th, 2018, and September 9, 2019.

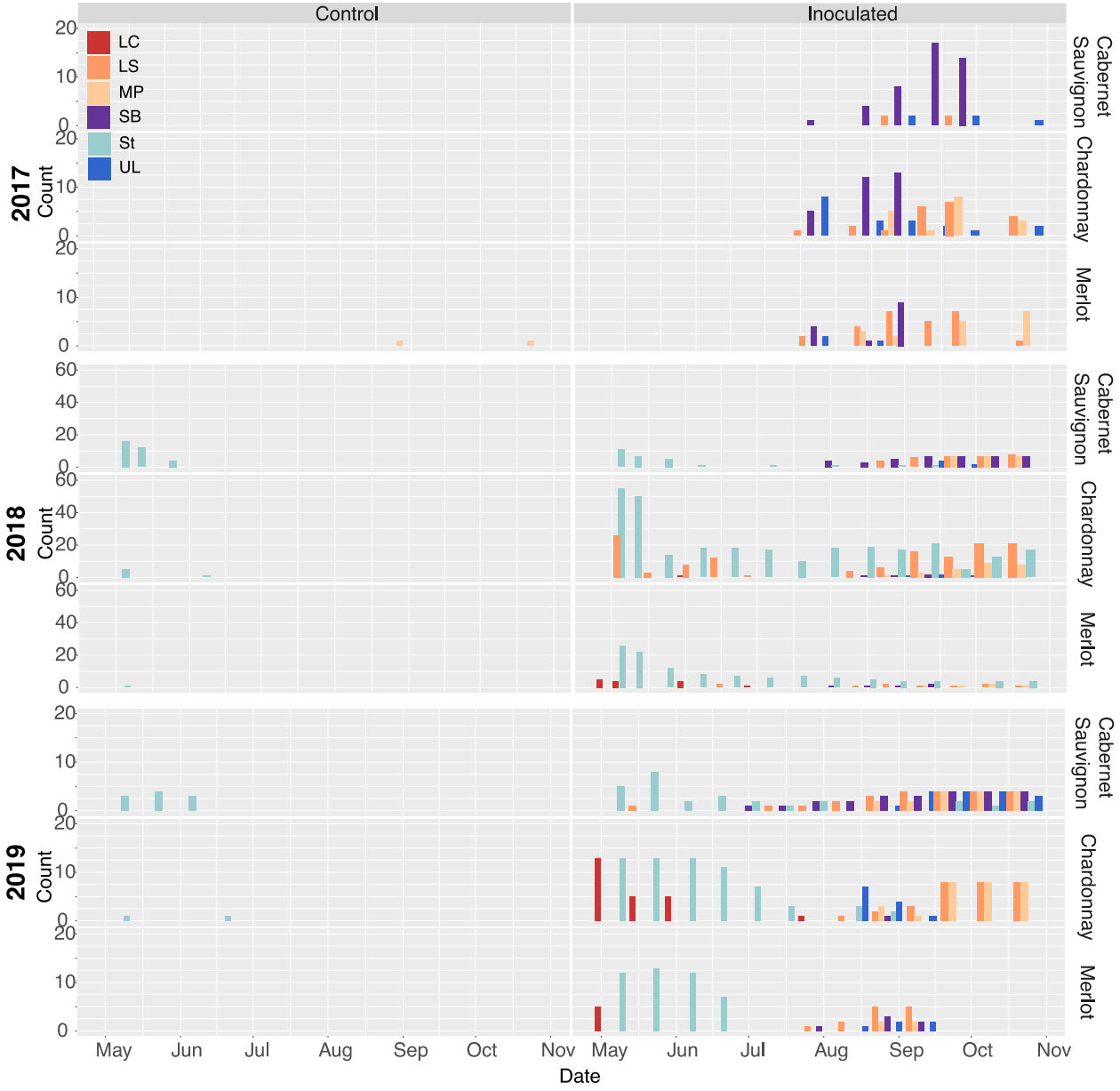


Supplementals – Chapter 3

S1 Bacterial population sizes by sampling date and treatment: Data are split between inoculated and control samples so as to see if any control plants tested positive.

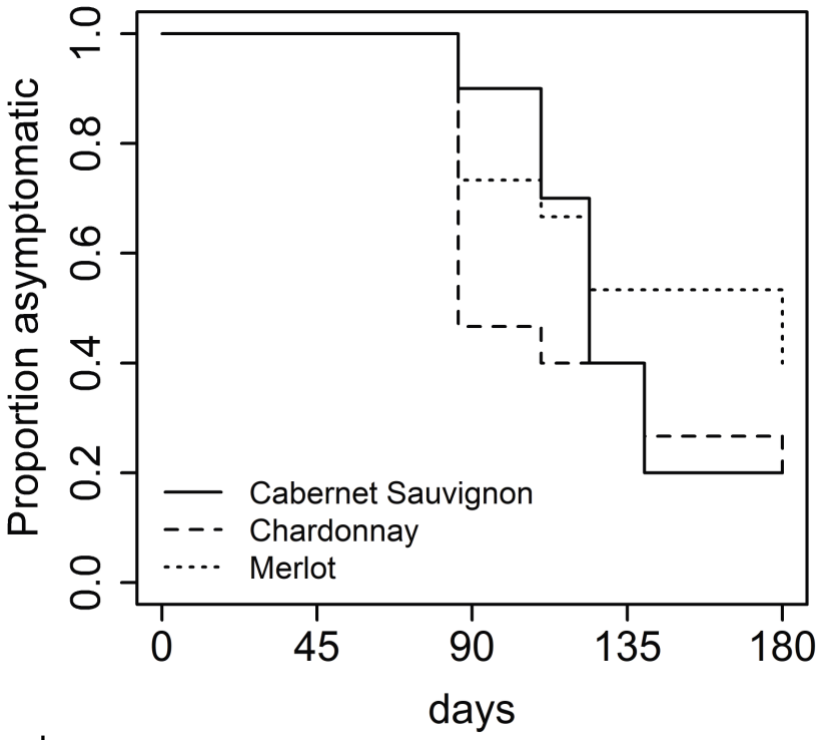


S2 Symptoms split by treatment and variety: All symptoms are displayed, broken up by variety so that controls where symptoms were detected can be identified by variety. Symptom names are shortened to LC (Leaf Chlorosis), LS (Leaf Scorch), MP (Matchstick Petioles), SB (Shriveled Berries), St (Stunting), and UL (Uneven Lignification).

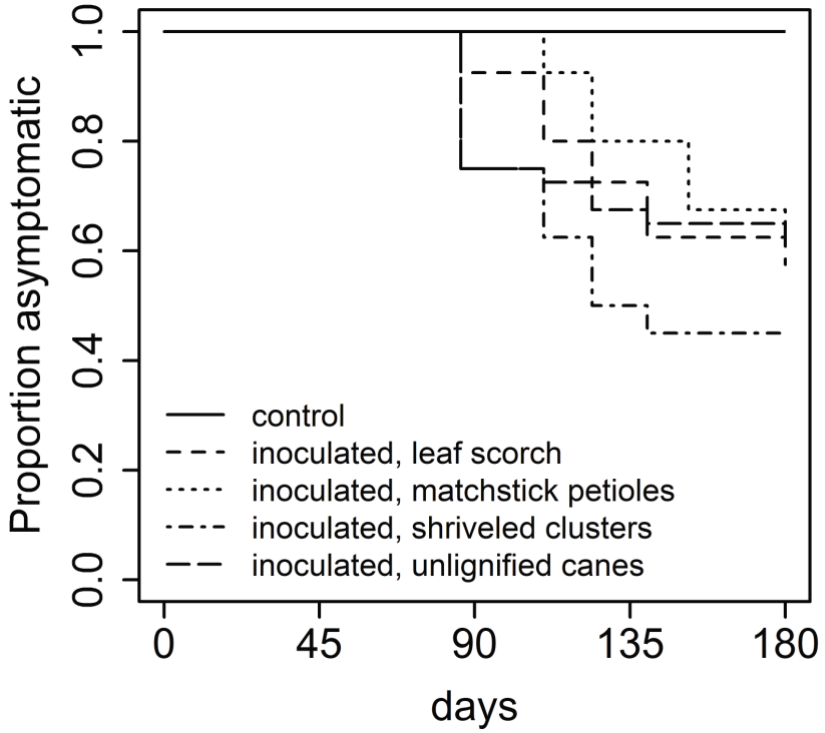


S3 Survival curves: Visualization of survival analysis of symptom onset in 2017 between (a) grape cultivars and (b) the four symptoms present during the inoculation year.

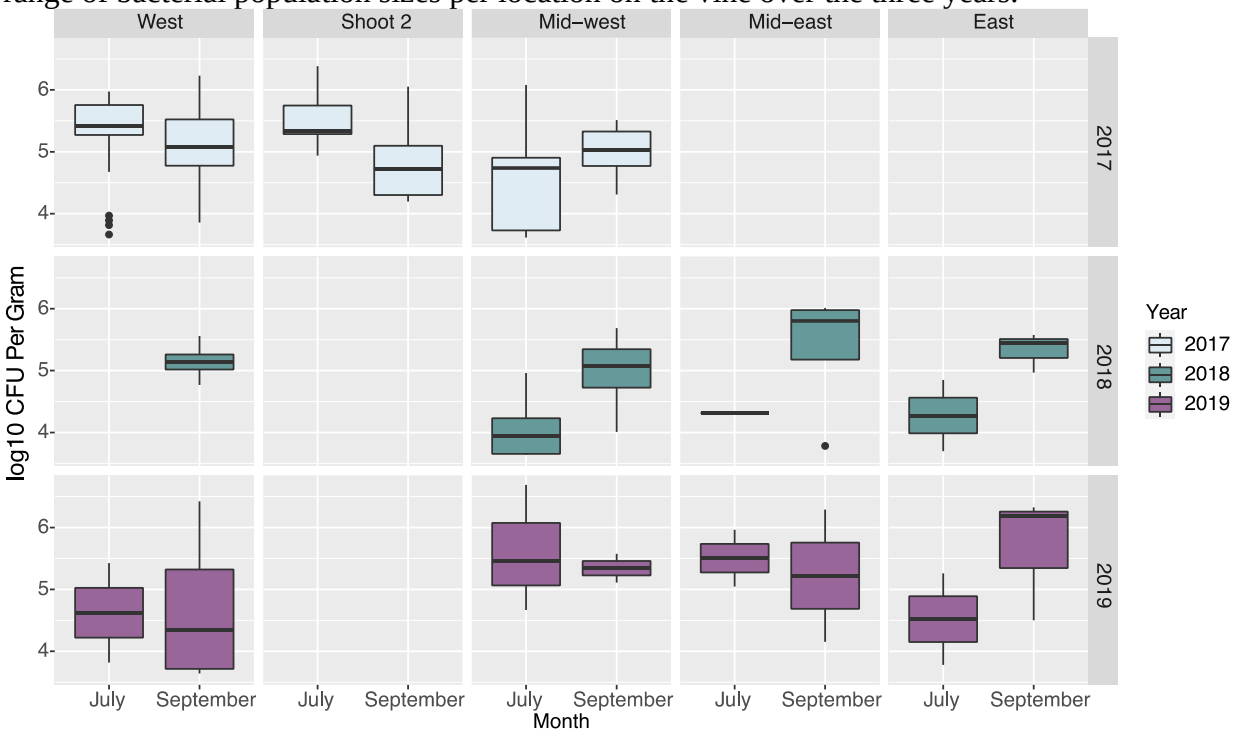
a



b



S4 Bacterial populations split by vine location and year: A boxplot is shown displaying the range of bacterial population sizes per location on the vine over the three years.



REFERENCES CITED

Aakre CD, Phung TN, Huang D, Laub MT, 2013. A bacterial toxin inhibits DNA replication elongation through a direct interaction with the β sliding clamp. *Molecular Cell* **52**, 617–628.

Abdelrazek S, Bush E, Oliver C *et al.*, 2023. A survey of *Xylella fastidiosa* in the US state of Virginia reveals wide distribution of subspecies fastidiosa and infection of grapevine by subspecies multiplex. *Phytopathology*.

Aguilar E, Moreira L, Rivera C, 2008. Confirmation of *Xylella fastidiosa* infecting grapes Vitis vinifera in Costa Rica. *Tropical Plant Pathology* **33**, 444–448.

Akaike H, 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*.

Almeida RPP, Nascimento FE, Chau J *et al.*, 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Applied and Environmental Microbiology* **74**, 3690–3701.

Almeida RPP, Purcell AH, 2003. Biological Traits of *Xylella fastidiosa* Strains from Grapes and Almonds. *Applied and Environmental Microbiology* **69**, 7447–7452.

Amanifar N, Babaei G, Mohammadi AH, 2019. *Xylella fastidiosa* causes leaf scorch of pistachio (Pistacia. *Phytopathologia Mediterranea* **58**, 369–378.

Barbosa RL, Benedetti CE, 2007. BigR, a transcriptional repressor from plant-associated bacteria, regulates an operon implicated in biofilm growth. *Journal of Bacteriology* **189**, 6185–6194.

Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J, 2009. Continua of specificity and virulence in plant host-pathogen interactions: Causes and consequences. *New Phytologist* **183**, 513–529.

Bates D, Mächler M, Bolker B, Walker S, 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48.

Benson D a, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, 2010. GenBank. *Nucleic acids research* **38**, D46-51.

Brooks M, Kristensen K, van Benthem K *et al.*, 2017. *glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling*.

Brun LA, Maillet J, Richarte J, Herrmann P, Remy JC, 1998. Relationships between extractable copper, soil properties and copper uptake by wild plants in vineyard soils. *Environmental Pollution* **102**, 151–161.

Brynildsrud O, Bohlin J, Scheffer L, Eldholm V, 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology* **17**, 1–9.

Burbank L, 2018. Winter climate and cultivar effects on severity of Pierce's Disease in table grapes. *Proceedings of the 2018 International Symposium on Proactive Technologies for Enhancement of Integrated Pest Management of Key Crops Winter*, 159–166.

Camacho C, Coulouris G, Avagyan V *et al.*, 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9.

Carbon S, Mungall C, 2023. Gene Ontology Data Archive. *Zenodo*.

Castillo AI, Almeida RPP, Kuo CH *et al.*, 2021a. Genetic differentiation of *Xylella fastidiosa* following the introduction into Taiwan. *Microbial genomics* **7**.

Castillo AI, Bojanini I, Chen H, Kandel PP, De La Fuente L, Almeida RPP, 2021b. Allopatric Plant Pathogen Population Divergence following Disease Emergence. *Applied and*

Environmental Microbiology **87**, 1–19.
Castillo, A.I., Chacón-díaz, C., Rodríguez-murillo, N., Coletta-, H.D., Almeida, R.P.P., Rica, C., Diaz, C.C., Colleta-filho, H., Almeida, R.P.P., 2020. Impacts of local population history and ecology on the evolution of a globally dispersed pathogen. BMC Genomics 21, 1–20. https://doi.org/https://doi.org/10.1186/s12864-020-06778-6
Chatterjee S, Almeida RPP, Lindow S, 2008. Living in two Worlds: The Plant and Insect Lifestyles of *Xylella fastidiosa*. *Annual Review of Phytopathology* **46**, 243–271.
Cicero DO, Contessa GM, Pertinhez TA *et al.*, 2007. Solution structure of ApaG from *Xanthomonas axonopodis* pv. *citri* reveals a fibronectin-3 fold. *Proteins: Structure, Function, and Bioinformatics* **67**, 490–500.
Coe MT, Evans KM, Gasic K, Main D, 2020. Plant breeding capacity in U.S. public institutions. *Crop Science* **60**, 2373–2385.
Coletta-Filho H Della, Castillo AI, Laranjeira FF *et al.*, 2020. Citrus Variegated Chlorosis: an Overview of 30 Years of Research and Disease Management. *Tropical Plant Pathology* **45**, 175–191.
Daugherty MP, Almeida RPP, Smith RJ, Weber EA, Purcell AH, 2018. Severe pruning of infected grapevines has limited efficacy for managing pierce's disease. *American Journal of Enology and Viticulture* **69**, 289–294.
Davis MJ, Purcell AH, Thomson S V., 1978. Pierce's disease of grapevines: Isolation of the causal bacterium. *Science*.
Delbianco A, Gibin D, Pasinato L, Boscia D, Morelli M, 2023. Update of the *Xylella* spp. host plant database – systematic literature search up to 30 June 2022. *EFSA Journal* **21**.
Denancé N, Briand M, Gaborieau R, Gaillard S, Jacques MA, 2019. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. *BMC Genomics* **20**, 1–21.
Deyett E, Rolshausen PE, 2019. Temporal Dynamics of the Sap Microbiome of Grapevine Under High Pierce's Disease Pressure. *Frontiers in Plant Science* **10**, 1–15.
Didelot X, Wilson DJ, 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Computational Biology* **11**, 1–18.
EFSA, 2018. Update of the *Xylella* spp. host plant database. *EFSA Journal* **16**.
EFSA Panel on Plant Health (PLH), 2015. Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. *EFSA Journal* **13**, n/a-n/a.
Faith DP, 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10.
Feil H, Feil WS, Purcell AH, 2003. Effects of date of inoculation on the within-plant movement of *Xylella fastidiosa* and persistence of Pierce's disease within field grapevines. *Phytopathology* **93**, 244–251.
Feil H, Purcell AH, 2001. Temperature-Dependent Growth and Survival of *Xylella fastidiosa* in Vitro and in Potted Grapevines. *Plant Disease* **85**, 1230–1234.
Felsenstein J, 1985. Confidence Limits on Phylogenies: an Approach Using the Bootstrap. *Evolution* **39**, 783–791.
Francisco CS, Ceresini PC, Almeida RPP, Coletta-Filho HD, 2016. Spatial Genetic Structure of Coffee-Associated *Xylella fastidiosa* Populations Indicates that Cross Infection Does Not Occur with Sympatric Citrus Orchards . *Phytopathology* **107**, 395–402.
French WJ, Kitajima EW, 1978. Occurrence of plum leaf scald in Brazil and Paraguay. *Plant Disease Reporter* **62**, 1035–1038.

Galán J, Collmer A, 1999. Type III Secretion Machines : Bacterial Devices for Protein Delivery into Host Cells. *Science* **284**, 1322–1329.
Ge Q, Cobine PA, De La Fuente L, 2021. The Influence of Copper Homeostasis Genes copA and copB on *Xylella fastidiosa* Virulence Is Affected by Sap Copper Concentration. *Phytopathology* **111**, 1520–1529.
Giampetruzzi A, Saponari M, Loconsole G *et al.*, 2017. Genome-Wide Analysis Provides Evidence on the Genetic Relatedness of the Emergent *Xylella fastidiosa* Genotype in Italy to Isolates from Central America. *Phytopathology* **107**, 816–827.
Goheen AC, Raju BC, Lowe SK, Nyland G, 1979. Pierce's Disease of Grapevine in Central America. *Plant Disease Reporter*, 788–792.
Hajri A, Brin C, Hunault G *et al.*, 2009. A «repertoire for repertoire» hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in Xanthomonas. *PLoS ONE* **4**.
Harms A, Liesch M, Körner J, Québatte M, Engel P, Dehio C, 2017. A bacterial toxin-antitoxin module is the origin of inter-bacterial and inter-kingdom effectors of Bartonella. *PLoS Genetics* **13**, 1–22.
Hedge J, Wilson DJ, 2014. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* **5**, 6–9.
Hill BL, Purcell AH, 1995. Multiplication and Movement of *Xylella fastidiosa* within Grapevine and Four Other Plants. *Ecology and Epidemiology*.
Hill BL, Purcell AM, 1997. Populations of *Xylella fastidiosa* in plants required for transmission by an efficient vector. *Phytopathology* **87**, 1197–1201.
Hopkins DL, 1984. Variability of virulence in grapevine among isolates of the Pierce's disease bacterium. *Phytopathology* **74**, 1395–1398.
Hopkins DL, Adlerz WC, 1988. Natural Hosts of *Xylella fastidiosa* in Florida. *Plant Disease* **72**, 429–431.
Huang W, Reyes-Caldas P, Mann M *et al.*, 2020. Bacterial Vector-Borne Plant Diseases: Unanswered Questions and Future Directions. *Molecular Plant* **13**, 1379–1393.
Hulme PE, 2009. Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* **46**, 10–18.
Jacques MA, Denancé N, Legendre B *et al.*, 2016. New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. *Applied and Environmental Microbiology* **82**, 1556–1568.
Jolley KA, Chan MS, Maiden MCJ, 2004. mlstdbNet - Distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics*.
Jombart T, Dray S, 2010. Adephylo: Exploratory Analyses for the Phylogenetic Comparative Method. *Bioinformatics*.
Kahn AK, Almeida RPPP, 2022. Phylogenetics of Historical Host Switches in a Bacterial Plant Pathogen. *Applied and Environmental Microbiology* **88**.
Kahn AK, Sicard A, Cooper ML, Daugherty MP, Donegan MA, Almeida RPP, 2023. Progression of *Xylella fastidiosa* infection in grapevines under field conditions. *Phytopathology*.
Kanehise M, Goto S, 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research* **28**, 27–30.
Katoh K, Kuma K, Toh H, Miyata T, 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511–518.

Killiny N, Almeida RPP, 2011. Gene regulation mediates host specificity of a bacterial pathogen. *Environmental Microbiology Reports* **3**, 791–797.
Killiny N, Martinez RH, Dumenyo CK, Cooksey D a, Almeida RPP, 2013. The exopolysaccharide of *Xylella fastidiosa* is essential for biofilm formation, plant virulence, and vector transmission. *Molecular plant-microbe interactions : MPMI* **26**, 1044–53.
Koide T, Vêncio RZN, Gomes SL, 2006. Global gene expression analysis of the heat shock response in the phytopathogen *Xylella fastidiosa*. *Journal of Bacteriology* **188**, 5821–5830.
Kosakovsky Pond SL, Frost SDW, 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208–1222.
Krivanek AF, Stevenson JF, Walker MA, 2005. Development and comparison of symptom indices for quantifying grapevine resistance to Pierce's disease. *Phytopathology* **95**, 36–43.
Lafforgue M, 1928. La Bouillie Bordelaise. 1er Congres International de la Vigne et du Vin.
Landa BB, Castillo AI, Giampetruzzi A *et al.*, 2019. Multiple intercontinental introductions associated with the emergence of a plant pathogen in Europe. *Applied and Environmental Microbiology* **86**, 1–15.
Letunic I, Bork P, 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*.
Levy A, Salas Gonzalez I, Mittelviefhaus M *et al.*, 2018. Genomic features of bacterial adaptation to plants. *Nature Genetics* **50**, 138–150.
Lieth JH, Meyer MM, Yeo KH, Kirkpatrick BC, 2011. Modeling cold curing of Pierce's disease in Vitis vinifera "Pinot Noir" and "Cabernet Sauvignon" grapevines in California. *Phytopathology* **101**, 1492–1500.
Lockwood JL, Cassey P, Blackburn T, 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution* **20**, 223–228.
Lury DA, Fisher RA, 1972. Statistical Methods for Research Workers. *The Statistician*.
Marcelletti S, Scortichini M, 2016. *Xylella fastidiosa* CoDIRO strain associated with the olive quick decline syndrome in southern Italy belongs to a clonal complex of the subspecies pauca that evolved in Central America. *Microbiology (United Kingdom)* **162**, 2087–2098.
Martins PMM, Machado MA, Silva N V., Takita MA, De Souza AA, 2016. Type II toxin-antitoxin distribution and adaptive aspects on Xanthomonas genomes: Focus on Xanthomonas citri. *Frontiers in Microbiology* **7**, 1–13.
Montero-Astúa M, Chacón-Díaz C, Aguilar E *et al.*, 2008. Isolation and molecular characterization of *Xylella fastidiosa* from coffee plants in Costa Rica. *Journal of Microbiology* **46**, 482–490.
Moralejo E, Gomila M, Montesinos M *et al.*, 2020. Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe. *Communications Biology* **3**, 1–13.
Morales-cruz A, Aguirre-liguori J, Massonnet M *et al.*, 2023. Multigenic resistance to *Xylella fastidiosa* in wild grapes grapes (Vitis sps.) and its implications within a changing climate. , 1–15.
Navaud O, Barbacci A, Taylor A, Clarkson JP, Raffaele S, 2018. Shifts in diversification rates and host jump frequencies shaped the diversity of host range among Sclerotiniaceae fungal plant pathogens. *Molecular Ecology* **27**, 1309–1323.
Neuwirth E, 2022. *RColorBrewer: ColorBrewer Palettes. R package version 1.1-3.*
Nunney L, Azad H, Stouthamer R, 2019. An Experimental Test of the Host-Plant Range of

Nonrecombinant Strains of North American *Xylella fastidiosa* subsp. multiplex . *Phytopathology* **109**, 294–300.
Nunney L, Yuan X, Bromley R *et al.*, 2010. Population genomic analysis of a bacterial plant pathogen: Novel insight into the origin of pierce's disease of grapevine in the U.S. *PLoS ONE* **5**.
Page AJ, Cummins CA, Hunt M *et al.*, 2015. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693.
Page AJ, Taylor B, Delaney AJ *et al.*, 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics* **2**, e000056.
Paradis E, Schliep K, 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528.
Park AW, Farrell MJ, Schmidt JP *et al.*, 2018. Characterizing the phylogenetic specialism-generalism spectrum of mammal parasites. *Proceedings of the Royal Society B: Biological Sciences* **285**.
Perlman RL, 2016. Mouse Models of Human Disease: An Evolutionary Perspective. *Evolution, Medicine, and Public Health*, eow014.
Poorter H, Fiorani F, Pieruschka R *et al.*, 2016. Pampered inside, pestered outside? Differences and similarities between plants growing in controlled conditions and in the field. *New Phytologist* **212**, 838–855.
Prado SDS, Lopes JRS, Demétrio CGB, Borgatto AF, De Almeida RPP, 2008. Host colonization differences between citrus and coffee isolates of *Xylella fastidiosa* in reciprocal inoculation. *Scientia Agricola* **65**, 251–258.
Purcell AH, 1975. Role of the Blue-green Sharpshooter, *Hordnia circellata*, in the Epidemiology of Pierce's Disease of Grapevines. , 745–752.
Purcell AH, 1977. Cold Therapy of Pierce's Disease of Grapevines. *Plant Disease Reporter* **61**, 514.
Purcell AH, 1981. Vector Preference and Inoculation Efficiency as Components of Resistance to Pierce's Disease in European Grape Cultivars. *Phytopathology* **71**, 429.
Purcell AH, Almeida RPPP, Purcell AH, 2003. Biological Traits of *Xylella fastidiosa* Strains from Grapes and Almonds Biological Traits of *Xylella fastidiosa* Strains from Grapes and Almonds. *Applied and Environmental Microbiology* **69**, 7447–7452.
R Core Team, 2023. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.
Raju BC, Goheen AC, 1981. Relative Sensitivity of Selected Grapevine Cultivars to Pierce's Disease Bacterial Inoculation. **32**, 0–3.
Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM, 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**, 62–66.
Ranwez V, Harispe S, Delsuc F, Douzery EJP, 2011. MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLoS ONE* **6**.
Rapicavoli JN, Blanco-Ulate B, Muszyński A *et al.*, 2018. Lipopolysaccharide O-antigen delays plant innate immune recognition of *Xylella fastidiosa*. *Nature Communications* **9**, 390.
Razo-Mendivil U, Pérez-Ponce de León G, 2011. Testing the evolutionary and biogeographical history of Glypthelmins (Digenea: Plagiorchiida), a parasite of anurans, through a simultaneous analysis of molecular and morphological data. *Molecular Phylogenetics and Evolution* **59**, 331–341.

Revell LJ, 2016. Package 'phytools.' *R topics documented*, 165.
Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT, 2009. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* **25**, 2071–2073.
Rocha JG, Zambolim L, Zambolim EM, Ribeiro do Vale FX, Junior WCJ, Filho AB, 2010. Quantification of yield loss due to coffee leaf scorch. *Crop Protection* **29**, 1100–1104.
Rodriguez-R LM, Grajales A, Restrepo S, Bernal A, Salazar C, Arrieta-Ortiz M, 2012. Genomes-based phylogeny of the genus Xanthomonas. *BMC Microbiology* **12**, 43.
Rodríguez CM, Obando JJ, Villalobos W, Moreira L, Rivera C, 2001. First Report of *Xylella fastidiosa* Infecting Coffee in Costa Rica. *Plant Disease* **85**, 1027.
Sanderlin RS, 2017. Host Specificity of Pecan Strains of *Xylella fastidiosa* subsp. *multiplex*. *Plant Disease* **101**, 744–750.
Santos CA, Janissen R, Toledo MAS *et al.*, 2015. Characterization of the TolB-Pal trans-envelope complex from *Xylella fastidiosa* reveals a dynamic and coordinated protein expression profile during the biofilm development process. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1854**, 1372–1381.
Scally M, Schuenzel EL, Stouthamer R, Nunney L, 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Applied and Environmental Microbiology* **71**, 8491–8499.
Schoch CL, Ciufo S, Domrachev M *et al.*, 2020. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* **2020**, 1–21.
Seemann T, 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.
Shubin N, Tabin C, Carroll S, 2009. Deep homology and the origins of evolutionary novelty. *Nature*.
Sicard A, Castillo AI, Voeltz M *et al.*, 2020. Inference of Bacterial Pathogen Instantaneous Population Growth Dynamics. *Molecular Plant-Microbe Interactions®* **33**, 402–411.
Sicard A, Saponari M, Vanhove M *et al.*, 2021. Introduction and adaptation of an emerging pathogen to olive trees in Italy. *Microbial Genomics*.
Sicard A, Zeilinger AR, Vanhove M *et al.*, 2018. *Xylella fastidiosa* : Insights into an Emerging Plant Pathogen. *Annual Review of Phytopathology* **56**, 181–202.
Soucy SM, Huang J, Gogarten JP, 2015. Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics* **16**, 472–482.
De Souza AA, Takita MA, Coletta-Filho HD *et al.*, 2004. Gene expression profile of the plant pathogen *Xylella fastidiosa* during biofilm formation in vitro. *FEMS Microbiology Letters* **237**, 341–353.
Stamatakis A, 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
Stevenson JF, Matthews MA, Greve LC, Labavitch JM, Rost TL, 2004. Grapevine susceptibility to Pierce's disease II: Progression of anatomical symptoms. *American Journal of Enology and Viticulture* **55**, 238–245.
Su C-C, Deng W-L, Jan F-J, Chang C-J, Huang H, Chen J, 2014. Draft Genome Sequence of *Xylella fastidiosa* Pear Leaf Scorch Strain in Taiwan. *Genome announcements* **2**, 2–3.
Szurek B, Couloux A, Poussier S *et al.*, 2009. The complete genome sequence of Xanthomonas albilineans provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* **10**, 616.
Therneau T, 2022. A Package for Survival Analysis in R.
Thines M, 2019. An evolutionary framework for host shifts – jumping ships for survival. *New*

Phytologist **224**, 605–617.
Tonkin-Hill G, MacAlasdair N, Ruis C *et al.*, 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology* **21**, 1–21.
Vanhove M, Retchless AC, Sicard A *et al.*, 2019. Genomic diversity and recombination among *Xylella fastidiosa* subspecies. *Applied and Environmental Microbiology* **85**, 1–17.
Vanhove M, Sicard A, Ezennia J, Leviten N, Almeida RPP, 2020. Population structure and adaptation of a bacterial pathogen in California grapevines. *Environmental Microbiology* **22**, 2625–2638.
Velásquez AC, Castroverde CDM, He SY, 2018. Plant–Pathogen Warfare under Changing Climate Conditions. *Current Biology* **28**, R619–R634.
Wickham H, 2016. *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.
Wistrom C, Purcell AH, 2005. The fate of *Xylella fastidiosa* in vineyard weeds and other alternate hosts in California. *Plant Disease* **89**, 994–999.
Wood GW, 2004. The wild blueberry industry-past. In: *Small Fruits Review*. 11–18.
Woolhouse MEJ, Haydon DT, Antia R, 2005. Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology and Evolution* **20**, 238–244.
Yang Z, Kumar S, Nei M, 1995. A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics* **141**, 1641–1650.
Zipfel C, Felix G, 2005. Plants and animals: A different taste for microbes? *Current Opinion in Plant Biology* **8**, 353–360.