Title

Parallel host shifts in a bacterial plant pathogen suggest independent genetic solutions.

Journal

Molecular plant pathology, 24(6)

ISSN

1464-6722

Authors

Donegan, Monica A
Coletta-Filho, Helvécio D
Almeida, Rodrigo PP

Publication Date

2023-06-01

DOI

10.1111/mpp.13316

Peer reviewed

Received: 15 November 2022 | Revised: 7 February 2023 | Accepted: 9 February 2023

DOI: 10.1111/mpp.13316

Molecular Plant Pathology WILEY

ORIGINAL ARTICLE

Parallel host shifts in a bacterial plant pathogen suggest independent genetic solutions

Monica A. Donegan[1] | Helvécio D. Coletta-Filho[2] | Rodrigo P. P. Almeida[1]

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

[2]Instituto Agronômico de Campinas, Centro de Citricultura Sylvio Moreira, Cordeirópolis, Brazil

Correspondence
Rodrigo P. P. Almeida, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94709, USA.
Email: rodrigoalmeida@berkeley.edu

Funding information
California Department of Food and Agriculture; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 313676/2017-8; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2016/02176-7; Horizon 2020 Framework Programme, Grant/Award Number: 101060593 and 727987

Abstract

While there are documented host shifts in many bacterial plant pathogens, the genetic foundation of host shifts is largely unknown. *Xylella fastidiosa* is a bacterial pathogen found in over 600 host plant species. Two parallel host shifts occurred—in Brazil and Italy—in which *X. fastidiosa* adapted to infect olive trees, whereas related strains infected coffee. Using 10 novel whole-genome sequences from an olive-infecting population in Brazil, we investigated whether these olive-infecting strains diverged from closely related coffee-infecting strains. Several single-nucleotide polymorphisms, many derived from recombination events, and gene gain and loss events separated olive-infecting strains from coffee-infecting strains in this clade. The olive-specific variation suggests that this event was a host jump with genetic isolation between coffee- and olive-infecting *X. fastidiosa* populations. Next, we investigated the hypothesis of genetic convergence in the host shift from coffee to olive in both populations (Brazil and Italy). Each clade had multiple mutations and gene gain and loss events unique to olive, yet no overlap between clades. Using a genome-wide association study technique, we did not find any plausible candidates for convergence. Overall, this work suggests that the two populations adapted to infect olive trees through independent genetic solutions.

KEYWORDS
convergent evolution, host shift, *Xylella fastidiosa*

1 | INTRODUCTION

An outstanding evolutionary question is whether parallel phenotypes evolve through the same genetic solution (Rosenblum et al., 2014; Sackton & Clark, 2019). This depends on the organism and the trait in question. For bacterial species, this question has been addressed in distinct traits in diverse pathogens: convergent genetic evolution has been documented in the evolution of a vascular lifestyle in plant-pathogenic *Xanthomonas* spp. (Gluck-Thaler et al., 2020) as well as in the emergence of multiple dysentery-causing *Shigella* spp. (Thomas et al., 2019). For bacterial plant pathogens, one trait of major economic and scientific interest is host specificity. The global plant trade has enabled pathogens to spread and infect new host plants, posing a threat to agriculture and global food security. Yet there is limited knowledge about the genetics of how bacterial pathogens adapt to novel host plants. In the case of parallel host shifts, understanding the potential for genetic convergence has important applications for both evolutionary biology and applied disease management.

One bacterial insect-transmitted pathogen of global concern, *Xylella fastidiosa*, infects over 600 plant species and causes diseases

in several economically critical crops: coffee, grapevine, citrus, and, recently, olive (Delbianco et al., 2021). The genetic basis of host specificity in *X. fastidiosa* is poorly understood. While there is no pathogen specificity to polyphagous insect vectors (Krugner et al., 2019), recent genomic studies provide evidence that there is a genetic component of host plant specificity (Batarseh et al., 2022; Kahn & Almeida, 2022). Yet specific genes and mechanisms have remained elusive. In Italy, 2013 marked the detection of a devastating outbreak of *X. fastidiosa* in olive trees. *X. fastidiosa* causes olive quick decline syndrome, in which the bacterium colonizes xylem vessels, resulting in reduced water availability, slower nutrient transport, and after a few years plant death (Martelli, 2016). More recently, a parallel host shift of *X. fastidiosa* to olive was reported in 2016, when disease was observed and associated with the bacterium in Brazilian olive groves (Coletta-Filho et al., 2016). The olive-infecting strains of *X. fastidiosa* in Italy and Brazil are from separate phylogenetic clades, that is, they are genetically distinct. The occurrence of a parallel host shift—the shift to olive trees in distinct *X. fastidiosa* strains in Italy and in Brazil—enables us to test hypotheses about genetic convergence.

Pathogen host shifts can occur in two distinct ways, host jumps and host range expansions, but the type of event that occurred in Brazil is currently unclear. In a host jump, a pathogen adapts to efficiently colonize a new host. This adaptation results in genetic isolation between the pathogen genotypes on the new host and the parent population on the old host (Thines, 2019). A host range expansion, on the other hand, does not affect the gene pool between the pathogen populations on the old and new host plants. Host range expansions are common where local pathogens are already able to efficiently colonize an introduced host (Thines, 2019). Previous host jumps in *X. fastidiosa* include a jump from coffee to citrus in Brazil, supported by genetic data and biological cross-inoculation experiments (Almeida et al., 2008; Coletta-Filho et al., 2020; Francisco et al., 2017). In Italy, available data suggest that the emergence of olive quick decline syndrome was a host jump, as the olive-colonizing strains from Italy show reduced fitness in coffee (Sicard et al., 2021). Given this history, we hypothesize that the shift to olive in Brazil was a host jump, involving genetic differentiation.

In distinct clades in Brazil and Italy, the bacterium appears to have independently shifted, or jumped, hosts from infecting coffee to infecting olive. The host jump in Italy was previously described (Giampetruzzi et al., 2017; Sicard et al., 2021); genomic analyses traced the outbreak back to a single introduction of infected plant material from Central America; the closest known strain infected coffee in Costa Rica (Castillo et al., 2020; Giampetruzzi et al., 2017; Vanhove et al., 2019). The outbreak in Brazil has not yet been described using whole-genome data. Multilocus sequence typing (MLST) data provided preliminary evidence to suggest that coffee-infecting *X. fastidiosa* strains may be ancestral to olive-infecting strains in Brazil (Safady et al., 2019). Previous analysis of MLST data from olive-infecting *X. fastidiosa* in Brazil indicated that most olive strains in the region are from one phylogenetic clade (sequence type [ST] 16) (Safady et al., 2019). This clade is genetically distinct from the olive-infecting group in Italy (ST53). Although ST53 (Italy and Central America) and ST16 (Brazil) are both in *X. fastidiosa* subsp. *pauca*, the two clades diverged thousands of years ago and have a separate evolutionary history (Vanhove et al., 2019). The split between ST53 strains currently in Central America and the rest of subsp. *pauca* in South America was estimated at 942 BCE (Vanhove et al., 2019), so these two clades have diverged allopatrically for thousands of years. *X. fastidiosa* has also been found in olive in California, although the presence of the bacterium was not correlated with disease, suggesting that this event was not a host shift (Krugner et al., 2014). There are also preliminary reports of distinct sequence types (ST69 and S78) of *X. fastidiosa* subsp. *pauca* in olive in Argentina (Tolocka et al., 2021). This event may represent a host shift, but whole-genome data are not yet available.

Genetic convergence may occur through point mutation, genome rearrangements, homologous recombination, or horizontal gene transfer, as *X. fastidiosa* is naturally competent (Kung & Almeida, 2014). During homologous recombination, bacteria swap alleles at the same loci, whereas in horizontal gene transfer accessory genes are gained or lost, but the bacterium does not have a homologous gene to begin with. For other cases of host shifts, there are documented examples of gene gain and loss (Richardson et al., 2018) or even a single single-nucleotide polymorphism (SNP) (Viana et al., 2015) underlying a host switch for the same bacterial pathogen, *Staphylococcus aureus*. In *Xanthomonas*, a genus closely related to *Xylella*, horizontal gene transfer was implicated in adaptation to common bean across distinct *Xanthomonas* lineages (Chen et al., 2018) The relative contributions of point mutation, homologous recombination, and horizontal gene transfer to bacterial evolution seems to depend on the species and the phenotype (Ambur et al., 2016).

We investigated two independent host shifts to olive in *X. fastidiosa* to understand the genetic basis of parallel host shifts in plant pathogens. First, we asked if olive-infecting strains in ST16 were genetically distinct from closely related coffee-infecting strains. We hypothesized that this host shift event was a host jump from coffee to olive, including changes in the gene pool. To test this, we used whole-genome sequencing data from 10 olive-infecting strains from Brazil to examine genetic diversity and ancestral relationships within ST16. Next, we programmatically searched for shared adaptations to olive (either genes gained and lost or SNPs) between olive-infecting clades in Brazil and Italy. Using genomic data from a large dataset of *X. fastidiosa* strains, we tested the hypothesis that there is genetic convergence between olive-infecting ST16 and ST53.

2 | RESULTS

2.1 | Olive-infecting strains in ST16 form a single group, divergent from coffee-infecting strains

The 10 olive-infecting and three coffee-infecting strains in ST16 form a monophyletic group within subsp. *pauca* (Figure 1a). This group includes both the nine olive-infecting strains from Minais

Gerais, as well as the olive-infecting strain from São Paulo. Although the citrus strains are geographically close, ST16 is phylogenetically distinct from the citrus clade. While the split between ST16 and the citrus clade was highly supported (bootstrap value of 100), the splits within ST16 were poorly supported (Figure S1e). To better infer relationships within ST16, we re-aligned only the 13 sequences within this group. Including only these sequences resulted in a larger core genome from which to build the phylogeny: 88.1% of genes were in the core genome for the ST16 alignment versus 50.8% of genes in the core for the subsp. *pauca* alignment (Figure S3). We then removed recombination from the core genome, which can obscure evolutionary relationships and change the tree topology (Figure S1b–e). After removing recombination from the core genome of ST16, there was a highly supported split in an unrooted phylogeny between the coffee-infecting strains (3124_BR, 32_BR, and RAAR15_co15_BR) and the 10 olive-infecting strains in ST16 (Figure 1b). A median-joining network built from the recombination-free core genome also mirrored this clear split between coffee-infecting and olive-infecting strains in ST16 (Figure 1c). The majority of mutations (depicted as notches in the network) are largely between the coffee- and olive-infecting split (Figure 1c). In both the unrooted phylogeny and joining network, the 10 olive-infecting strains were closely related, supporting a single host shift event.

2.2 | Several SNPs and gene gain/loss events differentiate ST16 olive-infecting strains

We examined genetic diversity within ST16; population genetics statistics were consistent with a very recent host jump to olive (Table 1). First, removing recombination from the core genome affected genetic diversity; after removing 11,507 inferred recombinant sites from the core genome, the number of segregating sites decreased from 3557 to 95. With or without recombination, the three coffee-infecting strains had greater nucleotide diversity than the olive-infecting strains (Table 1). Using the core genome of ST16 only, we examined all sites that differed between all olive-infecting and coffee-infecting strains. For this analysis, we included recombination because recombinant events may be involved in adaptation to the olive host plant. There were 469 core-genome SNPs shared by all olive-infecting strains that differentiated them from coffee-infecting strains. The vast majority of these (376 SNPs; 80%) occurred in inferred recombinant

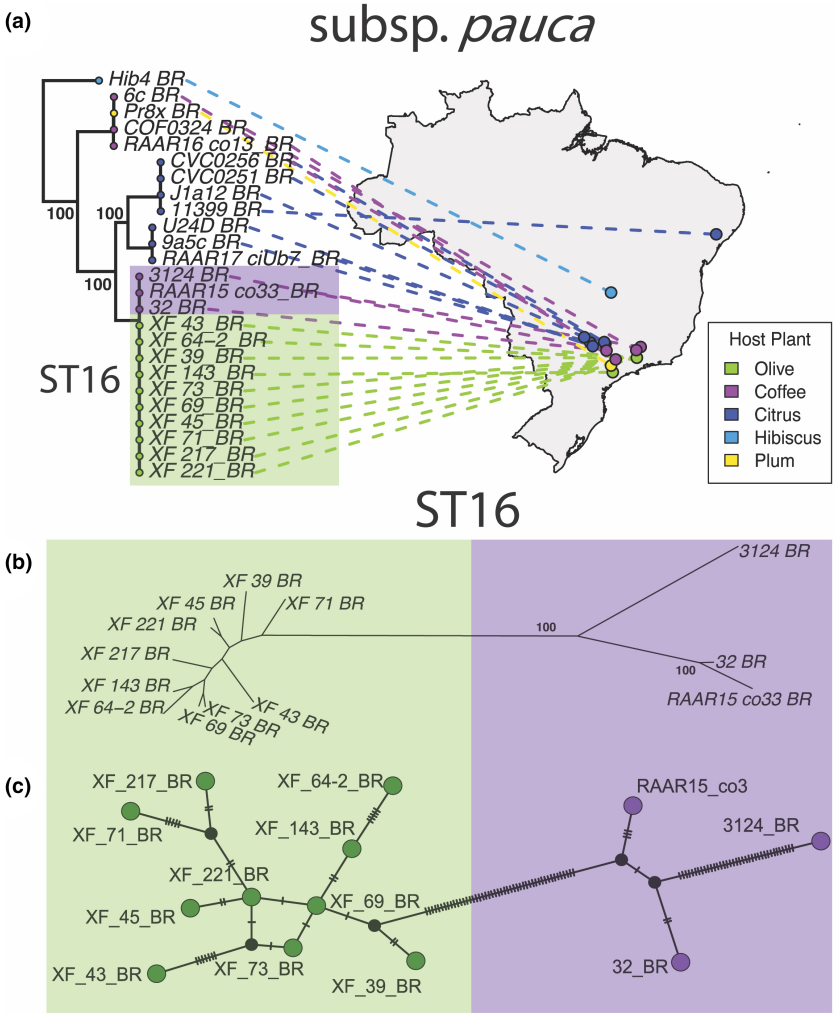

FIGURE 1 Distinct olive- and coffee-infecting clades with ST16. (a) The maximum-likelihood tree built from the nonrecombinant core genome of the subsp. *pauca*, mapped by sampling location. ST16 is highlighted (purple/green) and includes 10 olive strains and three coffee strains; the sequence types of all other strains are included in Table S4. (b) An unrooted, maximum-likelihood tree from the nonrecombinant core genome of ST16. There was support (bootstrap value of 100) for the split between the coffee cluster and the rest of the olive-infecting strains. (c) A median-joining network built from the nonrecombinant core genome of ST16. The notches on the tree depict single-nucleotide polymorphisms at that split.

TABLE 1 Population genetics statistics from an alignment of the core genome of 13 strains in ST16.

	Clade	*n*	SNPs	π	θ	Tajima's *D*
With recombination	All olive	10	1508	$2.473\times10e^{-4}$	$2.640\times10e^{-4}$	−0.317
	All coffee	3	2811	$9.280\times10e^{-4}$	$9.280\times10e^{-4}$	na[a]
	All ST16	13	3577	$5.021\times10e^{-4}$	$5.708\times10e^{-4}$	−0.557
Without recombination	All olive	10	27	$3.481\times10e^{-6}$	$4.747\times10e^{-6}$	−1.274
	All coffee	3	43	$1.426\times10e^{-5}$	$1.426\times10e^{-5}$	na[a]
	All ST16	13	95	$1.310\times10e^{-5}$	$1.522\times10e^{-5}$	−0.636

Note: There were 2,092,758 sites in the core genome before recombination was removed; following removal, there were 2,081,251 sites.
π, nucleotide diversity; θ, Watterson's θ, an estimate of population mutation rates; SNP, single-nucleotide polymorphism.
[a]Not applicable: Tajima's θ equals Watterson's θ in this case, so Tajima's *D* cannot be computed.

regions of the core genome. These SNPs were distributed across 13 genes, including 11 hypothetical proteins, *socA* (the antitoxin of the socAB toxin-antitoxin system), and *ssb* (single-stranded DNA-binding protein). Genetic changes in the *ssb* and *socA* genes resulted in changes to the protein, as each gene had amino acid insertions and nonsynonymous changes (Figure S2). In *ssb*, there was a 9 bp insertion, as well as six nonsynonymous and 15 synonymous SNPs; in *socA*, there was a 12 bp insertion with 16 synonymous and 13 nonsynonymous SNPs. Of the hypothetical proteins with SNPs, almost all genes included nonsynonymous SNPs, and four had phage-related products: tail fibre protein (group_169), baseplate J/gp47 family protein (group_173), Bro-N domain-containing protein (group_123), and phage terminase large subunit family protein (group_1516) (Table S5). In relation to the accessory genome of the coffee-infecting group, the olive strains shared six genes gained and 10 genes lost, the majority of which are hypothetical proteins (Table S5). The sole nonhypothetical protein was annotated as *lpxD_1*, an acyltransferase involved in lipid A biosynthesis. All 13 strains in ST16 retained *lpxD_2*, *lpxD_3*, and *lpxD_4*, but the coffee strains were missing *lpxD_1*.

2.3 | ST16 and ST53 do not share variation unique to olives

To examine potential molecular parallelism between ST16 and ST53, we aligned a set of 39 *X. fastidiosa* subsp. *pauca* strains (Table S4). Because the core genome (genes present in 95% of strains) decreased from 88.1% to 50.8% in alignments of ST16 to subsp. *pauca* (Figure S3), the counts of olive-specific SNPs/gene gain and loss events in ST16 shifted between alignments. This occurred because some genes shift from the core to the accessory genome when including additional strains in the alignment. Within subsp. *pauca*, we used Scoary to identify olive-specific variation in the core and accessory genome within ST16 and ST53, that is, genes or SNPs present/absent in all olive-infecting strains relative to all sister taxa

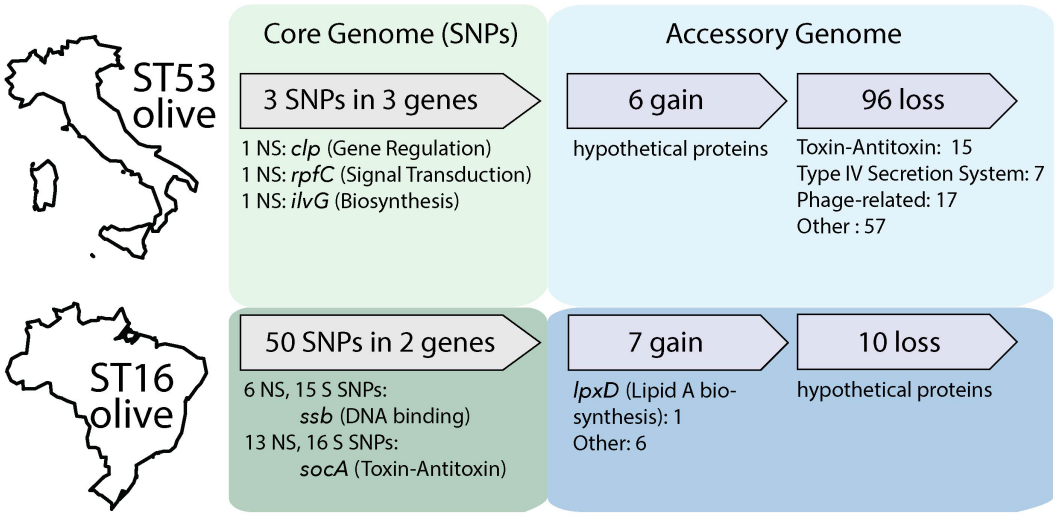

FIGURE 2 ST16 and ST53 do not share olive-specific variation. ST53 and ST16 each have olive-specific core genome single-nucleotide polymorphisms (SNPs). In ST53, the SNPs include several single nonsynonymous substitutions, while the variants in ST16 occurred in genes with inferred recombination. The hypothetical genes and genes with gapped alignments were excluded from the list of the core genome SNPs. None of the hypothetical proteins overlapped between the two clades. NS, nonsynonymous; S, synonymous substitution.

(i.e., coffee-infecting strains) within the clade. ST16 and ST53 both showed olive-specific gene gain and loss events, as well as olive-specific SNPs (Figure 2). The ST16 core genome mutations (*ssb*, *socA*) and gain/loss events (one copy of *lpxD* and hypothetical proteins) were detailed earlier. In ST53, we identified core-genome SNPs in seven genes; two of the genes were excluded because of gapped alignments and two of the genes were hypothetical proteins. The three remaining genes (*clp*, *rpfC*, and *ilvG*) each contained a single nonsynonymous mutation. None of these mutations occurred in inferred recombinant regions. In the ST53 accessory genome, 96 gene loss and six gene gain events were unique to olive (Figure 2). Of the 96 gene loss events, there were 15 genes involved in toxin-antitoxin systems, many of which were on a large plasmid deletion (Table S5). Additionally, seven genes were annotated as Type IV secretion system proteins, as well as 18 phage-related accessory gene losses. Neither genes with SNPs nor gene gain and loss events were shared between the two clades, that is, there were no genes with convergent mutations among olive-specific variants in ST53 and ST16.

2.4 | One gene loss event is associated with a host shift to olive

Using the same set of *X. fastidiosa* subsp. *pauca* genomes ($n = 39$), we used TreeWAS to identify SNPs and accessory genes associated with (rather than unique to) olive strains. We repeated this technique to detect core SNPs or accessory genes associated with citrus- and coffee-infecting strains as well. Neither core SNPs nor accessory genes were associated with coffee strains or citrus strains, and no core SNPs were associated with olive strains. One gene loss (hypothetical protein: *group_1969*) was significantly associated with olive-infecting strains (Figure 3). While *group_1969* was absent from all olive strains from ST53, it was present in some olive strains within ST16. A BLAST search revealed that this protein was a tail fibre domain-containing protein. Although statistically significant, this gene loss is probably not necessary for *X. fastidiosa* adaptation to olive due to its presence in ST16.

3 | DISCUSSION

We investigated the genetic and evolutionary history of the emergence of *X. fastidiosa* in olive trees in Brazil, as well as molecular parallelism between host shifts to olives in ST16 (Brazil) and ST53 (Italy). We found that olive-infecting strains in ST16 form a monophyletic clade with coffee-infecting strains within subsp. *pauca*, but the split between the two groups within ST16 was well-supported, suggesting a recent single host jump. There were several accessory genes and SNPs that differentiated all 10 olive-infecting strains from coffee-infecting strains. In examining the two olive-infecting clades (Brazil and Italy) for genetic convergence, we found that neither gene gains and losses nor SNPs converged between the Brazilian and Italian olive-infecting populations.

3.1 | Olive-infecting strains in Brazil (ST16) are genetically distinct from coffee-infecting strains

The genomic data from ST16 in Brazil are consistent with our hypothesis of a single host jump, as the observation of several SNPs and accessory genes unique to olive supports that this clade is genetically isolating. Our whole-genome data confirm the observation of a single emergence of olive-infecting strains from the MLST data (Safady et al., 2019). In the case of a host range expansion, the ancestral coffee-infecting population would also carry the genetic changes that enable olive colonization. While we do not know whether these genetic changes are responsible for olive adaptation, shared olive-infecting variants (469 SNPs and 16 gene gain and loss events) demonstrate that the olive population in ST16 is genetically distinct from related coffee strains. While the olive-infecting clade and three coffee-infecting strains in Brazil are still monophyletic within subsp. *pauca*, this may be because the host shift occurred very recently. Removing recombination from an alignment of only the 13 strains in ST16 revealed a clear split in this clade. The phylogeny and neighbour network both support a single split between olive- and coffee-infecting strains. Due to the prevalence of recombination in *X. fastidiosa*, removing recombination is important to resolve phylogenetic relationships. For example, removing recombination changes the tree topology of the entire *X. fastidiosa* species phylogeny (Kahn & Almeida, 2022). Plus, the two citrus clades in subsp. *pauca* shift from distinct clades to a monophyletic clade when recombination was removed (Coletta-Filho et al., 2020), a finding that was repeated.

While recombination can impair correct inference of phylogenetic relationships, homologous recombination may be involved in host plant adaptation. Of the olive-specific SNPs in the core genome of ST16, most of these were in inferred recombinant regions; 80% of 469 SNPs were annotated as recombinant. Homologous recombination may contribute to olive adaptation in ST16, as it has been implicated in past host jumps in South America. For example, the jump and subsequent virulence of citrus strains derived from coffee were linked to intersubspecific homologous recombination (Nunney et al., 2012). Given the sympatric distributions of multiple strains of *X. fastidiosa* within Brazil, homologous recombination may be important to host jumps in this region.

Biological data about the host range of the ST16 strains are pertinent to distinguish if olive-infecting strains and closely related coffee-infecting strains have speciated, that is, can coffee-infecting strains in ST16 colonize olive, and vice versa? Both biological and genetic data demonstrated speciation between coffee- and citrus-infecting strains in Brazil. Although coffee- and citrus-infecting strains in Brazil share a sympatric distribution, cross-infection experiments revealed that the groups are biologically distinct (Almeida et al., 2008). Even though coffee-infecting strains were ancestral, they do not infect citrus, and citrus strains only weakly infected coffee (Almeida et al., 2008; Francisco et al., 2017). Cross-infection experiments with olive and coffee strains will be important for determining whether these groups are biologically isolated, which has relevant management implications. For ST53 strains, preliminary cross-infection studies have demonstrated

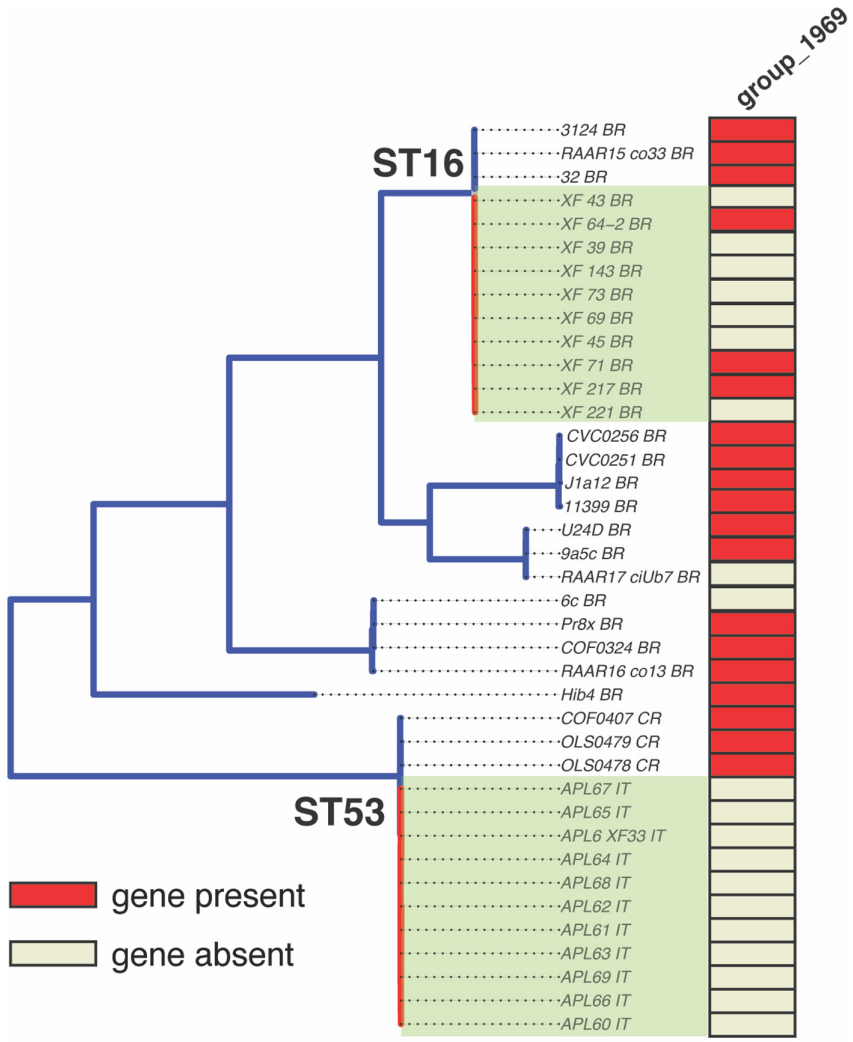

FIGURE 3 A heat map with gene presence/absence for the hypothetical protein associated with olive. Strains highlighted in green are olive-infecting strains. The tan rectangles denote gene absence, whereas red rectangles denote gene presence. The hypothetical protein identified, *group_1969*, was absent in all ST53 olive-infecting strains, but present in several ST16 olive-infecting strains.

that one olive-infecting strain showed slower colonization of coffee than olive, and that olive plants were a better source of *X. fastidiosa* for vector transmission (Sicard et al., 2021). Together, these findings support selection in olive-infecting *X. fastidiosa* for increased virulence and growth rate in olive, but further experimental research is necessary. Future biological data will build on the genomic evidence here to conclude if the coffee to olive host shift was indeed a host jump. Such in vivo data will also improve the predictive capacity of association-based genomic techniques, which rely on phenotype inputs.

3.2 | Lack of genetic convergence between two olive-infecting clades

Our results suggest that the host shift to olive evolved through independent genetic solutions in Brazil and Italy. While our study focused exclusively on coding regions, it is possible that convergent changes in noncoding genetic regions (e.g., promoters) could underlie these host shifts. Variation in noncoding regions is a promising area for future study. Within coding regions, a search for olive-specific genes and SNPs in ST53 and ST16 did not show widespread molecular parallelism between the two clades. Several of the variants identified here in ST53, including transcriptional regulatory gene *clp*, acetolactate synthase subunit-encoding gene *ilvG*, and quorum-sensing kinase *rpfC*, were previously described in Sicard et al. (2021). All three of these genes were found in the bacterial chromosome, and *clp* and *rpfC* were found to be under positive selection. Unlike core genome variants in ST16, these genes were not annotated as recombinant and only contained single nonsynonymous mutations. The contrasting nature of variants—nonsynonymous point mutations versus homologous recombination—may indicate that different evolutionary processes were involved for the two clades. In the accessory genome, many of the gene loss events in ST53, including several toxin-antitoxin genes, occurred due to a large deletion on the plasmid (Sicard et al., 2021). ST16 also included a mutation in *socA*, a toxin-antitoxin system. In *X. fastidiosa* subsp. *fastidiosa*, a knockout strain of the *dinJ/relE* toxin-antitoxin system exhibited a hypervirulent phenotype in grapevines, implicating these systems in controlling population growth (Burbank & Stenger, 2017). Given the diversity of toxin-antitoxin systems in prokaryotes (LeRoux & Laud, 2022), it is unlikely that this represents convergent functions.

The divergent history and ecology of *X. fastidiosa* in the two regions may explain the lack of molecular parallelism. No other known *X. fastidiosa* strains infecting olive trees exist in Apulia, where this pathogen adapted to a monoculture environment. The ancestral host, coffee, is not cultivated in Italy. However, the sympatric distribution of *X. fastidiosa* subsp. *pauca* genotypes in Brazil, as well as the widespread cultivation of coffee, may impact its adaptation and speciation from olive in Brazil due to gene flow. Thus, these two populations may represent different stages in the host jump trajectory; ST53 olive strains may be further along in speciation from ancestral strains due to their allopatric distribution and the lack of ancestral host plants in Italy. It may be interesting to compare pathogen virulence in ST53 to ST16, with the hypothesis that ecological isolation will select for more aggressive strains in ST53 (McDonald & Stukenbrock, 2016).

Because there was little evidence of genetic convergence, we cannot conclude whether point mutation, homologous recombination, or horizontal gene transfer drives adaptation to olive for *X. fastidiosa*. In fact, these findings suggest that different processes (point mutation, homologous recombination) may be implicated in olive adaptation in distinct clades. Horizontal gene transfer is still relevant, as analyses across all known *X. fastidiosa* subspecies found that several pan-genome gain and loss events were associated with host specificity to other plant genera in *X. fastidiosa* (Kahn & Almeida, 2022). Ultimately, there may not be one dominant mechanism across the species, subspecies, or even clade level. Rather, evolutionary history and local ecology may differentially shape the host range of *X. fastidiosa*. As worldwide plant pathogen introductions continue, there may be more opportunities for convergent pathogen host shifts. Managing and predicting pathogen outbreaks will require continued research into the molecular origins of host adaptation in bacterial pathogens.

4 | EXPERIMENTAL PROCEDURES

4.1 | *X. fastidiosa* sampling, whole-genome sequencing, and assembly

Ten *X. fastidiosa* subsp. *pauca* strains were collected from symptomatic olives from two geographically distant field sites in Brazil; each strain was sampled from a different olive variety (Table S1). The strains were obtained from January to April of 2017, following Safady et al. (2019). DNA was extracted using a commercial kit (Wizard DNA Genomic Kit; Promega) and sent for paired-end sequencing at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley. The genomes were sequenced using Illumina HiSeq2000. All raw reads and associated metadata are available on the NCBI SRA database (PRJNA715523). Draft genomes were assembled using the pipeline described in Castillo et al. (2020). After assembly, the quality was assessed with QUAST v. 5.0.2 (Gurevich et al., 2013). All 10 genomes showed an L50 of six or seven contigs and an N50 of between 98,870 and 149,454 bp. More detail on genome assembly for each strain is included in Table S2.

4.2 | Pan-genome alignment using Panaroo

Previous work comparing *X. fastidiosa* genomes (Castillo et al., 2020; Kahn & Almeida, 2022; Sicard et al., 2021) has used Roary (Page et al., 2015) for pan-genome alignment. However, we noted consistent issues with Roary, including fragmenting genes into separate clusters and splitting core genes into multiple separate clusters in the accessory genome, thus inflating the accessory genome. These issues with Roary are detailed in Tonkin-Hill et al. (2020). The underestimation of the core genome (due to gene fragmentation and gene splitting) can lead to erroneous estimates of nucleotide diversity and phylogenetic inference based on the core genome. Panaroo (Tonkin-Hill et al., 2020) is a recent package that was developed specifically to address these issues by using a graphical approach to cluster orthologous genes. We compared Roary v. 3.13.0 and Panaroo v. 1.2.9 using similar settings for each. With Roary, we used the default settings, along with the merge paralogs (−s) and allow paralogs in the core genome (−ap) flags. With Panaroo, we also used default settings along with the −clean mode strict and −merge paralogs flags. We used the same cut-offs for the core, shell, and cloud genome for both programs. A comparison of the pan-genome is included in Table S3. For alignments of ST16 ($n = 13$) and subsp. *pauca* ($n = 39$), Panaroo alignments were smaller than Roary results, matching comparisons published in Tonkin-Hill et al. (2020). Thus, moving forward, we used Panaroo for alignments to reduce accessory genome inflation.

4.3 | ST16 population structure and diversity

Along with the 10 genomes from olive, three previously reported genomes from coffee in Brazil (32: BioSample SAMN02471372, 3124: BioSample SAMN03166214, RAAR15_Co33: BioSample SAMN12994918) that are closely related to the olive samples were included to examine population structure in ST16. Using PubMLST, we confirmed that the sequence type of all 13 strains was ST16 (https://pubmlst.org/organisms/xylella-fastidiosa). The 13 strains (10 from symptomatic olive, three from symptomatic coffee plants) were annotated using Prokka v. 1.14.6 (Seemann, 2014), and the pan-genome was assembled with Panaroo v. 1.2.9 (Tonkin-Hill et al., 2020). To define the pan-genome, we used standard cut-offs (Page et al., 2015): the core genes at frequency between 95% and 100%, shell between 15% and 95%, and cloud under 15%. Originally, we included another isolate (COF0324: BioSample SAM038621222) from outside of ST16 as an outgroup. However, due to the low sequence diversity in ST16, a core genome phylogeny including an outgroup resulted in poor bootstrap support for splits in the ST16 clade (Figure S1a). Thus, we excluded COF0324 from the subsequent alignment of the 13 ST16 strains. We used ClonalFrameML v. 1.12 (Didelot & Wilson, 2015) to infer recombinant regions of the core genome; inferred recombinant regions, totalling 11,507 bp, were removed to build an alignment of the nonrecombinant core. Using the core genome from Panaroo with (2,092,758 bp) and without (2,081,251 bp) recombinant regions, separate maximum-likelihood trees were built with RAxML v. 8.2.11

using 1000 bootstraps (Stamatakis, 2014); trees were visualized in R. Additionally, using the nonrecombinant core genome, a median-joining neighbour net was constructed in PopART (Leigh & Bryant, 2015). Population genetics statistics were calculated using the PopGenome package v. 2.7.5 in R (Pfeifer et al., 2014).

4.4 | Evolution within subsp. *pauca*

The same pipeline as above (Prokka, Roary, ClonalFrameML, RAxML) was applied to a dataset of 39 *X. fastidiosa* subsp. *pauca* genomes. In addition to the 13 strains in ST16, this dataset included 12 other strains from Brazil and 14 strains from ST53, including 11 from a library of olive-infecting strains from Italy (Table S4). Again, we inferred recombinant regions using ClonalFrameML and subsequently removed 252,685 bp from the core genome. We built two maximum-likelihood trees, using the core genome with and without recombination, which were visualized using the Phytools package v. 0.7.70 in R (Revell, 2012), using longitude and latitude metadata obtained during sampling.

Next, we used the Fisher's exact *t* test for trait association implemented in Scoary v. 1.6.16 (Brynildsrud et al., 2016) to investigate core genome SNPs and accessory genes that were gained and lost within clades that olive-infecting strains evolved. For the SNPs input, the core genome was converted to a variant call file (vcf) using the snp-sites package (Page et al., 2016). For gene gain and loss, the accessory gene presence/absence file from Panaroo was used an input. Scoary was run separately on a subset of (a) ST16 only ($n = 13$) and (b) ST53 only ($n = 14$). That allowed us to look for core SNPs/gene association within that clade and eliminate spurious associations based on shared ancestry. Reported associations (SNPs or accessory genes) are genes that showed a significant *p* value, using an additional cut-off so that the genes/SNPs were perfectly split into olive/not olive strains (i.e., the gene or SNP had a perfect association with trait).

4.5 | Genome-wide association at the subspecies level

To test for association of mutations (SNPs, gene gain and loss) with the olive-infecting trait, we used a dataset of 39 *X. fastidiosa* genomes from *X. fastidiosa* subsp. *pauca* (Table S4). TreeWAS performs a genome-wide association study-like analysis that accounts for microbial genomic features, including recombination and clonal population structure by simulating a phylogenetic-aware null dataset using a tree (Collins & Didelot, 2018). Then, the simulated dataset is compared to the actual data to find correlations not expected by chance. Using TreeWAS v. 1.0, we searched for association with olive host plant with (1) core genome SNPs and (2) accessory genes; the maximum-likelihood tree from RAxML was used as an input. This analysis was repeated for association with coffee and citrus by amending the binary phenotype mapping file. TreeWAS outputs three correlation scores: (1) the terminal score, which simply uses phenotype association at terminal nodes; (2) the simultaneous score, which uses inferred ancestral state at internal branches to identify simultaneous evolution of the genotype/phenotype; and (3) the subsequent score, which uses both internal and terminal nodes (Collins & Didelot, 2018).

ACKNOWLEDGEMENTS

The authors thank Andreina Castillo Siri, Alexandra Kahn, and Isabel Bojanini for input on bioinformatics methods and helpful discussions. Many thanks to Anusha Bishop and Gabriel Doble for revising this manuscript. This work was partially funded by the California Department of Food and Agriculture Pierce's Disease Control Program. Additional funding was from Horizon 2020 (XF Actors project number 727987), BeXyl Project (HORIZON, id: 101060593), and FAPESP (São Paulo Research Foundation 2016/02176-7). H.D.C.F. received CNPq research fellowships (proc. no. 313676/2017-8).

DATA AVAILABILITY STATEMENT

The genomic data generated here were deposited in and are publicly available on the NCBI Sequence Reads Archive (PRJNA715523) at https://dataview.ncbi.nlm.nih.gov/object/PRJNA715523?reviewer=q6kvfh2hs43e0o2csc41c9hmcj.

ORCID

Monica A. Donegan https://orcid.org/0000-0002-6780-7744
Helvécio D. Coletta-Filho https://orcid.org/0000-0002-1382-4629
Rodrigo P. P. Almeida https://orcid.org/0000-0003-2888-9617

REFERENCES

Almeida, R.P.P., Nascimento, F.E., Chau, J., Prado, S.S., Tsai, C.W., Lopes, A.S. 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.

Ambur, O.H., Engelstädter, J., Johnsen, P.J., Miller, E.L. & Rozen, D.E. (2016) Steady at the wheel: conservative sex and the benefits of bacterial transformation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 1706.

Batarseh, T., Morales-Cruz, A., Ingel, B., Roper, M.C. & Gaut, B.S. (2022) Using genomes and evolutionary analyses to screen for host-specificity and positive selection in the plant pathogen *Xylella fastidiosa*. *Applied and Environmental Microbiology*, 88, e01220.

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, 238.

Burbank, L.P. & Stenger, D.C. (2017) The DinJ/RelE toxin-antitoxin system suppresses bacterial proliferation and virulence of *Xylella fastidiosa* in grapevine. *Phytopathology*, 107, 338–394.

Castillo, A.I., Chacón-Díaz, C., Rodríguez-Murillo, N., Coletta-Filho, H.D. & Almeida, R.P.P. (2020) Impacts of local population history and ecology on the evolution of a globally dispersed pathogen. *BMC Genomics*, 21, 369.

Chen, N.W.G., Serres-Giardi, L., Ruh, M., Briand, M., Bonneau, S., Darrasse, A. et al. (2018) Horizontal gene transfer plays a major role in the pathological convergence of Xanthomonas lineages on common bean. *BMC Genomics*, 19, 606.

Coletta-Filho, H.D., Francisco, C.S., Lopes, J.R.S., De Oliveira, A.F. & Da Silva, L.F.O. (2016) First report of olive leaf scorch in Brazil, associated with *Xylella fastidiosa* subsp. *pauca*. *Phytopathologia Mediterranea*, 55, 130–135.

Coletta-Filho, H.D., Castillo, A.I., Laranjeira, F.F., Chumbinho De Andrade, E., Silva, N.T., Alves De Souza, A. et al. (2020) Citrus variegated chlorosis: an overview of 30 years of research and disease management. *Tropical Plant Pathology*, 45, 175–101.

Collins, C. & Didelot, X. (2018) A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Computational Biology*, 14, e1005958.

Delbianco, A., Gibin, D., Pasinato, L. & Morelli, M. (2021) Update of the *Xylella* spp. host plant database – systematic literature search up to 31 December 2020. *EFSA Journal*, 19, 6674.

Didelot, X. & Wilson, D.J. (2015) ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Computational Biology*, 11, e1004041.

Francisco, C.S., Ceresini, P.C., Almeida, R.P.P. & Coletta-Filho, H.D. (2017) Spatial genetic structure of coffee-associated *Xylella fastidiosa* populations indicates that cross infection does not occur with sympatric citrus orchards. *Phytopathology*, 107, 395–402.

Giampetruzzi, A., Saponari, M., Loconsole, G., Boscia, D., Savino, V.N., Almeida, R.P.P. et al. (2017) Genome-wide analysis provides evidence on the genetic relatedness of the emergent *Xylella fastidiosa* genotype in Italy to strains from Central America. *Phytopathology*, 107, 816–827.

Gluck-Thaler, E., Cerutti, A., Perez-Quintero, A.L., Butchacas, J., Roman-Reyna, V., Madhavan, V.N. et al. (2020) Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles. *Science Advances*, 6, eabc4516.

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

Kahn, A.K. & Almeida, R.P.P. (2022) Phylogenetics of historical host switches in a bacterial plant pathogen. *Applied and Environmental Microbiology*, 88, e0235621.

Krugner, R., Sisterson, M.S., Chen, J., Stenger, D.C. & Johnson, M.W. (2014) Evaluation of olive as a host of *Xylella fastidiosa* and associated sharpshooter vectors. *Plant Disease*, 98, 1186–1193.

Krugner, R., Sisterson, M.S., Backus, E.A., Burbank, L.P. & Redak, R.A. (2019) Sharpshooters: a review of what moves *Xylella fastidiosa*. *Austral Entomology*, 58, 248–267.

Kung, S.H. & Almeida, R.P.P. (2014) Biological and genetic factors regulating natural competence in a bacterial plant pathogen. *Microbiology*, 160, 37–46.

Leigh, J.W. & Bryant, D. (2015) POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116.

LeRoux, M. & Laud, M.T. (2022) Toxin-antitoxin systems as phage defense elements. *Annual Review of Microbiology*, 76, 21–43.

Martelli, G.P. (2016) The current status of quick decline syndrome of olive in southern Italy. *Phytoparasitica*, 44, 1–10.

McDonald, B.A. & Stukenbrock, E.H. (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20160026.

Nunney, L., Yuan, X., Bromley, R.E. & Stouthamer, R. (2012) Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Applied and Environmental Microbiology*, 78, 4702–4714.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G. et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31, 3691–3693.

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A. et al. (2016) SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, 2, e000056.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E. & Lercher, M.J. (2014) PopGenome: an efficient Swiss Army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31, 1929–1936.

Revell, L.J. (2012) Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.

Richardson, E.J., Bacigalupe, R., Harrison, E.M., Weinert, L.A., Lycett, S., Vrieling, M. et al. (2018) Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature Ecology and Evolution*, 2, 1468–1478.

Rosenblum, E.B., Parent, C.E. & Brandt, E.E. (2014) The molecular basis of phenotypic convergence. *Annual Review of Ecology, Evolution, and Systematics*, 45, 203–229.

Sackton, T.B. & Clark, N. (2019) Convergent evolution in the genomics era: new insights and directions. *Philosophical Transactions of the Royal Society B*, 374, 20190102.

Safady, N.G., Lopes, J.R.S., Francisco, C.S. & Coletta-Filho, H.D. (2019) Distribution and genetic diversity of *Xylella fastidiosa* subsp. *pauca* associated with olive quick syndrome symptoms in southeastern Brazil. *Phytopathology*, 109, 257–264.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30, 2068–2069.

Sicard, A., Saponari, M., Vanhove, M., Castillo, A.I., Giampetruzzi, A., Loconsole, G. et al. (2021) Introduction and adaptation of an emerging pathogen to olive trees in Italy. *Microbial Genomics*, 7, 000735.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Thines, M. (2019) An evolutionary framework for host shifts – jumping ships for survival. *New Phytologist*, 224, 605–617.

Thomas, A.K., Preetha, S., Omanakuttan, A., Vidyullata, L., Ashokan, A., Rajachandran, V. et al. (2019) Mutational convergence acts as a major player in adaptive parallel evolution of *Shigella* spp. *Scientific Reports*, 9, 3252.

Tolocka, P.A., Guzmán, F.A., Paccioretti, M.D., Roca, M.E., Otero, M.L. & Haelterman, R.M. (2021) Identificación de la secuencia tipo ST78 de *Xylella fastidosa* en olivo de Argentina [poster]. *5th Congreso Argentino de Fitopatologia*. Available at: https://repositorio.inta.gob.ar/xmlui/handle/20.500.12123/10519 [Accessed 28th February 2023].

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A. et al. (2020) Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, 21, 180.

Vanhove, M., Retchless, A.C., Sicard, A., Rieux, A., Coletta-Filho, H.D., De La Fuente, L. et al. (2019) Genomic diversity and recombination among *Xylella fastidiosa* subspecies. *Applied and Environmental Microbiology*, 85, e2972–e2990.

Viana, D., Comos, M., McAdam, P.R., Ward, M.J., Selva, L., Guinane, C.M. et al. (2015) A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nature Genetics*, 47, 361–366.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Donegan, M.A., Coletta-Filho, H.D. & Almeida, R.P.P. (2023) Parallel host shifts in a bacterial plant pathogen suggest independent genetic solutions. *Molecular Plant Pathology*, 24, 527–535. Available from: https://doi.org/10.1111/mpp.13316