environmental microbiology

Environmental Microbiology (2020) **22**(7), 2625–2638 doi:10.1111/1462-2920.14965

Population structure and adaptation of a bacterial pathogen in California grapevines

Mathieu Vanhove, Anne Sicard, Jeffery Ezennia, Nina Leviten and Rodrigo P.P. Almeida *

Department of Environmental Science, Policy and Management, University of California-Berkeley, Berkeley, CA, 94720.

Summary

***Xylella fastidiosa* subsp. *fastidiosa* causes Pierce's disease of grapevine (PD) and has been present in California for over a century. A singly introduced genotype spread across the state causing large outbreaks and damaging the grapevine industry. This study presents 122 *X. fastidiosa* subsp. *fastidiosa* genomes from symptomatic grapevines, and explores pathogen genetic diversity associated with PD in California. A total of 5218 single-nucleotide polymorphisms (SNPs) were found in the dataset. Strong population genetic structure was found; isolates split into five genetic clusters divided into two lineages. The core/soft-core genome constituted 41.2% of the total genome, emphasizing the high genetic variability of *X. fastidiosa* genomes. An ecological niche model was performed to estimate the environmental niche of the pathogen within California and to identify key climatic factors involved in dispersal. A landscape genomic approach was undertaken aiming to link local adaptation to climatic factors. A total of 18 non-synonymous polymorphisms found to be under selective pressures were correlated with at least one environmental variable highlighting the role of temperature, precipitation and elevation on *X. fastidiosa* adaptation to grapevines in California. Finally, the contribution to virulence of three of the genes under positive selective pressure and of one recombinant gene was studied by reverse genetics.**

Received 23 September, 2019; revised 3 January, 2020; accepted 26 February, 2020. *For correspondence. E-mail rodrigoalmeida@berkeley.edu; Tel. (510) 642-1603; Fax (510) 643-5438.

Introduction

In agricultural ecosystems, bacterial plant pathogens offer largely untested models to measure a phenotypic trait such as virulence and map associated genomic loci. A long standing paradigm in crop–pathogen interactions is that hosts and pathogens are engaged in gene-for-gene co-evolutionary dynamics (Keen, 1990). However, with the development of high-throughput genomics, multiple studies reported that these interactions might also be influenced by abiotic factors acting on genetic loci (Croll and McDonald, 2017). While the study of microbial biogeography has expanded in the recent years, the local adaptation of plant pathogens has not been expansively examined, despite the fact that these are biologically amenable systems to study (Kraemer and Boynton, 2017). Furthermore, pathogens in agroecosystems can have devastating effects on crop yields and epidemics remain a major concern (Stukenbrock and McDonald, 2008; Fisher *et al*., 2012). Understanding pathogen evolution and the origin of pathogenicity and virulence remain central to mitigate impacts and risks of plant pathogens.

In the last two decades, landscape genetics has emerged as a discipline aimed at linking population genetics, spatial statistics and landscape ecology in order to quantify the effects of landscape features on gene flow and adaptation (Manel *et al*., 2003; Manel and Holderegger, 2013). The heterogeneous space or 'landscape' affects microevolutionary dynamics at various scales, leaving genomic signatures that may be identified (Biek and Real, 2010). To date, in this emerging field, Dudaniec and Tesson (2016) noted that little attention has been paid to the linkage between microorganism dispersal and environmental factors, despite evidence of non-random distributions and patterns of isolation by distance (IBD) within populations (Martiny *et al*., 2006). Because of their high dispersal abilities (Taylor *et al*., 2006), microbes were long thought to display little genetic biogeographic differentiation. However, numerous studies support the idea that microbial species indeed exhibit biogeographic patterns (Martiny *et al*., 2006).

Local adaptation occurs when different biotic and/or abiotic selective pressures lead to higher fitness in a

focal population compared to other (Giraud *et al*., 2017). Different considerations must be taken into account when investigating spatially structured microbial populations compared to macro-organisms, such as large population sizes, fast generation time, colonization bottlenecks and seasonal variations (Prosser *et al*., 2007; Hanson *et al*., 2012; Hahn *et al*., 2015). Recombination also influences local adaptation by introducing foreign genes and increasing genetic variance, either allowing for adaptation or disturbing locally adapted gene combinations (Bürger, 1999).

Xylella fastidiosa subsp. *fastidiosa* in California represents an opportunity to study local adaptation of a plant pathogen in an agroecosystem. This pathogen is responsible for Pierce's disease of grapevine (PD), a devastating disease first described in 1892 in California (Pierce, 1892). Multiple epidemics have been reported across the state over the past century. The disease is caused by a blockage of xylem vessels, probably due both to bacterial populations and/or secretions and plant defence responses (Sicard *et al*., 2018). The following reduction in xylem sap flow leads to leaf scorch and stunted growth and can result in vine death. The only natural means of pathogen spread is via xylem sap-feeding insect vectors (Sicard *et al*., 2018). *X. fastidiosa* has been classified into four subspecies, namely, *fastidiosa*, *multiplex*, *pauca* and *sandyi*. A fifth subspecies isolated in mulberry, subsp. *morus*, is thought to be the results of inter-subspecific homologous recombination between subsp. *fastidiosa* and *multiplex* (Nunney *et al*., 2014; Vanhove *et al*., 2019). *X. fastidiosa* populations have been historically isolated due to geographical and host barriers, but the emergence of the pathogen in Europe in 2013 (subsp. *pauca*) is an example of the potential impacts associated with human-mediated invasions (Sicard *et al*., 2018). The clade of subsp. *fastidiosa* causing PD is not native to California; available data suggest a single introduction from Central America (Nunney *et al*., 2010). This singly introduced genotype then spread in Californian vineyards and is now found across the major grape-growing regions of the state (Tumber *et al*., 2014). With favourable climatic conditions plant pathogens are expected to extend their range (Garrett *et al*., 2006), and the impact of climate on *X. fastidiosa* has been extensively reported (Bosso *et al*., 2016a). Low winter temperatures are known to be the primary limitation of the geographical range of *X. fastidiosa* causing PD (Purcell, 1974), suggesting that the disease may increase its distribution due to climate change. As such, this disease system offers a broad set of spatial scales to study abiotic factors that affect plant pathogen distribution.

In the present study, five geographic locations were sampled across California resulting in the sequencing of 122 subsp. *fastidiosa* genomes obtained from symptomatic grapevines. The genetic diversity and population structure of the plant pathogen were quantified and the genomic basis of adaptation to abiotic factors at the scale of California was investigated by using two different approaches. The first approach consisted of detecting outlier loci that deviate from genome-wide patterns of diversity (Vitti *et al*., 2013). This approach uses large numbers of single nucleotide polymorphisms (SNPs) and detects markers that exhibit higher level of genetic differentiation than expected under neutrality (Holderegger and Wagner, 2008). The second approach, known as ecological association, detects significant statistical associations between potential genetic markers and environmental variables (Mita *et al*., 2013; Manel *et al*., 2016). The list of polymorphisms uncovered from these environmental associations and potentially involved in local adaptation can then be compared to genomic regions under selection to provide additional supports of environmental adaptation and reduce false positives (Branco *et al*., 2017).

We first modelled the ecological niche of the PD-causing bacterium in California, and used whole genome sequence data to explore the genetic structure of this population. Then, we examined patterns of selection by investigating genomic signatures of positive selection and correlate them with altitude, temperature and precipitation variables. While comparative genomic studies may reveal genes that contribute to pathogen virulence (Griswold, 2008), reverse genetics enable to experimentally test whether these genes are indeed involved in its pathogenesis. We also selected three genes under positive selective pressure and one recombinant gene and tested their effect on *X. fastidiosa* virulence on grapes.

Results

Population subdivision, variant detection and spatial structure

An average of 1 908 446 reads per isolate was obtained and the mapping of reads to the reference *X. fastidiosa* subsp. *fastidiosa* Temecula1 (ASM724v1) averaged 98.05% (Supporting Information Table S1), with a depth of coverage of 132.30 $\pm$ 60.2 SD. Genetic comparisons were possible due to a conservative variant-calling strategy resulting in a set of high confidence SNPs (see Materials and Methods). A total of 5218 SNPs were identified among the 122 isolates sequenced in California (Table 1). A Bayesian Analysis of Population Structure (BAPS) revealed the presence of two lineages and five genetic clusters: cluster 1 in Santa Barbara, cluster 2 in Temecula, cluster 3 in Bakersfield, cluster 4 in Sonoma and cluster 5 in Napa (Fig. 1, Supporting Information

Table 1. Population genetic statistics for clusters of Pierce's disease causing *Xylella fastidiosa* strains causing disease in California grapevines.

Cluster	SNP[a]	η	π	θ	Tajima's *D*
Santa Barbara (cluster 1, *n* = 7)	727	297	5.97×10^{-5}	6.64×10^{-5}	−0,549
Temecula (cluster 2, *n* = 16)	2316	219	3.38×10^{-5}	4.83×10^{-5}	−1.170
Bakersfield (cluster 3, *n* = 26)	1362	184	6.15×10^{-5}	5.07×10^{-5}	0.842
Sonoma (cluster 4, *n* = 28)	2875	211	5.76×10^{-5}	5.18×10^{-5}	0.450
Napa (cluster 5, *n* = 43)	2479	135	3.36×10^{-5}	4.35×10^{-5}	−0.844
Total (*n* = 120)	5218	5240	0.275	0.187	1.580

Note: A total of 5218 SNPs were identified. In the core genome Santa Barbara isolates (cluster 1) displayed the highest number of mutations (η = 297), where Napa only harboured 135 mutations. Each cluster displayed similar nucleotide diversity (π) and population mutation rates (θ), but Tajima's *D* values were negative for clusters 1, 2 and 5, indicative of a recent population expansion.

a. Single nucleotide polymorphism mapped to the *X. fastidiosa* subsp. *fastidiosa* Temecula1 reference.

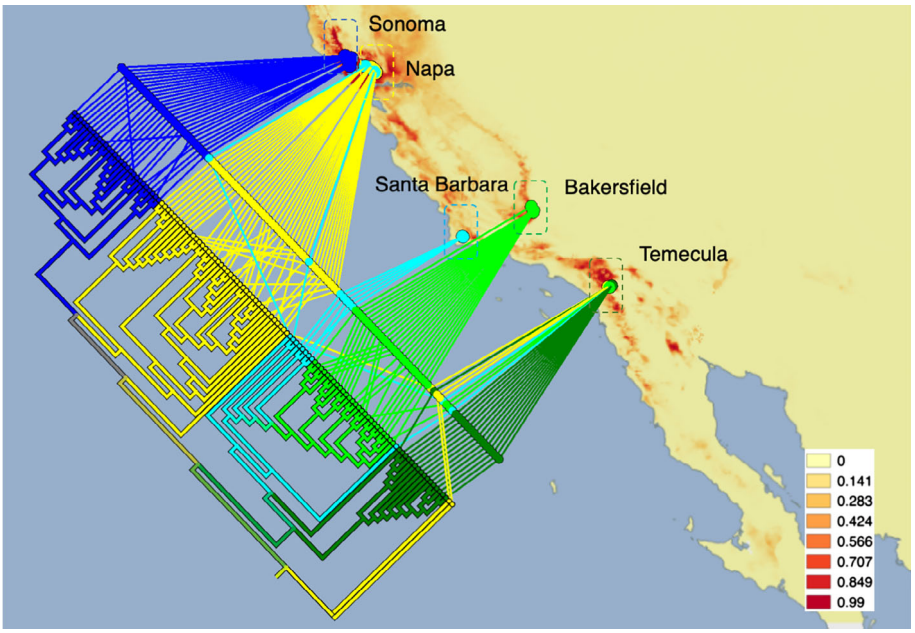

Fig. 1. Distribution of and phylogenetic placement of the Pierce's disease-causing *Xylella fastidiosa* subsp. *fastidiosa* populations within California. The predicted niche of the population was estimated using MAXENT; area values closer to 1 (red) indicate higher likelihood of pathogen occurrence. [Color figure can be viewed at wileyonlinelibrary.com]

Table S1), harbouring different amount of genetic diversity ranging from 727 to 2875 SNPs (Table 1). Each cluster was roughly associated with its geographic origin, but a few outliers were present in each genetic cluster indicative of exchange among these subpopulations: cluster 1 (57.1% of isolates were isolated in Santa Barbara, 3 isolates were outliers), cluster 2 (Temecula, 94.7%, 1 outlier), cluster 3 (Bakersfield, 84.6%, 4 outliers), cluster 4 (Sonoma, 92.6%, 2 outliers) and cluster 5 (Napa, 88.3%, 5 outliers). Phylogenetic analysis revealed strongly supported clades. However, gene flow has occurred among subpopulations as portrayed by Wright's F_{ST} (Table 2), with values ranging from 0.108 to 0.218. Signs of isolation by distance were also identified (Mantel *r* test = 0.392, $P \leq 0.001$), indicative of genetic disparities over the landscape. The realized niche of subsp. *fastidiosa* infecting grapevines across California was predicted using *MaxEnt*, and selected environmental variables, precipitation in the coldest quarter (bio19; 46.0%) and altitude (41.8%) contributed the most to the model (Fig. 1).

Pan-genome analysis

To investigate the pan-genome of this population, we performed *de novo* assemblies on the entire dataset (Supporting Information Table S2). Isolates had an average of 2 517 953 bp, and a N50 and L50 of 51 903 bp and 20.97 respectively. The genetic diversity in the five different clusters ranged from 6.15×10^{-5} to 3.36×10^{-5} (Table 1). Due to quality issues, two isolates were removed (Je9 and Je17) from the dataset. Analysis of the core and accessory genomes revealed the presence of 4583 genes, with 1073 (23.4% in ≥99% of isolates) core genes and 816 (17.8% in 99%–95% of isolates) soft-core genes, 756 shell genes shared by 15%– 95% of the population (16.5%) and 1938 cloud genes shared by less than 15% of the population (42.5%; Fig. 2C). Presence of

Table 2. F_{ST} statistics for *X. fastidiosa* genetic clusters from grapevines in California.

F_{ST}	Cluster 1 (n = 7)	Cluster 2 (n = 16)	Cluster 3 (n = 26)	Cluster 4 (n = 28)	Cluster 5 (n = 43)
Santa Barbara (cluster 1, n = 7)	0.000	0.108	0.146	0.218	0.137
Temecula (cluster 2, n = 16)	0.108	0.000	0.179	0.132	0.111
Bakersfield (cluster 3, n = 26)	0.146	0.179	0.000	0.197	0.212
Sonoma (cluster 4, n = 28)	0.218	0.132	0.197	0.000	0.155
Napa (cluster 5, n = 43)	0.137	0.111	0.212	0.155	0.000

homologous recombination was investigated using *ClonalFrameML* and fastGEAR on the core genome alignment (725 750 bp). The relative effect of recombination to mutation was: r/m = 6.797 (i.e. recombination generated more substitutions than mutation), the relative rate of recombination to mutation was R/θ = 0.524, and the average length of imports equal to δ = 406 bp. *ClonalFrameML* and fastGEAR analyses identified 98 and 47 recombining segments respectively (Supporting Information Fig. S6). Recombining elements were mapped to the reference genome; 64 and 28 of the recombining segments found respectively by *ClonalFrameML* and fastGEAR were identified and mapped to known genes (Supporting Information Tables S9 and S10). Ten of these genes were found using both methods, including an ABC transporter (*cvaB*), a cardiolipin synthase (*cls*), a transcriptional regulator (*attO*) and a cation acetate symporter (*ppa_1*).

Analysis of depth coverage variation has the potential to reveal duplications. A total of 30 genes were found to have an average coverage ≥2 (Supporting Information Table S12). These genes encoded mostly for hypothetical proteins (n = 16) or phage-related proteins (n = 12). Interestingly, each cluster seemed to have a gene duplication for PD_0789, a resolvase/integrase-like protein (GO:0003677; GO:0000150; GO:0006310; Supporting Information Table S12). Additionally, Clusters 1 and 3, which are part of the same lineage, had a higher average coverage for PD_1184 (mean = 1.977 and 1.874 respectively), a toxin-like protein. We interpreted these as loci duplications conserved in these populations.

To investigate the temporal evolution of the grapevine genotype, an ML phylogeny was constructed using the core genome of the 120 isolates with an additional 24 previously published subsp. *fastidiosa* genomes. Published isolates include the following regions: eastern United States (n = 3) and Mexico (n = 2) with known isolation times for a time total time period dating from 1987 to 2015 (28 years, Supporting Information Fig. S7, Table S2). Tip-dating inference using BEAST led to the inference of a substitution rate of 6.37724×10^{-7} per site per year (95 Confidence Interval (CI): 3.9277×10^{-7}, 9.0912×10^{-7}). The evolutionary rate was then extrapolated to the whole subspecies using BEAST (Fig. 2A). The split between the Mexican outgroup and the rest of the USA isolates was estimated at 1269 CE (CI: 850 CE–1668 CE). A divergence between eastern and western isolates was estimated at 1827 CE (1676 CE–1946 CE). Based on these estimations, the time to most common recent ancestor (TMRCA) of subsp. *fastidiosa* in California dates to 1960 CE (1851 CE–1976 CE).

Selection

In order to investigate signs of natural selection, two gene-based methods were used: d_N/d_S (ω) and the McDonald–Kreitman (MK) test. In addition, one univariate outlier test, X_TX, based on high confidence SNPs, was also used. The core genome was generated based on *de novo* assemblies and the d_N/d_S ratio was estimated using *codeml*. The SNP outlier X_TX method identified 190 SNPs that were mapped to the reference genome leading to the identification of 60 genes and a total of 64 non-synonymous mutations (Supporting Information Table S8). Some of the gene products under selection encoded for proteins involved in pathogenesis (GO:0009405) such as a hemolysin-type calcium binding domain (cluster 3, 4 and 5), DNA recombination (GO:0006310) and proteolysis (GO:0006508; Supporting Information Table S11). Additionally, selection was investigated within each genetic cluster to assess whether the different clusters were under different selective pressures. The results of that analysis are summarized in the Supporting Information.

Selection, recombination and virulence

Three genes encoding for hypothetical proteins and displaying high values of d_N/d_S in at least one of the five clusters (PD_0516 > 2.8 in cluster 4 and 5; PD_2073 > 2.6 in cluster 4; PD_0616 > 1 in cluster 3) were knocked out to determine their effect on *X. fastidiosa* virulence in grapes. One recombinant gene with unknown function (PD_0579, Supporting Information Table S9) was also selected. The PD_2073 mutant did not survive on selective growth media pointing towards an essential physiological role. The other three knockout

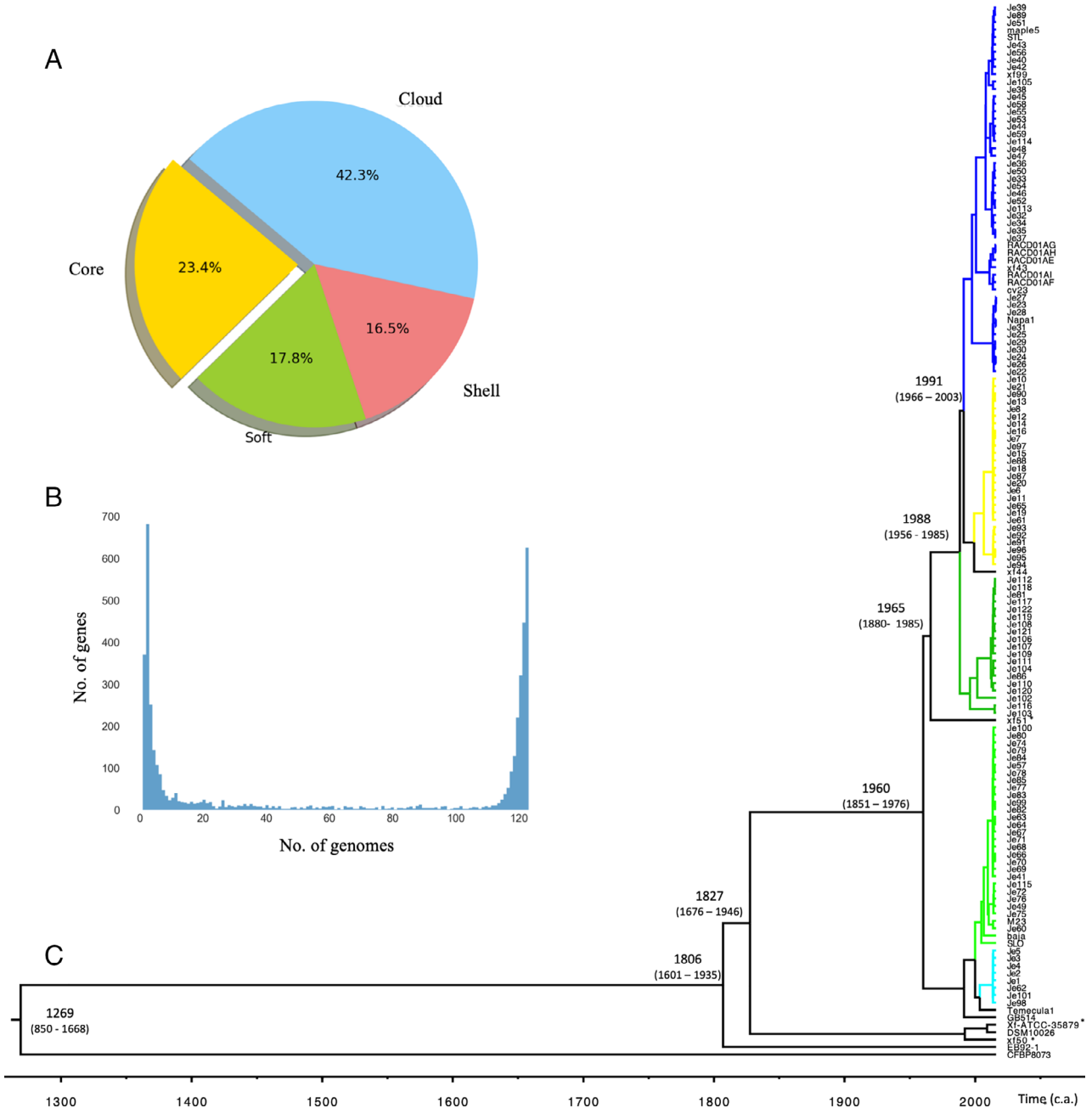

Fig. 2. Characterization of *Xylella fastidiosa* subsp. *fastidiosa* in California.A. Clustering of subsp. *fastidiosa* genetic clusters; a core genome phylogeny of the 120 isolates with tips coloured by region of isolation, Santa Barbara (cyan), Bakersfield (dark green), Temecula (green), Napa (yellow), Sonoma (blue), other isolates are left in black. Asterisks highlight isolates from outside of California, with the outgroup CFBP8073 being from Mexico.B. Chart showing proportion of shared genes among the genomes.C. Representation of the gene in the core, soft-core and accessory genome for the 120 Californian isolates. [Color figure can be viewed at wileyonlinelibrary.com]

mutants (PD_0516, PD_0579 and PD_0616 mutants), and the virulent wild-type (WT) PD strain STL, were mechanically inoculated into susceptible grapevines. No difference in disease severity was observed among the different strains. A more detailed summary of the results is available as Supporting Information.

Association analysis of climatic variables

The association analyses between SNPs and environmental variables were performed using the Bayes' factors and non-parametric Spearman's Rho methods implemented in Bayenv2, and latent factor mixed model

(LFMM). Thirty genes, which were found to be responding to positive selective pressure, were significantly correlated with a climatic variable using two Environmental Association Analysis (EAA) methods. A total of 59 SNPs were significantly associated with at least one environmental variable, including 18 non-synonymous (NSY) mutations (Table 3, Fig. 3, Supporting Information Fig. S2). Among the 18 correlated NSY loci, four were associated with altitude, three with annual mean temperature (bio1), four with mean temperature in the warmest quarter (bio10) and three with mean temperature in the wettest quarter (bio8). Precipitation variables were also correlated with 81 SNPs including four NSY mutations that were found to be associated with precipitation in the wettest month (bio13). Only one NSY mutation was detected with all three methods (snp_977382; Fig. 3F, Table 3 and Fig. S2): a mutation on the PD_0789 gene encoding for a recombinase protein. The C to T transition on the first codon led to a change from leucine to phenylalanine and was significantly correlated with mean temperature in the warmest quarter (bio10). Finally, one SNP (snp_1295815, Fig. 3D) on gene PD_1095 (hypothetical protein) was associated with four different variables. Lower altitude and temperature in the northern part of California seemed to have an effect on these isolates.

Among Sonoma isolates, which experience lower annual mean temperature compared to the rest of the dataset ($z = 4.115$; $P \leq 0.001$, calculated in a Wilcoxon rank sum test), a mutation inducing a glutamine to histidine change was observed in the PD_1517, a gene encoding for an arginine deaminase (Fig. 3B, snp_629619). Similar associations were observed for the PD_0515 and PD_1095 genes (Fig. 3A and Supporting Information Fig. S2). Higher temperature in Southern California compared to Napa and Sonoma regions ($z = 2.872$; $P \leq 0.005$) also led to significant association between snp_161654 on PD_0127 and mean temperature in the wettest quarter (bio13). Lower elevation levels in Napa and Sonoma counties ($z = 3.240$; $P \leq 0.001$) were associated with SNPs in the PD_0515, PD_0764, PD_1095 and PD_1243 genes. Precipitation in the warmer quarters (bio18), and in the wettest month (bio13), which are more abundant in Northern California ($z = 5.671$; $P \leq 0.001$), were correlated with a NSY mutation on gene PD_0790 encoding for a DNA primase (snp_978167; Fig. 3E) and PD_0620, a glycine decarboxylase (snp_765080; Fig. 3C). One NSY mutation change in gene PD_0744 (encoding for a surface protein), was present in the genetic cluster 2 (Temecula at 93.75%), and was associated with elevation and found to be under positive selection (MK test = 0.788, Supporting Information Table S7). The average elevation for this cluster was higher than for other genetic clusters.

Discussion

This study used data on the genetic diversity of a population of *X. fastidiosa* subsp. *fastidiosa* causing disease in grapevines at a regional scale (i.e. state of California, USA). The sampling design provided a broad set of spatial scales to study the influence of physical (abiotic) factors on the distribution of this plant–pathogen. Selection pressures are acting on this population derived from a single introduction event, and that lineage-specific selection is at play and can be identified within the different genetic clusters. A total of 5218 SNPs were uncovered in the dataset, which represents as much genetic diversity as that of the four ST53 genomes (4076 SNPs) described in Giampetruzzi *et al*. (2017), which are associated with an epidemic in olive trees in Italy. This finding highlights the clonal character of subsp. *fastidiosa* in California grapevines. However, signs of population structure were detected in the form of two lineages, one formed by strains from Bakersfield and Santa Barbara, the other formed by Temecula and northern California strains. Population subdivisions were more marked between isolates from Bakersfield and the northern regions, with F_{ST} values reaching 0.197 and 0.212 for Sonoma and Napa counties respectively. Gene flow remained limited between the northern clusters ($F_{ST} = 0.155$), probably due to the Mayacamas Mountains acting as a physical barrier between the two counties, limiting insect vector dispersal. It is not clear what ecological process has led to the geographic structuring of pathogen populations observed here.

The tip dating approach displayed significant temporal signal, allowing for evolutionary rate estimation (6.3772×10^{-7} mutation per site per year; CI: 3.9277×10^{-7}, 9.0912×10^{-7}). This value approaches the rate previously estimated for a subsp. *pauca* clade (7.6204×10^{-7} mutation per site per year; Vanhove *et al*., 2019). We estimated the introduction date of subsp. *fastidiosa* in the USA to be between 850 CE and 1668 CE, and in California to be between 1851 CE and 1976 CE. This study confirms that PD-causing subsp. *fastidiosa* is not native to the United States, as hypothesized by Nunney *et al*. (2010). Hewitt (1958) argued that the pathogen originated from the Gulf Coastal Plain of the USA based on the resistance of wild grapevines in that region, a hypothesis not supported by our results. Each main population was estimated to have a most recent common ancestor in the mid-1900s, although it is known that PD occurred in these regions in the early 1900s (Hewitt, 1958). One explanation, assuming the dating estimates are correct, is that genetic sweeps occurred when the area dedicated to grapevines in California doubled in the 1970s (Geisseler and Horwath, 2016), which could be associated with landscape

Table 3. Non-synonymous mutations in the Pierce's disease *X. fastidiosa* Californian population associated with climate variables.

SNP	Selection Method	Gene	Function	Landscape Genomics Methods	Bioclim	Variable	GO term	GO Names
161 654	XtX; ω_{C3}	PD_0127	DUF2326 domain-containing	BF & ρ*	bio8	Temperature		
463 139	X_TX	PD_0378 (*gp4*)	phage-related portal protein	BF & LFMM	bio13	Precipitation	GO:0019068	virion assembly
598 600	X_TX	PD_0501 (yadG)	ABC transporter ATP-binding protein	BF & ρ*	bio10	Temperature	GO:0016887	ATPase activity
627 294	X_TX; $\omega_{C4;}$ ω_{C5}	PD_0515	Unknown	BF & LFMM	bio8	Temperature	GO:0016021	integral component of membrane
627 395	X_TX; ω_{C4}; ω_{C5}	PD_0515	Unknown	ρ* & LFMM	bio1	Temperature	GO:0016021	integral component of membrane
627 509	X_TX; ω_{C4}; ω_{C5}	PD_0515	Unknown	BF & LFMM	Altitude	Altitude	GO:0016021	integral component of membrane
628 191	ω_{C4}; ω_{C5}	PD_0516	Unknown	BF & LFMM	bio1	Temperature	GO:0016021	integral component of membrane
629 619	X_TX	PD_0517	Arginine deaminase	BF & LFMM	bio1	Temperature	GO:0016021	integral component of membrane
765 080	X_TX	PD_0620 (gcvP)	Glycine decarboxylase	BF & ρ*	bio13	Precipitation	GO:0006546	glycine catabolic process
946 774	X_TX	PD_0764 (int)	Phage-related integrase	ρ* & LFMM	Altitude	Altitude	GO:0015074	DNA integration
977 362	X_TX; ω_{C1}; ω_{C2}	PD_0789	Recombinase family	BF & LFMM	bio10	Temperature	GO:0003677; GO:0000150; GO:0006310	DNA binding; recombinase activity; DNA recombination
977 382	X_TX; ω_{C1}; ω_{C2}	PD_0789	Recombinase family	BF & LFMM & ρ*	bio10	Temperature	GO:0003677; GO:0000150; GO:0006310	DNA binding; recombinase activity; DNA recombination
978 167	X_TX	PD_0790 (traC)	DNA primase	ρ* & LFMM	bio18	Precipitation		
1 068 152	X_TX	PD_0858	DUF4440 domain-containing	BF & ρ*	bio13	Precipitation	GO:0005516; GO:0004683; GO:0006468	calmodulin binding; calmodulin-dependent protein kinase activity; protein phosphorylation
1 295 506	X_TX	PD_1095	Unknown	BF & ρ*	Altitude	Altitude		
1 295 815	X_TX	PD_1095	Unknown	ρ* & LFMM	bio9; bio10; bio13; bio14; bio19	Temperature		
1 446 174	X_TX	PD_1243	DUF596 domain-containing	LFMM & ρ*	bio8	Temperature		
1 446 184	X_TX	PD_1243	DUF596 domain-containing	LFMM & ρ*	Altitude; bio6	Altitude; Temperature		

Each SNP listed was found to be under positive selection with least one method (MK test, d_N/d_S or X_TX) and was found to be correlated with one environmental variable by two of the three EEA methods: Bayes' factors and non-parametric Spearman's Rho methods implemented in Bayenv2 or latent factor mixed model (LFMM). ρ*, non-parametric Spearman's Rho (ρ) distribution.

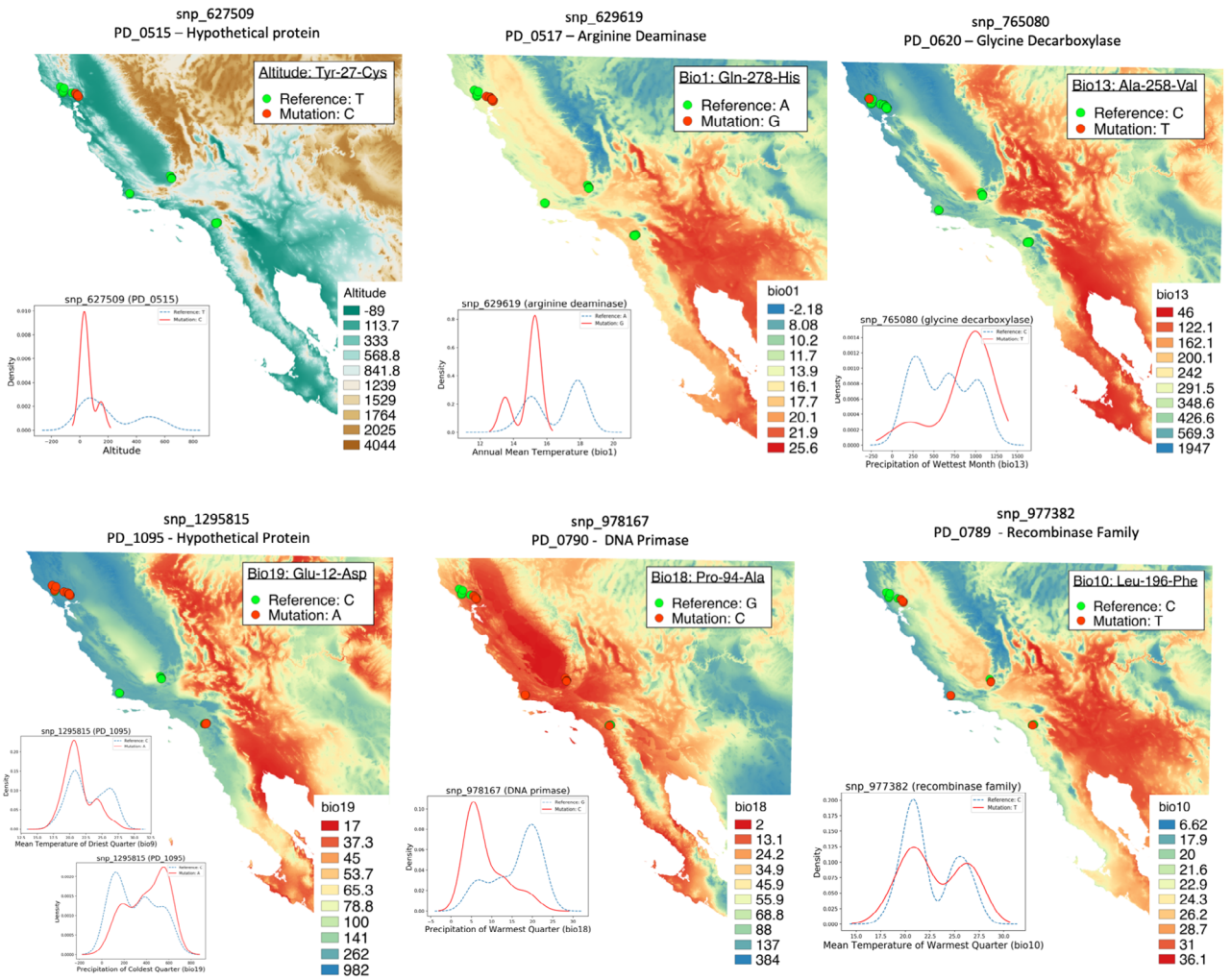

Fig. 3. Six SNPs displaying the highest differences between the population carrying the reference nucleotide and the population with the mutation are shown. Each map is composed of the environmental variable associated with the correlated SNP. The change in aminoacid is indicated and density plots portray the values of the environmental variable for the two populations of isolates (wild-type and mutant). The full maps of the 18 non-synonymous mutations associated with a climatic variable can be found on Supporting Information Fig. S2. [Color figure can be viewed at wileyonlinelibrary.com]

changes and modifications to farming practices, or transportation and establishment of novel genotypes in various regions (Sicard *et al*., 2018). Regardless, while the inferred dates and respective CI are ecologically reasonable, considering the history of PD in the USA, additional sampling efforts from other locations are expected to provide more robust data on the origin of PD in the USA.

The core/soft-core genome constituted 41.2% of the 4583 genes in the population, similar to what has been observed for other bacteria and *X. fastidiosa* (Lapierre and Gogarten, 2009; Mira *et al*., 2010). The large accessory genome, 1938 cloud genes shared by <15% of the population (42.3%), appeared similar to other study. In a study of 205 multidrug-resistant (MDR) *Serratia marcescens* in the United Kingdom and Ireland, Moradigaravand *et al*. (2016) found an accessory genome value of 61.3%. Analysis of depth of coverage variation revealed duplications that could contribute to pathogenesis. In the first lineage, 10 genes had double coverage including PD_1184, a toxin protein and PD_0789 (an integrase involved in DNA recombination). On the other hand, depth of coverage variation was more pronounced in the second lineage, with over 35 genes found in each of the three genetic clusters. These are expected to be conserved duplications, and are potentially important for host (plant or insect) colonization.

Various selective forces are acting on the population of subsp. *fastidiosa* analysed. Tajima's *D* values were negative in genetic clusters 1, 2 and 5. This metric searches for genomic regions undergoing a selective sweep (Tajima, 1989). Negative Tajima's *D* values are indicative of a surplus of rare alleles, signatures of positive selection or recent population expansion (Charlesworth, 2006). ClonalFrameML and fastGEAR analyses revealed the

exchange of genetic material occurring among the genetic clusters. Recombination is recognized as a main driver of genetic diversity in this species (Nunney *et al*., 2014). In this population, the relative effect of recombination and mutation to substitution accumulation was 6.797. Previous work found that recombination contributed twice as much (Vanhove *et al*., 2019), indicating that recombination within subsp. *fastidiosa* in California is more frequent than in other populations of *X. fastidiosa* studied. It is possible that allele exchange has limited fitness consequences, given the population is host-limited and relatively young. Additionally, the singly-introduced genotype affecting grapevines in California has recombined with endemic subsp. *multiplex*, potentially facilitating subsp. *fastidiosa* adaptation to grapevines and environmental conditions (Vanhove *et al*., 2019).

Recombining genes that persist within populations may confer benefits to the strains that harbour them. We here tested whether a recombining gene, PD_0579, had an effect on *X. fastidiosa* virulence. The deletion of this gene did not have any effect on multiplication or movement of the bacterium within grapevines or on disease symptom development, pointing towards its lack of effect on *X. fastidiosa* virulence in this plant species. Similarly, the three genes under positive selection tested biologically had no effect on disease symptom development or *in planta* movement. These strains might have already been selected for their fast multiplication and movement in grapes and as a consequence, may already be well adapted to this host. These genes may be associated with insect colonization, for example, or abiotic stresses among other possibilities. A NSY mutation in one of the candidate genes, PD_0516, was significantly associated with the annual mean temperature, suggesting that genes involved in local adaptation might not necessarily be associated with pathogenicity. PD_2073 encodes a hypothetical protein within an operon including six other genes with homology to the type I restriction and modification system (ProOpDB; Taboada *et al*., 2011), its mutant did not grow in vitro. This gene may also be part of this large, multi-functional enzyme complex.

Selective pressures imposed by biotic and abiotic factors may provide higher fitness to local populations in relation to those from other regions. Comparing the rate of synonymous and non-synonymous mutations has been widely used in microbial genomics (Giraud *et al*., 2017). Other studies have previously identified genes in bacterial plant pathogens using this approach. For instance, Richard *et al*. (Richard *et al*., 2017) reported selection in genes involved in resistance to copper-based insecticide in *Xanthomonas citri* pv. *citri*. In the present study, the reverse-ecology approaches identified signatures of adaptation in genes involved in several functions including response to antibiotics (*mrcB*), pathogenesis (*btaE* and *upaG)*, heme transport (*ccmC*) and cell adhesion (*pilA_1*). These findings highlight the adaptive potential of *X. fastidiosa*, a concern for regions outside its historical range in the Americas associated with recent disease outbreaks (Sicard *et al*., 2018).

We used an ecological niche modelling approach to understand the abiotic niche requirements of subsp. *fastidiosa* in California grapevines. The predictive environmental model described where PD has been reported over past years (Tumber *et al*., 2014). These analyses confirmed the influence of climate on the chronic establishment of the bacterium, as previously suggested (Purcell, 1997). This approach has been previously used to predict the *X. fastidiosa* ecological niche in Italy and across the Mediterranean basin (Bosso *et al*., 2016a). Minimum temperature in the coldest month (Bio6) and altitude were identified as important factors in shaping *X. fastidiosa* distribution. A colder winter climate has been previously shown to be a limiting factor for *X. fastidiosa* survivorship in grapevines (Purcell, 1977).

The ecological determinants leading to subsp. *fastidiosa* adaptation to grapevines in California were studied in relation to the spatial genomic structure observed in this dataset. This approach aimed to characterize the association between genetic loci, selection pressures and abiotic factors. Eighteen NSY mutations were detected using EAAs methods. GO term analyses revealed that these genetic changes were primarily associated with genes involved in recombination, glycine catabolism and protein phosphorylation. A SNP in a recombinase protein (PD_0789) was found using all three landscape genomic methods and may play a role in the ability of the pathogen to adapt to its environment. Both an ABC transporter (PD_0501) and an arginine deaminase (PD_0517) had SNPs correlated with a temperature variable. In these cases, few isolates acquired the mutation, which then spread within focal populations. One NSY (snp_1446184) was significantly correlated with altitude and minimum temperature in the coldest month, which appeared as a potential limiting factor for pathogen expansion. PD has been reported in areas where winter temperature reaches 1–4°C (Purcell, 1997) but freezing exposures led to the elimination of the disease (Purcell, 1980). These findings provide evidence that epidemiological dynamics can be directly influenced by abiotic factors over short timescales as suggested by Biek and Real (2010). The role of abiotic factors in agroecosystems might be relevant to pathogen adaptation and future studies should consider such interactions.

The present study revealed the genomic structure of a PD-causing *X. fastidiosa* subsp. *fastidiosa* population in California, and shed light onto the environmental factors associated with the adaptation of this plant pathogen. Evidence of local adaptation was observed and despite

the presence of a robust population structure, the study supports the single-introduction hypothesis (Nunney *et al*., 2010). The relatively young age of this population resulted in the formation of distinct genetic clusters displaying signs of homologous recombination. The study of selection and gene–environment associations revealed the presence of traits associated with climatic variables (Branco *et al*., 2017). Whether local adaptation is favourable or detrimental to pathogenicity remains to be examined.

Experimental procedures

Environmental sampling

We collected samples of European grapevine (*Vitis vinifera*) plants expressing Pierce's disease symptoms in commercial vineyards across California (Supporting Information Table S2). A total of 122 isolates were obtained across 900 km in California from five different counties, Temecula (n = 23), Santa Barbara (n = 5), Bakersfield (n = 27), Napa (n = 41) and Sonoma (n = 28). The age of the vines, when available, ranged from 1994 to 2014; all samples were collected in 2015. Isolation in the laboratory was performed on PD3 solid medium (Davis *et al*., 1981a), followed by triple cloning on PD3. For long-term storage, strains were stored at −80°C in PW broth (Davis *et al*., 1981b) with 30% glycerol. DNA extraction was performed using commercial kit (DNeasy Blood & Tissue Kit; Qiagen) according to instructions by the manufacturer.

Whole-genome sequencing, SNP calling and ploidy analysis

High molecular weight DNA was extracted, and DNA libraries were prepared for Illumina MiSeq paired-end sequencing. DNA from all samples was sent for sequencing at the QB3 Vincent J. Coates Genomics Sequencing Laboratory. Raw reads and other information regarding each isolate have been submitted to the NCBI database (MiSeq project: SUB3867588). Reads were mapped to *X. fastidiosa* subsp. *fastidiosa* Temecula1 (*ASM724v1*; ENA assembly: GCA_000007245.1). Alignments were performed with the Burrows-Wheeler Aligner (BWA) 0.7.15 aln (Li and Durbin, 2009) with a quality threshold of 15 (Rhodes *et al*., 2014). FastQs were converted to SAM format using BWA and converted to BAM files, and the BAM files were then sorted and indexed with SAMTOOLS version 1.3.1 (Li *et al*., 2009). Duplicate reads were marked with PICARD TOOLS (v.2.4.1). The BAM files were processed around insertions or deletions (INDELs) using the GATK RealignerTargetCreator and IndelRealigner (McKenna *et al*., 2010). Single nucleotide polymorphisms (SNPs) and INDELs was identified using GATK UNIFIEDGENOTYPER version 3.6 in haploid mode (DePristo *et al*., 2011; Auwera *et al*., 2013). SNPs and INDELs were filtered to call only high-confidence variants, according to whether they were present in 80% of reads. Resulting variants were mapped to genes using VCF-annotator (Broad Institute, Cambridge, MA) and the latest release of *X. fastidiosa* subsp. *fastidiosa* Temecula1 (*ASM724v1.36*). Mapped reads for each isolate are given on Supporting Information Table S1.

De novo assembly and pan-genome analyses

Data processing was similar as previously done (Vanhove *et al*., 2019). Genomes were assembled using SPAdes 3.6.0 using the *careful* parameter (Bankevich *et al*., 2012); total contig length is summarized in Table S2. progressiveMauve was used to order contigs (Darling *et al*., 2010) and Prokka used for annotation (Seemann, 2014). A pan genome of the 122 isolates was constructed using Roary (Page *et al*., 2015). Recombination events were identified by finding regions of enriched SNP density by using ClonalFrameML (Didelot and Wilson, 2015) and fastGEAR (Mostowy *et al*., 2017). To detect change in ploidy (i.e. duplications), the mean coverage for each isolate was determined using 'DepthOfCoverage' from the GATK pipeline under default setting; the subsp. *fastidiosa* Temecula1 genome was used as a reference and coverage was normalized and averaged over a 500 bp window. Average coverage for each genetic cluster was computed and regions displaying a normalized coverage ≥2 were considered diploid events. The ontology of genes of interests was investigated. GO terms were then assigned using Blast2GO (Conesa and Götz, 2008) using a minimum E value of 1×10^{-10}.

Phylogeny, population assignment and molecular dating

Whole-genome SNP files were converted to Nexus and Phylip formats. A maximum likelihood tree was generated with *RAxML* (1000 bootstrap replicates and a generalized time reversible (GTR) substitution matrix (Stamatakis, 2006), and visualized with FIGTREE v. 1.4 (Rambaut, 2012). Bayesian Analysis of Population Structure (BAPS; Corander *et al*., 2004) was used to assign isolates to genetic clusters. We investigated the presence of a temporal signal in the data set by using our 120 isolates and an additional 24 previously published subsp. *fastidiosa* genomes. The non-recombining core genome of strains with known isolation dates (1987–2015, 27 years of evolution) was obtained and used for this analysis (Supporting Information Fig. S7; https://localtemporalsignal.shinyapps.io/LocalTemporalSignal/). We refer to Vanhove *et al*. (2019) for details on this analysis, as the

procedures used here are the same as performed in that study.

Population genetics and detection of regions under positive selection

Population level statistics were generated for each *X. fastidiosa* subsp. *fastidiosa* genetic cluster. The number of segregating sites (S), total number of mutations (η), nucleotide diversity (π), Waterson's estimator (θ) and Tajima's *D* were estimated using VARISCAN V.2.0 (Vilella *et al*., 2005) on the core genome alignment without the recombination regions (Table 1). Fixation index (F_{ST}) statistics were calculated using BEDASSLE (Bradburd *et al*., 2013). In addition, several statistical methods are available (Vitti *et al*., 2013) to detect signatures of Darwinian selection. Candidate loci under selection were investigated using X_TX, a population differentiation statistic analogous to F_{ST} that accounts for variance–covariance of the population using Bayenv2 (Günther and Coop, 2013). On the other hand, genebase methods compare the rate of synonymous (d_S) and nonsynonymous (d_N) mutations in protein-coding genes (Yang and Bielawski, 2000). The ratio of the rates of synonymous and non-synonymous substitution (d_N/d_S) is commonly used to characterize microbial adaptation (Hurst, 2002). The McDonald–Kreitman (MK) test identifies patterns of selection by comparing the number of silent (d_N) and non-silent substitutions (d_S) with the number of silent (p_S) and non-silent polymorphism (p_N) of an outgroup species (Stoletzki and Eyre-Walker, 2011). A d_N/d_S analysis and a McDonald–Kreitman (MK) test (McDonald and Kreitman, 1991) were performed using annotated high-confidence SNP mapped to the reference strain Temecula1. SNPs with allele frequency <20% were removed and only genes with ≥5 SNPs were considered to improve test performance as recommended by Liti *et al*. (2009)). The *X. fastidiosa* subsp. *multiplex* M12 strain (Chen *et al*., 2010) was used as an outgroup for the MK test. The d_N/d_S (ω) ratio was estimated using *de novo* assemblies after removing recombination events (identified with ClonalFrameML), and *Codeml* from the PAML4.1 package (runmode 0, model 0 was used assuming constant d_N/d_S; Yang, 2007); gene clusters were generated by Roary using gene annotation from Prokka (Seemann, 2014).

Ecological niche modelling, biogeography and environmental association analysis

California provides an appropriate setting to model the distribution of *X. fastidiosa* subsp. *fastidiosa* infecting grapevines. *MaxEnt* is a species habitat modelling software that uses maximum entropy to model the geographic distribution of a species (Phillips and Dud'ik, 2008). The software uses presence-only data and climatic variables. The ecological niche of the *X. fastidiosa* distribution was modelled using WorldClim layers v.2 and altitude, which were obtained from the WORDCLIM database at 30 arc-seconds resolution (Fick and Hijmans, 2017). Each variable was tested for colinearity using a Pearson's $|r| \leq 0.80$ implemented in the R package *ppcor* as described by Bosso *et al*. (2016b). To test model prediction, 25% of the samples were randomly set aside (Supporting Information Fig. S1).

The Mantel test was used to assess the association between genetic and geographic distance among individuals, and to detect spatial autocorrelation (Mantel, 1967). Genetic variation was calculated as the Bray–Curtis distances between loci. The geographic distances were the Euclidean distances between the sampling localities. Mantel tests were performed using the *ecodist* package (Goslee and Urban, 2007) in R using 10 000 permutations. Environmental factors were extracted from WorldClimv.2.0 layers (Fick and Hijmans, 2017). The information for each sample was extracted in R (version 3.1.1) using the *raster* (Hijmans and van Etten, 2012) and *dismo* (Hijmans *et al*., 2012) packages.

Detection of loci correlated with physical variables was performed using Latent Factor Mixed Model (LFMM; Frichot *et al*., 2013) and Bayenv2 (Coop *et al*., 2010). LFMM is a Bayesian approach used to detect selection in landscape genomics. The method investigates the influence of population structure on allele frequencies by introducing unobserved variables as latent factors (Stucki *et al*., 2014). LFMM provides a way to investigate signatures of local adaptation by identification of high degrees of correlation between polymorphism and environmental variables. To detect signatures of selection, a positive false discovery rate of 0.05 was also applied using the *qvalue* package (Dabney *et al*., 2004) in R. BayEnv2 was also used to detect selection using Bayes' factors (BF; BF ≥3 and within top 5%) and non-parametric Spearman's Rho distribution (top 5%). To estimate the covariance matrix, three replicates were performed and averaged using 100,000 Monte Carlo Markov Chain (MCMC). To ensure independence between SNPs (Bayenv2 Manual) when computing the covariance matrix, loci identified using LFMM and loci found using the program LDhat, which identifies patterns of linkage disequilibrium using Hudson's composite likelihood method (McVean *et al*., 2004), were removed. For both methods, LFMM and BayEnv2, three independent runs were performed using 100 000 MCMC cycles and resulting scores were averaged for each of the climate variable. To perform environmental association analysis (EAA), each variable was averaged, standardized and mean-centred across the population as described in the

Bayenv2 manual. Potential SNP candidates were mapped to the Temecula1 reference genome (*ASM724v1*) using a Basic Local Alignment Search Tool (BLAST) service obtained from the Universal Protein Resource (UniProt ID: 183190; UniProt, 2017) and Ensembl (Aken *et al*., 2016).

Biological testing of mutant strains of genes under positive selection or with evidence of recombination

The Materials and Methods section associated with these experiments is described in the Supporting Information.

Acknowledgements

We thank farmers and colleagues Mark Battany, Monica Cooper, Matthew Daugherty, David Haviland, Rhonda Smith, Lucia Varela, from University of California Cooperative Extension for assistance in site selection and sample collection. Funding supporting this research was provided by the California Department of Food and Agriculture PD/GWSS program and Horizon 2020 XF-ACTORS consortium. Anne Sicard is funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 707013. Genome sequencing was performed at the UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory, which is supported by an NIH instrumentation grant (S10 OD018174).

References

Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., *et al*. (2016) Ensembl 2017. *Nucleic Acids Res* **45**: D635–D642.

Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., *et al*. (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**: 11.10.1–11.10.33.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Biek, R., and Real, L.A. (2010) The landscape genetics of infectious disease emergence and spread. *Mol Ecol* **19**: 3515–3531.

Bosso, L., Di Febbraro, M., Cristinzio, G., Zoina, A., and Russo, D. (2016a) Shedding light on the effects of climate change on the potential distribution of *Xylella fastidiosa* in the Mediterranean basin. *Biol Invasions* **18**: 1759–1768.

Bosso, L., Russo, D., Di Febbraro, M., Cristinzio, G., and Zoina, A. (2016b) Potential distribution of *Xylella fastidiosa* in Italy: a maximum entropy model. *Phytopathol Mediterr* **55**: 62–72.

Bradburd, G.S., Ralph, P.L., and Coop, G.M. (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution (N Y)* **67**: 3258–3273.

Branco, S., Bi, K., Liao, H., Gladieux, P., Badouin, H., Ellison, C.E., *et al*. (2017) Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Mol Ecol* **26**: 2063–2076.

Bürger, R. (1999) Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* **153**: 1055–1069.

Charlesworth, D. (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* **2**: e64.

Chen, J., Xie, G., Han, S., Chertkov, O., Sims, D., and Civerolo, E.L. (2010) Whole genome sequences of two Xylella fastidiosa strains (M12 and M23) causing almond leaf scorch disease in California. *J Bacteriol* **192**: 4534.

Conesa, A., and Götz, S. (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* **2008**: 1–12.

Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J.K. (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423.

Corander, J., Waldmann, P., Marttinen, P., and Sillanpää, M. J. (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**: 2363–2369.

Croll, D., and McDonald, B.A. (2017) The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol Ecol* **26**: 2027–2040.

Dabney, A., Storey, J.D., and Warnes, G. (2004) Q-value estimation for false discovery rate control. *Medicine (Baltimore)* **344**: 539–548.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Davis, M.J., French, W.J., and Schaad, N.W. (1981a) Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald. *Curr Microbiol* **6**: 309–314.

Davis, M.J., Whitcomb, R.F., and Gillaspie, A.G., Jr. (1981b) Fastidious bacteria of plant vascular tissue and invertebrates (including so called rickettsia-like bacteria). In *The Prokaryotes*, Balows, A., Trüper, H.G., Dworkin, M., Harder, W., and Schleifer, K.H. (eds). New York, NY: Springer, pp. 2172–2188.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., *et al*. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491–498.

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

Dudaniec, R.Y., and Tesson, S.V.M. (2016) Applying landscape genetics to the microbial world. *Mol Ecol* **25**: 3266–3275.

Fick, S.E., and Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* **37**: 4302–4315.

Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., and Gurr, S.J. (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**: 186–194.

Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**: 1687–1699.

Garrett, K.A., Dendy, S.P., Frank, E.E., Rouse, M.N., and Travers, S.E. (2006) Climate change effects on plant disease: genomes to ecosystems. *Annu Rev Phytopathol* **44**: 489–509.

Geisseler, D. and Horwath, W.R. (2016) Alfalfa production in California. Available at https://apps1.cdfa.ca.gov/FertilizerResearch/docs/Alfalfa_Production_CA.pdf.

Giampetruzzi, A., Saponari, M., Loconsole, G., Boscia, D., Savino, V.N., Almeida, R., *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.

Giraud, T., Koskella, B., and Laine, A. (2017) Introduction: microbial local adaptation: insights from natural populations, genomics and experimental evolution. *Mol Ecol* **26**: 1703–1710.

Goslee, S.C., and Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw* **22**: 1–19.

Griswold, A. (2008) Genetic origins of microbial virulence. *Nat Educ* **1**: 81.

Günther, T., and Coop, G. (2013) Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205–220.

Hahn, M.W., Koll, U., Jezberová, J., and Camacho, A. (2015) Global phylogeography of pelagic Polynucleobacter bacteria: restricted geographic distribution of subgroups, isolation by distance and influence of climate. *Environ Microbiol* **17**: 829–840.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Hewitt, W.B. (1958) The probable home of Pierce's disease virus. *Plant Dis Report* **42**: 241–245.

Hijmans, R.J. and van Etten, J. (2012) Raster: geographic analysis and modeling with raster data. R Packag version 1.9-92.

Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J. (2012) dismo: Species distribution modeling. *R Packag version 07-17*.

Holderegger, R., and Wagner, H.H. (2008) Landscape genetics. *Bioscience* **58**: 199–207.

Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* **18**: 486–487.

Keen, N.T. (1990) Gene-for-gene complementarity in plant-pathogen interactions. *Annu Rev Genet* **24**: 447–463.

Kraemer, S.A., and Boynton, P.J. (2017) Evidence for microbial local adaptation in nature. *Mol Ecol* **26**: 1860–1876.

Lapierre, P., and Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. *Trends Genet* **25**: 107–110.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., *et al*. (2009) Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341.

Manel, S., and Holderegger, R. (2013) Ten years of landscape genetics. *Trends Ecol Evol* **28**: 614–621.

Manel, S., Schwartz, M.K., Luikart, G., and Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* **18**: 189–197.

Manel, S., Perrier, C., Pratlong, M., Abi-Rached, L., Paganini, J., Pontarotti, P., and Aurelle, D. (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol Ecol* **25**: 170–184.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R. K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

McDonald, J.H., and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in drosophila. *Nature* **351**: 652–654.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al*. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303.

McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science (80-)* **304**: 581–584.

Mira, A., Martín-Cuadrado, A.B., D'Auria, G., and Rodríguez-Valera, F. (2010) The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* **13**: 45–57.

Mita, S., Thuillet, A., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., and Vigouroux, Y. (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol* **22**: 1383–1399.

Moradigaravand, D., Boinett, C.J., Martin, V., Peacock, S.J., and Parkhill, J. (2016) Recent independent emergence of multiple multidrug-resistant Serratia marcescens clones within the United Kingdom and Ireland. *Genome Res* **26**: 1101–1109.

Mostowy, R., Croucher, N.J., Andam, C.P., Corander, J., Hanage, W.P., and Marttinen, P. (2017) Efficient inference of recent and ancestral recombination within bacterial populations. *Mol Biol Evol* **34**: 1167–1182.

Nunney, L., Yuan, X., Bromley, R., Hartung, J., Montero-Astúa, M., Moreira, L., *et al*. (2010) Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the US. *PLoS One* **5**: e15488.

Nunney, L., Schuenzel, E.L., Scally, M., Bromley, R.E., and Stouthamer, R. (2014) Large-scale intersubspecific recombination in the plant-pathogenic bacterium Xylella fastidiosa is associated with the host shift to mulberry. *Appl Environ Microbiol* **80**: 3025–3033.

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.
Phillips, S.J., and Dud'ik, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography (Cop)* **31**: 161–175.
Pierce, N.B. (1892) *The California Vine Disease: A Preliminary Report of Investigations*. Washington, DC: US Government Printing Office.
Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al*. (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**: 384–392.
Purcell, A.H. (1974) Spatial patterns of Pierce's disease in the Napa Valley. *Am J Enol Vitic* **25**: 162–167.
Purcell, A.H. (1977) Cold therapy of Pierce's disease of grapevines [Viral diseases, insect vectors]. *Plant Dis Rep* **61**: 514–518.
Purcell, A.H. (1980) Environmental therapy for Pierce's disease of grapevines. *Plant Dis* **64**: 388–390.
Purcell, A.H. (1997) Xylella fastidiosa, a regional problem or global threat? *J Plant Pathol* **79**: 99–105.
Rambaut, A. (2012) *Figtree v1. 4. Molecular Evolution, Phylogenetics and Epidemiology*. Edinburgh, UK: Institute of Evolutionary Biology, University of Edinburgh.
Rhodes, J., Beale, M.A., and Fisher, M.C. (2014) Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. *PLoS One* **9**: e113501.
Richard, D., Ravigné, V., Rieux, A., Facon, B., Boyer, C., Boyer, K., *et al*. (2017) Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements. *Mol Ecol* **26**: 2131–2149.
Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.
Sicard, A., Zeilinger, A.R., Vanhove, M., Schartel, T.E., Beal, D.J., Daugherty, M.P., and Almeida, R.P.P. (2018) Xylella fastidiosa: insights into an emerging plant pathogen. *Annu Rev Phytopathol* **56**: 181–202.
Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
Stoletzki, N., and Eyre-Walker, A. (2011) Estimation of the neutrality index. *Mol Biol Evol* **28**: 63–70.
Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., *et al*. (2014) High performance computation of landscape genomic models integrating local indices of spatial association. *Mol Ecol Resour* **17**: 1072–1089.
Stukenbrock, E.H., and McDonald, B.A. (2008) The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol* **46**: 75–100.
Taboada, B., Ciria, R., Martinez-Guerrero, C.E., and Merino, E. (2011) ProOpDB: pro karyotic Op eron D ata B ase. *Nucleic Acids Res* **40**: D627–D631.
Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
Taylor, J.W., Turner, E., Townsend, J.P., Dettman, J.R., and Jacobson, D. (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom fungi. *Philos Trans R Soc Lond B Biol Sci* **361**: 1947–1963.
Tumber, K., Alston, J., and Fuller, K. (2014) Pierce's disease costs California $104 million per year. *Calif Agric* **68**: 20–29.
UniProt, C. (2017) The universal protein resource (UniProt). *Nucleic Acids Res* **36**: D190–D195.
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. *Appl Environ Microbiol* **85**: e02972-18.
Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005) VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* **21**: 2791–2793.
Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013) Detecting natural selection in genomic data. *Annu Rev Genet* **47**: 97–120.
Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
Yang, Z., and Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**: 496–503.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supporting Information