

Genetic differentiation of *Xylella fastidiosa* following the introduction into Taiwan

Andreina I. Castillo^{1†}, Chi-Wei Tsai^{2†}, Chiou-Chu Su³, Ling-Wei Weng², Yu-Chen Lin⁴, Shu-Ting Cho⁴, Rodrigo P. P. Almeida¹ and Chih-Horng Kuo^{4,*}

Abstract

The economically important plant pathogen *Xylella fastidiosa* has been reported in multiple regions of the globe during the last two decades, threatening a growing list of plants. Particularly, *X. fastidiosa* subspecies *fastidiosa* causes Pierce's disease (PD) of grapevines, which is a problem in the USA, Spain, and Taiwan. In this work, we studied PD-causing subsp. *fastidiosa* populations and compared the genome sequences of 33 isolates found in Central Taiwan with 171 isolates from the USA and two from Spain. Phylogenetic relationships, haplotype networks, and genetic diversity analyses confirmed that subsp. *fastidiosa* was recently introduced into Taiwan from the Southeast USA (i.e. the PD-I lineage). Recent core-genome recombination events were detected among introduced subsp. *fastidiosa* isolates in Taiwan and contributed to the development of genetic diversity. The genetic diversity observed includes contributions through recombination from unknown donors, suggesting that higher genetic diversity exists in the region. Nevertheless, no recombination event was detected between *X. fastidiosa* subsp. *fastidiosa* and the endemic sister species *Xylella taiwanensis*, which is the causative agent of pear leaf scorch disease. In summary, this study improved our understanding of the genetic diversity of an important plant pathogenic bacterium after its invasion to a new region.

DATA SUMMARY

All raw reads have been deposited in NCBI under BioProject PRJNA715299. Complete genome sequence of the reference isolate GV230 has been deposited in GenBank/ENA/DDBJ under the accession CP060159.

INTRODUCTION

Southeast Asia is a region of intricate tectonic and climatic evolution, complex biogeography, and a hotspot for global biodiversity [1–4]. Oceanic islands like Taiwan, are important contributors to plant richness within the region, with local diversity reflecting both palaeogeographic events and human-mediated plant movement [5]. Though there are

many species endemic to Taiwan, approximately 73% of vascular plants found in the island originated from neighbouring regions [6]. In recent decades, a significant number of plants from the Americas, Europe, and Africa have also been introduced. By the start of the new millennium, it is estimated that 341 plant species have been naturalized in Taiwan [7], many of them crops. In association with these novel crop species, numerous pests and pathogens have also been introduced [8, 9].

The economic losses caused by phytopathogen infections are estimated as \$1 billion dollars worldwide every year [10]. In addition to the economic losses, emerging plant pathogens represent a threat to food security, not only by affecting a region's capacity to meet its nutritional needs, but also by

Received 24 September 2021; Accepted 22 October 2021; Published 13 December 2021

Author affiliations: ¹Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA; ²Department of Entomology, National Taiwan University, Taipei 106, Taiwan, ROC; ³Division of Pesticide Application, Taiwan Agricultural Chemicals and Toxic Substances Research Institute, Taichung 413, Taiwan, ROC; ⁴Institute of Plant and Microbial Biology, Academia Sinica, Taipei 115, Taiwan, ROC.

*Correspondence: Chih-Horng Kuo, chk@gate.sinica.edu.tw

Keywords: *Xylella fastidiosa*; genomic diversity; grapevines; Pierce's disease; introduction event; Taiwan.

Abbreviations: ANI, average nucleotide identity; ML, maximum likelihood; MLST, multilocus sequence typing; N, sample size; PD, pierce disease; ST, sequence type; USA, United States of America.

All raw reads have been deposited in NCBI under BioProject PRJNA715299. Complete genome sequence of the reference isolate GV230 has been deposited in GenBank/ENA/DDBJ under the accession CP060159.

†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table and three supplementary figures are available with the online version of this article.

000727 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

negatively impacting local economies and social structures dependent on trade [11–20]. The difficulties associated with the control and management of introduced plant diseases are well documented [21–23]. Specifically, the genetic homogeneity of crop monocultures favours the emergence of host-specialized pathogen genotypes and the increment of virulence [24, 25]. Furthermore, pathogens can quickly adapt to geographical and ecological conditions [26–29]. These problems are particularly concerning in the context of introduced plant pathogens with wide host ranges and demonstrated capacity to spill over to sympatric flora. In this regard, the xylem-limited bacterial plant pathogen *Xylella fastidiosa*, transmitted by various insect vectors, represents an important topic of study [30–33].

The species *X. fastidiosa* is taxonomically divided into three main monophyletic subspecies with allopatric origins: *X. fastidiosa* subsp. *fastidiosa*, native to Central America [34–36]; *X. fastidiosa* subsp. *multiplex*, native to temperate and subtropical North America [37, 38]; and *X. fastidiosa* subsp. *pauca*, thought to be native to South America [37]. Each of these subspecies has experienced a recent geographical expansion mediated by multiple intra- and inter-continental introduction events [30, 36, 39–42]. In the particular case of subsp. *fastidiosa*, that includes one introduction to the USA that led to the emergence of Pierce's disease (PD) [36]. Within the USA, PD-causing isolates diversified into three phylogenetically supported clades. The clade PD-I is composed of isolates collected from the Southeast USA. The clades PD-II and PD-III are both composed of mostly Californian isolates, but each one includes isolates from a different location in the Southeast USA (i.e. Texas for PD-II and Georgia for PD-III) [36]. Subsequently, the introduction of PD-II-related isolates into the island of Mallorca in Spain has led to new disease outbreaks [39]. In addition, PD-causing isolates of subsp. *fastidiosa* have also invaded Taiwan [43].

The invasion of subsp. *fastidiosa* in Central Taiwan dates back to at least 2002, when the pathogen was first isolated from *Vitis vinifera* L. plants showing symptoms of PD [43]. Two genomic studies suggested that this invasion originated via the introduction of PD-I lineage from the Southeast USA [44, 45]. However, this hypothesis was based on analyses including only two isolates collected in Taiwan and requires further tests. The local studies of PD in Taiwan identified two xylem feeders (*Kolla paulula* and *Bothrogonia ferruginea*) as putative vectors [46, 47]. The local PD control strategies have mainly been centred around vector control, assessment of the habitats suitable for these vectors, eradication of infected plants and alternative host plants, and planting of healthy seedlings [48]. Notably, alternative hosts may function as genetic diversity hotspots of plant pathogens and contribute to the maintenance of vector populations, thus requiring attention in the control strategies [49].

There is another xylem-limited bacterium only found in Taiwan. Leaf scorch symptoms described in pear plants (*Pyrus pyrifolia* cv. Hengshan) in Central Taiwan were

Impact Statement

The plant pathogen *Xylella fastidiosa* subspecies *fastidiosa* causes Pierce's disease (PD) of grapevines and is a threat in multiple regions across North America, Europe, and Asia. In this work, we conducted population genomics analysis of this pathogen in Taiwan. Through sampling of districts with high disease prevalence and comparative analysis with isolates from elsewhere in the world, we confirmed that the Taiwanese populations were all originated from the Southeast USA. Following the invasion, mutation and recombination have contributed to the local expansion of genetic diversity. Intriguingly, although recombination between different subspecies of *X. fastidiosa* have been reported, we found no evidence of recombination between the Taiwanese *X. fastidiosa* and the endemic sister species *Xylella taiwanensis*. These findings provide important information regarding the global transmission of an important plant-pathogenic bacterium and improve our understanding of its evolution following invasion into a new region. Additionally, the newly reported genome sequences provide valuable resources for future studies of this pathogen.

documented in the 1980s, and the appearance of disease symptoms was correlated with the presence of a xylem-limited fastidious bacterium [50]. Isolate PLS229 was obtained from symptomatic pear and sequencing of its 16S rRNA showed 99% similarity with published *X. fastidiosa* isolates [51]. Subsequent analyses found that the average nucleotide identity (ANI) between PLS229 and *X. fastidiosa* representatives ranged between 83.4–83.9%, suggesting that PLS229 represents a new species [52]. This species, named *Xylella taiwanensis*, is the only known sister taxon to *X. fastidiosa* within the same genus. Analysis of the complete genome of *X. taiwanensis* shows that this species shares only ~66–70% of its gene content with *X. fastidiosa* [53].

Both *X. taiwanensis* and *X. fastidiosa* subsp. *fastidiosa* have a negative effect on Taiwanese crops. In the case of subsp. *fastidiosa*, vineyards located in hilly terrains are under high risk of PD infection. In an analysis of survey data from 2002 to 2012, infection rates of grapevines were highest in Waipu (71%), Tongxiao (73%), and Houli (85%) [48]. For comparison, in an adjacent district Zhuolan, only sporadic PD infections were reported and the infection rate was ~1%. Regardless of the infection rates, the genetic diversity of subsp. *fastidiosa* isolates found in these districts and the degree of genotypic divergence among districts are unknown. Moreover, the prevalence of evolutionary events such as gene gain/loss, homologous recombination and mutation rate remain to be investigated. Addressing these points is important to understand how subsp. *fastidiosa* diversifies following an introduction event, and can aid

Table 1. Summary of assembly statistics for the newly sequenced Taiwanese *Xylella fastidiosa* subsp. *fastidiosa* isolates

Isolate	Geographic origin	Grapevine variety	Collection date	N50 (bp)	Genome size (bp)	Coverage (x)
GV210	Tongxiao, Miaoli	Kyoho	2015.08.21	87 160	2 463 940	294
GV215	Houli, Taichung	Kyoho	2017.06.30	99 605	2 463 006	331
GV216	Houli, Taichung	Kyoho	2017.08.03	97 514	2 468 540	285
GV219	Waipu, Taichung	Kyoho	2017.07.24	92 425	2 460 764	237
GV220	Waipu, Taichung	Kyoho	2017.07.24	99 605	2 461 512	288
GV221	Waipu, Taichung	Kyoho	2017.07.24	94 057	2 461 485	258
GV222	Waipu, Taichung	Kyoho	2017.07.24	113 299	2 461 579	254
GV225	Waipu, Taichung	Kyoho	2017.07.24	92 981	2 470 584	272
GV229	Waipu, Taichung	Golden Muscat	2017.07.24	99 551	2 465 605	232
GV230*	Waipu, Taichung	Black Queen	2018.06.24	2 514 993	2 514 993	277
GV231	Waipu, Taichung	Golden Muscat	2018.06.24	101 467	2 490 832	274
GV232	Waipu, Taichung	Golden Muscat	2018.06.24	90 280	2 467 990	272
GV233	Waipu, Taichung	Golden Muscat	2018.06.24	100 549	2 461 948	209
GV234	Waipu, Taichung	Golden Muscat	2018.06.24	100 555	2 463 675	213
GV235	Waipu, Taichung	Golden Muscat	2018.06.24	100 555	2 466 447	197
GV236	Waipu, Taichung	Golden Muscat	2018.06.24	100 555	2 465 286	215
GV237	Waipu, Taichung	Golden Muscat	2018.06.24	92 561	2 468 425	207
GV238	Waipu, Taichung	Golden Muscat	2018.06.24	90 280	2 516 844	285
GV239	Waipu, Taichung	Golden Muscat	2018.06.24	99 551	2 490 161	250
GV240	Waipu, Taichung	Golden Muscat	2018.06.24	124 734	2 469 727	237
GV241	Waipu, Taichung	Golden Muscat	2018.06.24	90 280	2 467 728	219
GV244	Waipu, Taichung	Golden Muscat	2018.12.25	80 059	2 774 440	57
GV245	Waipu, Taichung	Golden Muscat	2018.12.25	92 729	2 600 059	83
GV248	Waipu, Taichung	Black Queen	2019.07.03	98 380	2 456 198	19
GV249	Waipu, Taichung	Black Queen	2019.07.03	86 393	2 456 603	15
GV252	Waipu, Taichung	Golden Muscat	2019.07.03	90 026	2 455 757	16
GV253	Waipu, Taichung	Golden Muscat	2019.07.03	99 557	2 458 547	15
GV263	Waipu, Taichung	Black Queen	2019.12.04	97 977	2 460 836	14
GV264	Waipu, Taichung	Black Queen	2019.12.04	95 089	2 461 237	17
GV265	Waipu, Taichung	Black Queen	2019.12.04	99 087	2 457 696	16
GV266	Zhuolan, Miaoli	Kyoho	2019.12.31	86 635	2 459 085	16

*The reference isolate GV230 whose complete genome sequence is available.

the design and implementation of disease management strategies.

In this study, we conducted an exploratory sampling of the subsp. *fastidiosa* isolates across locations reported to have high PD incidence in Central Taiwan. Specifically, we aim to: (1) provide a first look of the genomic diversity of this pathogen in this region and infer the patterns of

diversification, (2) confirm the source population of these Taiwanese isolates, (3) compare the genetic diversity and diversification patterns of the Taiwanese population with those sampled from elsewhere in the world, and (4) infer the levels of recombination among the Taiwanese subsp. *fastidiosa* isolates as well as with the endemic sister species *X. taiwanensis*.

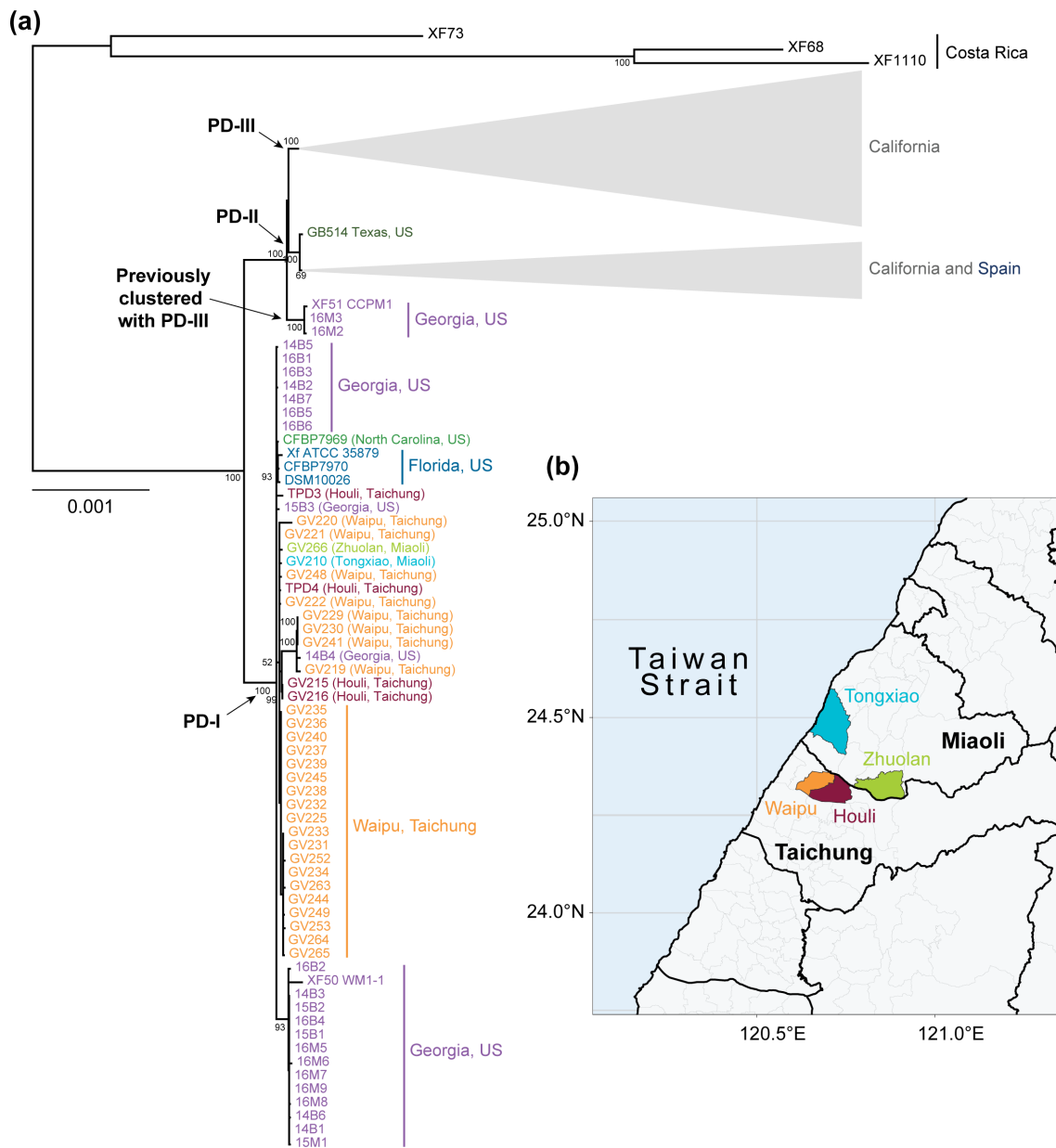


Fig. 1. Relationship of PD-causing isolates. (a) Maximum-likelihood (ML) tree of worldwide PD-causing subsp. *fastidiosa* isolates. The phylogenetic inference was based on the core genome (i.e. shared by >99% of the isolates) without removal of recombinant segments. Costa Rica isolates were used to root the tree. Clades encompassing California and Spain isolates have been compressed to their most recent common ancestor. The points of divergence for the PD-I to PD-III clades are indicated by arrows. Distinct colours are used to differentiate isolates from different geographic regions. (b) Collection locations of the Taiwan isolates.

METHODS

Biological samples and genome sequencing

Thirty-one field isolates were collected from symptomatic grapevines in Central Taiwan (Table 1, Fig. S1, available in the online version of this article). Most of these isolates were from Waipu ($n=27$), while few were from Houli ($n=2$), Tongxiao ($n=1$), and Zhuolan ($n=1$) (Fig. 1b, Table 1). The bias reflects PD prevalence in these regions during our collection trips between 2015 and 2019. Notably, while PD infection

rates were reported to be >70% in Houli and Tongxiao during 2002–2012 [48], we had difficulties in finding infected grapevines in these two districts, suggesting that the local PD control strategies have been effective in recent years. After the infected plant samples were collected, the isolates were cultivated as described [43]. In short, petioles were surface-sterilized with 1% bleach, rinsed with sterile water and then minced in PD2 broth to extract bacterial cells from xylem. The extract was streaked onto PD2 agar medium

and incubated at 28°C for 2–3 weeks in preparation for DNA extraction.

The procedures for whole-genome shotgun sequencing and data processing were based on those described in our previous studies [45, 53]. All bioinformatics tools were used with the default settings unless stated otherwise. Briefly, the DNA samples of all 31 isolates were prepared using the Wizard Genomic DNA Purification Kit (A1120; Promega, USA). The Illumina sequencing libraries were prepared by the Genomic Technology Core (Institute of Plant and Microbial Biology, Academia Sinica) using KAPA LTP Library Preparation Kit (KK8232; Roche, Switzerland) and KAPA Dual-Indexed Adapter Kit (KK8722; Roche, Switzerland) with a target insert size of ~550 bp. The Illumina MiSeq paired-end sequencing service was provided by the Genomics Core (Institute of Molecular Biology, Academia Sinica) using the MiSeq Reagent Nano Kit v2 (MS-103–1003; Illumina, USA). For the reference isolate GV230, additional sequencing was conducted using the Oxford Nanopore Technologies (ONT) MinION platform (FLO-MIN106; R9.4 chemistry and MinKNOW Core v3.6.0) (Oxford Nanopore Technologies, UK). The sequencing library was prepared using the ONT Ligation Kit (SQK-LSK109) without shearing or size selection. Guppy v3.4.5 was used for base calling.

To obtain the complete genome sequence of the reference isolate GV230, the Illumina and ONT reads were combined for *de novo* assembly by using Unicycler v0.4.8-beta [54]. For validation, the Illumina and ONT raw reads were mapped to the assembly using BWA v0.7.12 [55] and Minimap2 v2.15 [56], respectively. The mapping results were programmatically checked using SAMtools v1.2 [57] and manually inspected using IGV v2.3.57 [58]. Gene prediction and annotation were performed using the NCBI prokaryotic genome annotation pipeline [59].

For the remaining 30 isolates, draft genome assemblies were prepared using only Illumina reads. The quality of raw paired FASTQ reads was evaluated using FastQC [60]. Low-quality reads and adapter sequences were removed from all paired raw reads using seqtk v1.2 (<https://github.com/lh3/seqtk>) and cutadapt v1.14 [61], respectively. After pre-processing, SPAdes v3.13 [62, 63] was used for *de novo* assembly with the *-careful* parameter and a *-k* of 21, 33, 55, and 77. Contigs were reordered with Mauve's contig mover function [64] using the complete assembly of GV230 as a reference. Assembled and reordered genomes were then individually annotated using the Prokka pipeline [65].

Additional genome sequences for comparative analysis

In addition to the 31 newly sequenced Taiwanese isolates, 175 genome sequences available from GenBank were included for comparative analysis (Table S1). These include two additional Taiwanese isolates from Houli (i.e. TPD3 and TPD4) [44], 171 isolates from the USA [45], and two isolates from Spain [45]. Those 171 USA isolates include 140 from California and 31 from Southeast USA, representing all three phylogenetic

clades (i.e. PD-I, PD-II, and PD-III) established previously [45]. The sequence types (STs) of all isolates were determined based on an established multilocus sequence typing (MLST) scheme [66, 67] using BIGSdb [68].

Other than the combined data set of 206 *X. fastidiosa* subsp. *fastidiosa* isolates obtained from infected grapevines with PD symptoms, three Costa Rican subsp. *fastidiosa* isolates from non-*Vitis* hosts were included as the outgroup [34]. These three isolates are XF68 from *Psidium* sp., XF73 from *Coffea* sp., and XF1110 from *Vinca* sp. Furthermore, the type strain of *X. taiwanensis* (PLS229) [51–53] was included for between-species comparisons.

Core genome alignments, construction of ML trees and recombination detection

Roary v3.11.2 [69] was used to identify the core, soft-core and accessory genomes among the isolates compared. The core genome was defined as genes shared by >99% of the isolates, the soft-core was defined as genes shared by 95–99% of the isolates, and the accessory genome was defined as genes shared by <95% of the isolates. The analysis was performed separately for three sets of isolates, including: (1) a worldwide set containing all 206 PD-causing isolates (Table S1), (2) a Taiwan-only set containing 33 isolates, and (3) a PD-I/Taiwan set containing all 33 Taiwanese isolates and 28 PD-I isolates from Southeast USA (i.e. the putative source of the population in Taiwan). The ancestor/descendant relationships among populations were established using the worldwide PD-causing subsp. *fastidiosa* maximum-likelihood (ML) tree (see Results) as well as based on previous studies [44, 45]. For each set, genes in the core genome were aligned using Roary to produce a core-genome alignment. The alignment was used for ML phylogeny inference using RAXML [70]. The GTRCAT substitution model was used on tree construction, while tree topology and branch support were assessed with 1000 bootstrap replicates.

The core-genome alignment for the worldwide set was also used to estimate the location of recombinant events using fastGEAR [71]. Lineage-specific recombinant segments (ancestral) and isolate-specific recombinant segments (recent) were identified. Recombinant regions were removed from the alignment to generate a non-recombinant core genome of the worldwide set. This was done to determine the frequency of recombination events, to measure their impact on genetic diversity and the estimated mutation rate, and to estimate the contribution of recombination to genetic differentiation. After the recombinant regions were removed, this non-recombinant core was realigned using MAFFT [72] and used to construct a ML non-recombinant tree of the worldwide set as described. It should be noted that fastGEAR was designed to test recombination in individual gene alignments instead of core-genome alignments; yet, a previous study found that fastGEAR was more conservative than other recombination detection methods such as ClonalFrameML [36]. The number and location of recombinant regions was also estimated for the Taiwan-only and the PD-I/Taiwan sets. In both instances,

recombinant segments were also removed. The location and origin of recombinant segments in the Taiwan-only set were visualized using fastGEAR. In addition, the presence of homologous recombination between the Taiwan-only set and *X. taiwanensis* was evaluated following the same protocol.

Population genetic analyses, haplotype network assessment and population structure

Global measures of genetic diversity were estimated using the R package ‘PopGenome’ [73]. Briefly, PopGenome converts alignments into a biallelic matrix with rows corresponding to sequences and columns corresponding to single nucleotide polymorphism (SNP) positions. All SNPs were classified as 0 (major alleles) or 1 (minor alleles), while missing data and unknown variants were labelled as NA. The number of SNPs, nucleotide diversity (π), Tajima’s D [74], and Watterson’s estimator (θ) [75] were computed. Nucleotide diversity (π) measures the average number of nucleotide differences per site in pairwise comparisons among DNA sequences. Tajima’s D measures the frequency of polymorphisms present in a population and compares that value to the expectation under neutrality. The Watterson θ estimator measures the mutation rate of a population. The analyses were performed using core-genome alignments for the worldwide and PD-I/Taiwan sets, and then repeated using their corresponding non-recombinant cores. For each set, isolates were divided based on their geographic origins.

Haplotype networks were built for: (1) the non-recombinant PD-I/Taiwan set, (2) the recombinant Taiwan-only set, and (3) the non-recombinant Taiwan-only set. Core-genome haplotypes were calculated based on the number of mutations among the analysed isolates. The haplotype network was built using the ‘HaploNet’ function in the R package ‘pegas’ [76]. Haplotype shapes and colours were coded by location. In addition, *SNP-sites* [77] was used to build a SNP alignment from the non-recombinant core-genome alignment for worldwide PD-causing isolates. This alignment was then used to evaluate if panmictic subpopulation clusters existed within the larger context of worldwide PD-causing isolates and to establish which isolates belonged to these potential subpopulations. The existence of genetically distinct subpopulations was assessed using a Bayesian analysis of population structure (BAPS) as implemented in the R package ‘hierBAPS’ [78]. This tool identifies population structure in haploid genomes by integrating over the allele frequencies, then uses this information to partition individual sequences into subpopulation clusters, followed by using the ‘ggtree’ package to annotate these clusters on a phylogenetic tree.

RESULTS

The PD outbreak in Taiwan is linked to at least one introduction from the USA

The core genome phylogenetic analyses of 206 worldwide isolates (Figs 1 and S2) provided strong support to the hypothesis that the introduction of subsp. *fastidiosa* into Taiwan originated from the PD-I lineage in the USA [44, 45].

All 33 Taiwanese isolates, including those two used in the previous studies to propose this hypothesis and an additional 31 newly sequenced in this study, form a monophyletic clade with the PD-I lineage. This clade is robust with 100% bootstrap support when the entire core genome was used (Fig. 1) or when the possible recombinant segments were excluded (Fig. S2), suggesting a single source of introduction from the Southeast USA into Taiwan. Consistent with the phylogenetic relationships, nearly all isolates within this PD-I/Taiwan clade were classified as ST2 based on an established MLST scheme [66, 67]. Isolate GV225 is the only exception that was unclassified (Table S1). For closer examinations within this clade, the branch lengths are mostly very short, which demonstrates that there is little sequence differentiation. This finding suggests that the introduction event into Taiwan occurred recently, which is consistent with local PD monitoring [43, 48]. Comparatively longer branch lengths among GV219, GV229, GV230, and GV241 were observed in the phylogenetic tree. These four isolates form a monophyletic clade with isolate 14B4 from Georgia (Fig. 1a). Nonetheless, this clade is not observed when recombinant segments are removed from the core-genome alignment of all worldwide PD-causing isolates (Fig. S2). When the PD-I/Taiwan set was independently analysed using a haplotype network, two clearly differentiated groups between the Southeast USA (PD-I) and Taiwan populations were observed (Fig. 2a).

A population structure analysis using a non-recombinant core SNP alignment of all worldwide PD-causing isolates (Fig. S3) supports the conclusion that isolates from the PD-I clade and Taiwan are closely related. However, isolates GV215, GV216, GV230, GV235, and GV240 formed a distinct population (Fig. S3, shown in lime green) related to isolate CFBP7969 from North Carolina (USA). In addition, isolates TPD3, TPD4, and GV249 were also assigned to other populations. TPD3 was assigned to the same population as Je57 (PD-II from California, USA), while TPD4 was assigned to the same population as Hopland (PD-III from California, USA), XYL1732 (PD-II from Mallorca, Spain), GB514 (PD-II from Texas, USA) and 16M2 (previously assigned to PD-III from Georgia, USA). No other isolate in this data set was assigned to the same population as GV249. Overall, the population structure analysis identified 18 different populations among the worldwide set of 206 isolates. Nonetheless, most isolates were assigned to five major groups: two for the PD-I clade (Fig. S3, shown in dark blue and lime green), one for the PD-II clade (Fig. S3, shown in orange) and two in the PD-III clade (Fig. S3, shown in red and yellow). Other identified populations were usually composed of three or fewer isolates and could largely be explained by higher sequence divergence.

The PD-causing population in Taiwan is diversifying

The core genome (i.e. shared by >99% of isolates) of the Taiwan-only set included 2041 genes, while the soft-core genome (95–99% of isolates) included 67 genes (Table 2). No genes were found to be uniquely present or absent in all Taiwanese isolates in respect to the Southeast USA (i.e. PD-I). In other words, there is no gene present in all (or most)

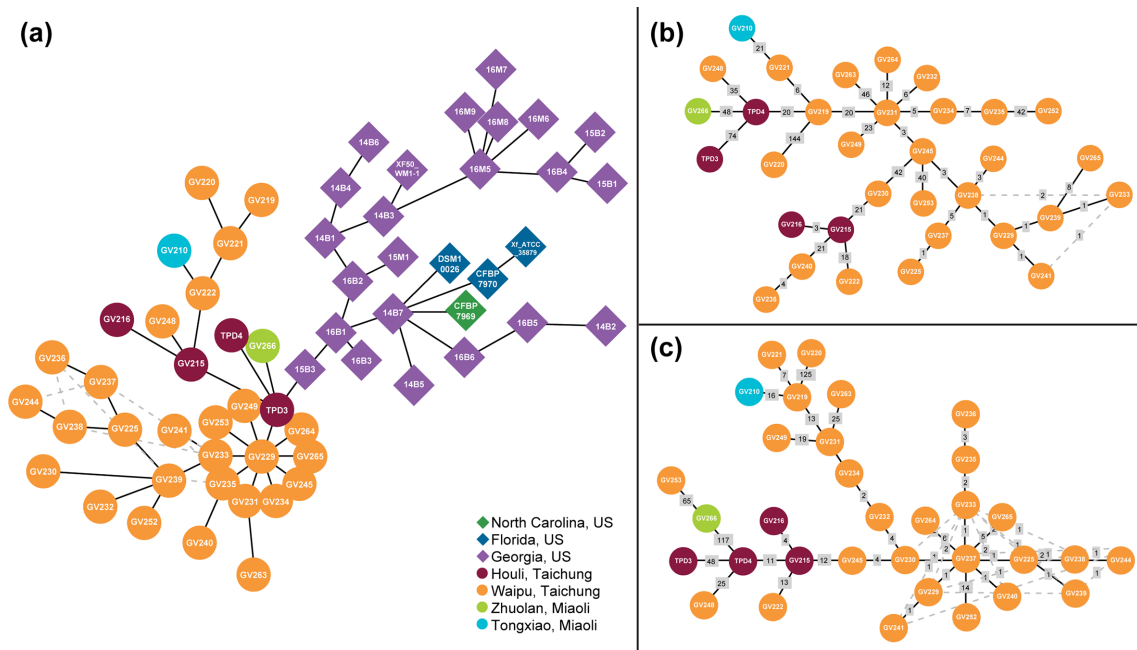


Fig. 2. Haplotype networks showing relationships among PD-causing isolates. Isolate ID is included inside each circle/diamond. (a) Haplotype network between Taiwan (descendant population) and PD-I/Southeast USA (ancestral population) isolates. Network was built based on the core-genome alignment following removal of recombinant segments of the PD-I/Taiwan data set. (b) Haplotype network of Taiwan isolates. Network was built based on the core-genome alignment of the Taiwan-only data set. (c) Non-recombinant haplotype network of Taiwan isolates. Network was built based on the core-genome alignment following removal of recombinant segments of the Taiwan-only data set. Numbers of mutations between isolates are shown in grey boxes. The minimum spanning tree is illustrated with solid black lines. Alternative relationships among isolates are illustrated with dashed grey lines.

Taiwanese isolates that is absent in all (or most) PD-I isolates. Similarly, there is no gene absent in all (or most) Taiwanese isolates that is present in all (or most) PD-I isolates. This finding suggests that there is no clear change in gene content linked to the introduction event. The core-genome haplotype network for the Taiwan-only population (Fig. 2b) showed that several mutations have accumulated within the region without a clear geographic pattern. The non-recombinant core-genome haplotype network confirms this trend (Fig. 2c), but it also suggests that many of these mutations are likely a product of recombinant events.

Overall, the genetic diversity between Southeast USA (PD-I) (336 SNPs, $\pi=2.645 \times 10^{-05}$) and Taiwan (362 SNPs, $\pi=3.045 \times 10^{-05}$) was comparable. Likewise, the mutation rate was similar between these geographic areas (Southeast USA's

$\theta=4.986 \times 10^{-05}$ and Taiwan's $\theta=5.101 \times 10^{-05}$). The similarities remained following removal of recombinant segments from the PD-I/Taiwan set (Table 3). The negative Tajima's D values detected are indicative of a recent population introduction; however, values were less negative in Taiwan (-1.547) compared to the Southeast USA (-1.857). This trend was flipped when recombinant segments were removed (-2.368 for Southeast USA and -2.717 for Taiwan).

Based on the fastGEAR inference of core-genome recombination among Taiwanese isolates, all isolates belonged to one single lineage and there was no clear geographic structuring (Fig. 3a). This observation may or may not be attributed to the sampling bias that 82%(27/33) of these isolates were collected from the same district. Some recent recombination events involving unknown donors (illustrated by the black blocks in Fig. 3a) were inferred. Our BLASTN analysis found that these recombinant segments have 100% sequence identity with other subsp. *fastidiosa* isolates and ~98% identity with isolates belonging to subsp. *pauca* or *multiplex*. This result suggests that not all sequence diversity found within the region has been sampled. When the sympatric sister species *X. taiwanensis* was included in the inference, the size of core genome is greatly reduced (i.e. from 1.84 to 0.1 Mb) and no inter-species recombination was identified (Fig. 3b).

Table 2. Summary of pan-genome analysis of worldwide PD-causing *X. fastidiosa* data and the Taiwanese population alone. Threshold for inclusion indicates the proportion of isolates harbouring the genes

Threshold for inclusion	Worldwide (N=206)	Taiwan-only (N=33)
Core (>99%)	1715	2041
Soft-core (95–99%)	258	67
Shell (15–95%)	652	272
Cloud (<15%)	7379	757

Table 3. Genome-wide genetic diversity measurements

Analysis	Population (N)	Alignment length (bp)	SNPs	Pi	Watterson	Tajima
With recombination						
PD-causing isolates+Costa Rica	Southeast USA (32)	1 562 792	926	1.346×10 ⁻⁰⁴	1.471×10 ⁻⁰⁴	-0.329
	Taiwan (33)		215	1.056×10 ⁻⁰⁵	3.390×10 ⁻⁰⁵	-2.629
PD-I/Taiwan	Southeast USA (28)	1 748 449	336	2.645×10 ⁻⁰⁵	4.986×10 ⁻⁰⁵	-1.857
	Taiwan (33)		362	3.045×10 ⁻⁰⁵	5.101×10 ⁻⁰⁵	-1.547
Without recombination						
PD-causing isolates+Costa Rica	Southeast USA (32)	1 410 222	864	1.331×10 ⁻⁰⁴	1.521×10 ⁻⁰⁴	-0.485
	Taiwan (33)		193	1.044×10 ⁻⁰⁵	3.372×10 ⁻⁰⁵	-2.633
PD-I/Taiwan	Southeast USA (28)	1 624 189	272	1.739×10 ⁻⁰⁵	4.345×10 ⁻⁰⁵	-2.368
	Taiwan (33)		398	1.767×10 ⁻⁰⁵	6.038×10 ⁻⁰⁵	-2.717

DISCUSSION

PD-causing subsp. *fastidiosa* was introduced into Taiwan from the Southeast USA

Despite attempts at disease surveillance, quarantine and control, novel plant pathogen outbreaks are still observed in numerous locations worldwide [23, 79–82]. Following introduction, the reconstruction of invasion routes, the identification of a pathogen's establishment and spread mechanisms becomes crucial for developing mitigation strategies [83]. Here, we have confirmed previous studies suggesting that PD-causing subsp. *fastidiosa* was introduced to Taiwan from the Southeast USA (PD-I) [44, 45] using multiple lines of evidence based on whole-genome sequence data. The results support that the ancestor/descendant relationship between these populations was not an artefact caused by limited sampling of only two isolates from Houli in the previous study [45].

All except one of the PD-I and Taiwan isolates included here were classified as ST2 based on an established MLST scheme [66, 67] (Table S1). In addition, whole-genome-based phylogenetic analyses show that the Taiwan and Southeast USA (PD-I) isolates form one strongly supported monophyletic clade. Moreover, short branch length and lack of phylogenetically supported geographic clusters are indicative of few nucleotide changes differentially accumulating in both populations. This finding is consistent with reports of other introduction events involving *X. fastidiosa* [30, 37, 38, 42, 84–88]. Likewise, nucleotide diversity and mutation rate were comparable between Southeast USA (PD-I) and Taiwan isolates. Previous studies have found that introduced populations of subsp. *pauca* and subsp. *fastidiosa* have lower genetic diversity than native ones [34]. Yet, with time, nucleotide differences accumulate as a product of ecological and environmentally mediated pressures as well as the result of non-adaptive evolution [45]. This has been observed in the introduction of subsp. *fastidiosa* into the USA ~150 years ago [36]. Therefore, the identification of the same MLST sequence type in both Taiwan and the Southeastern USA (PD-I), the lack of genome-wide

nucleotide differences among the two populations, and the similar genetic diversity and mutation rate between Taiwan and the Southeast USA (PD-I) are all indicative that this was a recent introduction event.

A previous study conducted using only two isolates from Taiwan found four putative gene gains and five putative losses in respect to the ancestral Southeast USA (PD-I) population [45]. Our updated pangenome analysis used 33 isolates from Taiwan but failed to find a gene gain/loss pattern that could be linked to this introduction. This suggests that the previous observation was an artefact of the small sample size and that no clear gene gain/loss trends can be attributed to a founder effect. Furthermore, since gene gain/loss often precedes nucleotide substitutions and indels in bacterial genome evolution [89], this is consistent with the record of invasion of *X. fastidiosa* to Taiwan in 2002. However, one caveat is that most of the genome sequences included in this gene content analysis are incomplete draft assemblies, such that the finding should be interpreted with caution.

Although a single source population (i.e. PD-I in Southeast USA) was inferred, it is difficult to determine if a single or multiple introduction events into Taiwan have taken place. Phylogenetic analyses of all worldwide PD-causing isolates and the haplotype network analysis conducted on the PD-I/Taiwan clade are indicative of a single introduction event. On the other hand, population structure analyses show that two isolates from Houli (GV215 and GV216), and three isolates from Waipu (GV230, GV235 and GV240) form a distinct population with CFBP7969 from North Carolina (USA), hinting to multiple introduction events. These isolates are neither phylogenetically nor geographically aggregated, so it is difficult to confidently ascertain if this finding is due to the differentiation of a genetically unique population in Taiwan, or if it is the product of an additional introduction event. In previous studies assessing the introduction of subsp. *multiplex* into Europe, both phylogenetic and genetic diversity data supported the hypothesis of multiple introductions from the USA [40]. Here, it is possible that our conflicting results are

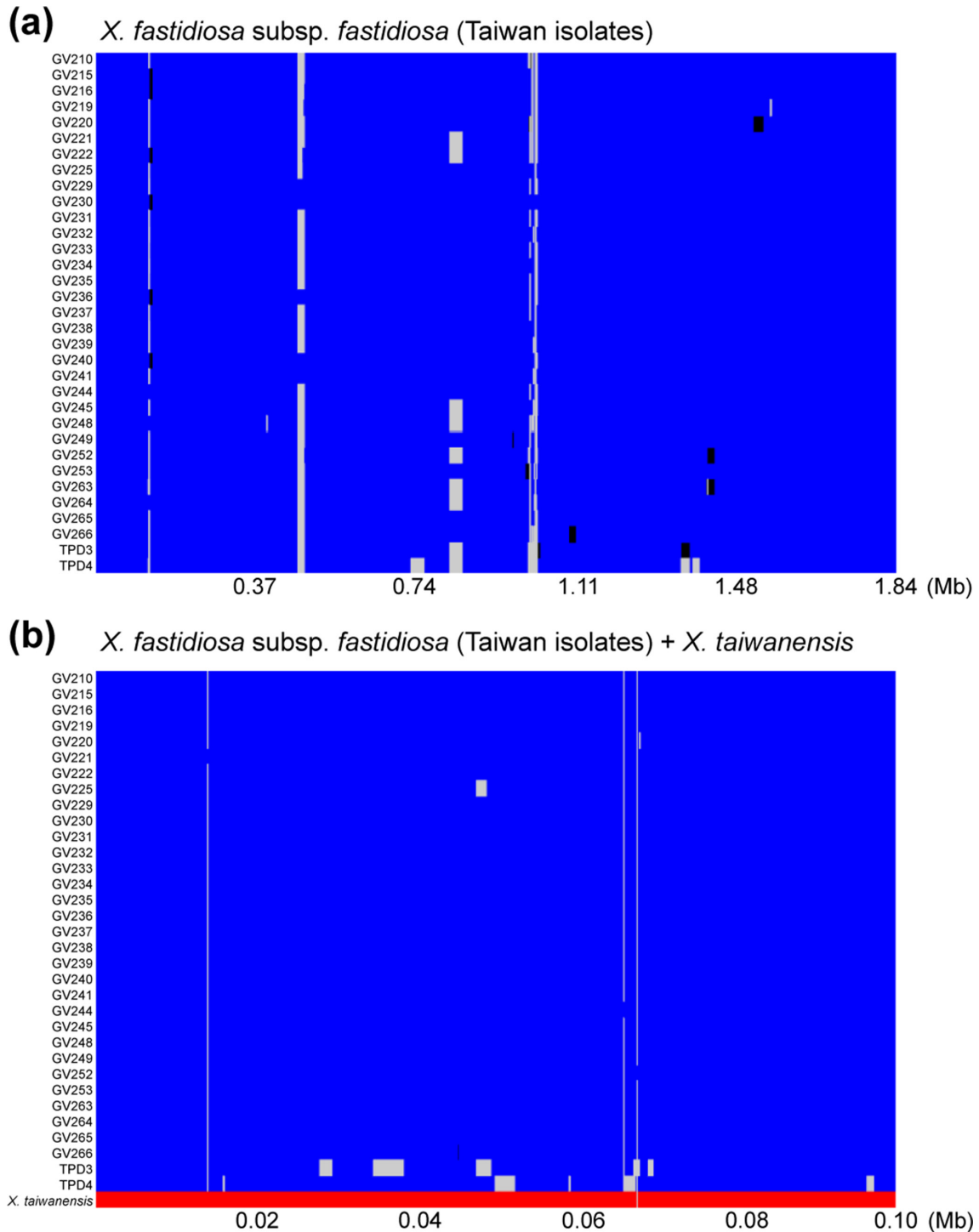


Fig. 3. Recombination events and lineage assignments based on the core genome. Recombination events are shown across the length of the core-genome alignments. Larger areas in blue (i.e. a single lineage for all subsp. *fastidiosa* isolates in Taiwan) or red (i.e. a separate lineage for *X. taiwanensis*) represent recipient sequences. Recombinant segments from unidentified lineages are shown in black and alignment gaps are shown in grey. (a) All 33 subsp. *fastidiosa* isolates in Taiwan. The core-genome alignment contains 1 842 804 aligned nucleotide sites. (b) All 33 subsp. *fastidiosa* isolates in Taiwan and *X. taiwanensis*. The core-genome alignment contains 98 310 aligned nucleotide sites.

linked to genetic diversity losses associated with a founder event masking the presence of simultaneously introduced strains. Moreover, since most isolates were collected from the two neighbouring districts in Taichung City, it is possible that additional introduction events might have not been detected. Although additional samples from other regions in Taiwan can help to further investigate this question, such effort may be difficult following the successful PD control in recent years.

Taiwanese isolates are differentiating via multiple mechanisms

It should be noted that despite evidence suggesting a short evolutionary time since the introduction of PD-causing subsp. *fastidiosa* to Taiwan, sequence divergence in core-genome genes is already observed in the region. It is not possible to confidently determine if this is due to variable evolutionary pressures across locations in Taiwan or an intrinsic genetic characteristic carried over from the ancestor population(s). Waipu and Houli are neighbouring districts that are <10km apart and share similar warm humid subtropical climate conditions, so it would be expected that environmental pressures would be similar among populations. Likewise, Zhuolan and Tongxiao are separated from Waipu and Houli by only ~20km. This proximity could also explain the lack of clear phylogenetic differentiation among isolates from distinct locations.

Another notable aspect is how the haplotype network analysis responded to the removal of recombinant regions. When recombinant segments were included, the minimum spanning tree (solid black lines) was defined and showed no clear separation among Taiwanese isolates from different locations. Following the removal of recombinant segments, Houli isolates clustered within the network and the relationships among Waipu isolates became less clear (dashed grey lines). This is evidence that recombination significantly contributes to the genetic differentiation of PD-causing subsp. *fastidiosa* in Taiwan; however, this result could also be sensitive to low sample size in certain regions. The removal of recombinant segments from the worldwide PD-causing core-genome alignments also had an effect in the topology of the non-recombinant phylogenetic tree of worldwide PD-causing isolates. Mainly, a monophyletic clade grouping GV219, GV229, GV230, and GV241 with I4B4 from Georgia was absent from the non-recombinant tree of worldwide PD-causing isolates. This provides further support that recombination plays a significant role in the diversification of the Taiwanese population. In addition, it also suggests that, to some degree, the recombinant capacity of some Taiwanese PD-causing isolates might be linked to the isolates introduced from the USA.

Previous studies have shown that natural competency and recombinant capacity are isolate dependent [37, 38, 40, 90]. Moreover, recombinant events can be carried over from ancestral populations into new introductions; particularly, in instances where these introductions are recent [40]. Therefore, our results are indicative that recombinant-prone subsp.

fastidiosa isolates might have been introduced to Taiwan from the Southeast USA or that they might have evolved there. Further sampling from additional regions in Taiwan would be required to evaluate this. Nonetheless, the existence of unknown recombining sequences in Taiwanese isolates suggest that genetic diversity within Taiwan is higher than what can be defined via our current sampling. Finally, even though homologous recombination is known to occur in sympatric subsp. *pauciflora* populations [91] and among different *X. fastidiosa* subspecies [38], no recombination events between PD-causing subsp. *fastidiosa* and *X. taiwanensis* were detected. This result is expected based on the low level of genome-wide nucleotide sequence identity between these two species at ~84% [52, 53] and further supports that these two species are ecologically and/or biologically isolated despite diseased grapevines and pears can be found in proximity in Central Taiwan.

CONCLUSION

In globally distributed pathogens, understanding the relationships among populations and the location-specific mechanisms of diversification has significant implications in the development of successful management techniques. Here, we used phylogenetic analysis to investigate the introduction of subsp. *fastidiosa* into Taiwan and confirm that the invasion originated from a single-source population (i.e. the Southeast USA). Haplotype network analysis, pan-genome trends, and global genetic diversity measures support this hypothesis. Following its recent establishment within Taiwan, PD-causing subsp. *fastidiosa* has differentiated across geographic locations. We propose that part of this diversification can be linked to homologous recombination and highlight the need for further studies within the region. Moreover, focus should be allocated in acquiring further samples from other location within Taiwan and addressing how natural selection, and other evolutionary forces, might impact local adaptation across regions.

Funding Information

Funding was provided by the Ministry of Science and Technology of Taiwan (MOST 106-2923-B-002-005-MY3) to CWT and Academia Sinica to CHK. The work also received funding from the PD/GWSS Research Program, California Department of Food and Agriculture, and the European Union's Horizon 2020 research and innovation program under grant agreement no. 727987, "Xylella fastidiosa Active Containment Through a Multidisciplinary-Oriented Research Strategy XF-ACTORS". The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgements

The Illumina and Oxford Nanopore sequencing library preparation services were provided by the Genomic Technology Core (Institute of Plant and Microbial Biology, Academia Sinica). The Illumina MiSeq sequencing service was provided by the Genomics Core (Institute of Molecular Biology, Academia Sinica). We thank Yi-Ming Tsai for technical assistance.

Author contributions

Conceptualization: A.I.C., C.W.T., R.P.P.A., C.H.K. Funding acquisition: C.W.T., R.P.P.A., C.H.K. Investigation: A.I.C., C.W.T., C.C.S., L.W.W., Y.C.L., S.T.C. Methodology: A.I.C., C.C.S., S.T.C., C.H.K. Project administration:

C.W.T., R.P.P.A., C.H.K. Resources: C.W.T., L.W.W., C.C.S., C.H.K. Supervision: C.W.T., R.P.P.A., C.H.K. Validation: A.I.C. Visualization: A.I.C., L.W.W., Y.C.L. Writing – original draft: A.I.C. Writing – review and editing: A.I.C., C.W.T., R.P.P.A., C.H.K.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- de Bruyn M, Stelbrink B, Morley RJ, Hall R, Carvalho GR, et al. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Syst Biol* 2014;63:879–901.
- Hughes AC. Understanding the drivers of Southeast Asian biodiversity loss. *Ecosphere* 2017;8:e01624.
- Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, et al. Biogeography of the Indo-Australian Archipelago. *Annu Rev Ecol Evol Syst* 2011;42:205–226.
- Sholihah A, Delrieu-Trottin E, Condamine FL, Wowor D, Rüber L, et al. Impact of Pleistocene eustatic fluctuations on evolutionary dynamics in Southeast Asian biodiversity hotspots. *Syst Biol* 2021;70:940–960.
- Liao C-C, Chen C-H. Investigation of floristic similarities between Taiwan and terrestrial ecoregions in Asia using GBIF data. *Bot Stud* 2017;58:15.
- Hsieh C-F. Composition, endemism and phytogeographical affinities of the Taiwan flora. *Taiwania* 2002;47:298–310.
- Wu S-H, Hsieh C-F, Rejmánek M. Catalogue of the naturalized flora of Taiwan. *Taiwania* 2004;49:16–31.
- Yeh Y. Status and management of invasive species in Taiwan. Food and Fertilizer Technology Center. Food and Fertilizer Technology Center. 2005. <https://www.fftcc.org.tw/en/publications/main/439>
- Wu ML. The threat of invasive alien pathogens to trees. *For Res News* 2006;13:15–17.
- Martins PMM, Merfa MV, Takita MA, De Souza AA. Persistence in phytopathogenic bacteria: do we know enough? *Front Microbiol* 2018;9:1099.
- Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, et al. Threats to global food security from emerging fungal and oomycete crop pathogens. *Nat Food* 2020;1:332–342.
- Avery SV, Singleton I, Magan N, Goldman GH. The fungal threat to global food security. *Fungal Biol* 2019;123:555–557.
- Bebber DP, Field E, Gui H, Mortimer P, Holmes T, et al. Many unreported crop pests and pathogens are probably already present. *Glob Change Biol* 2019;25:2703–2713.
- Bebber DP, Gurr SJ. Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet Biol* 2015;74:62–64.
- Oerke E-C. Crop losses to pests. *J Agric Sci* 2005;144:31–43.
- Paini DR, Sheppard AW, Cook DC, De Barro PJ, Worner SP, et al. Global threat to agriculture from invasive species. *Proc Natl Acad Sci U S A* 2016;113:7575–7579.
- Savary S, Bregaglio S, Willcoquet L, Gustafson D, Mason D'Croz D, et al. Crop health and its global impacts on the components of food security. *Food Sec* 2017;9:311–327.
- Savary S, Ficke A, Aubertot J-N, Hollier C. Crop losses due to diseases and their implications for global food production losses and food security. *Food Sec* 2012;4:519–537.
- Savary S, Willcoquet L, Pethybridge SJ, Esker P, McRoberts N, et al. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* 2019;3:430–439.
- Strange RN, Scott PR. Plant disease: A threat to global food security. *Annu Rev Phytopathol* 2005;43:83–116.
- He D, Zhan J, Xie L. Problems, challenges and future of plant disease management: from an ecological point of view. *Journal of Integrative Agriculture* 2016;15:705–715.
- Occhibove F, Chapman DS, Mastin AJ, Parnell SSR, Agstner B, et al. Eco-epidemiological uncertainties of emerging plant diseases: The challenge of predicting *Xylella fastidiosa* dynamics in novel environments. *Phytopathology* 2020;110:1740–1750.
- Parnell S, van den Bosch F, Gottwald T, Gilligan CA. Surveillance to inform control of emerging plant diseases: an epidemiological perspective. *Annu Rev Phytopathol* 2017;55:591–610.
- McDonald BA, Stukenbrock EH. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Phil Trans R Soc B* 2016;371:20160026.
- Read AF. The evolution of virulence. *Trends Microbiol* 1994;2:73–76.
- Croll D, McDonald BA. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol Ecol* 2017;26:2027–2040.
- Giraud T, Gladioux P, Gavritets S. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol* 2010;25:387–395.
- Mhedbi-Hajri N, Hajri A, Boureau T, Darrasse A, Durand K, et al. Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLOS ONE* 2013;8:e58474.
- Dutta A, Hartmann FE, Francisco CS, McDonald BA, Croll D. Mapping the adaptive landscape of a major agricultural pathogen reveals evolutionary constraints across heterogeneous environments. *ISME J* 2021;15:1402–1419.
- Sicard A, Zeilinger AR, Vanhove M, Schartel TE, Beal DJ, et al. *Xylella fastidiosa*: insights into an emerging plant pathogen. *Annu Rev Phytopathol* 2018;56:181–202.
- Chatterjee S, Almeida RPP, Lindow S. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu Rev Phytopathol* 2008;46:243–271.
- Rapicavoli J, Ingel B, Blanco-Ulate B, Cantu D, Roper C. *Xylella fastidiosa*: an examination of a re-emerging plant pathogen. *Mol Plant Pathol* 2018;19:786–800.
- Roper C, Castro C, Ingel B. *Xylella fastidiosa*: bacterial parasitism with hallmarks of commensalism. *Curr Opin Plant Biol* 2019;50:140–147.
- Castillo AI, Chacón-Díaz C, Rodríguez-Murillo N, Coletta-Filho HD, Almeida RPP. Impacts of local population history and ecology on the evolution of a globally dispersed pathogen. *BMC Genomics* 2020;21:369.
- Nunney L, Azad H, Stouthamer R. An experimental test of the host-plant range of nonrecombinant strains of North American *Xylella fastidiosa* subsp. *multiplex*. *Phytopathology* 2019;109:294–300.
- Vanhove M, Sicard A, Ezennia J, Leviten N, Almeida RPP. Population structure and adaptation of a bacterial pathogen in California grapevines. *Environ Microbiol* 2020;22:2625–2638.
- Nunney L, Yuan X, Bromley RE, Stouthamer R. Detecting genetic introgression: High levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Appl Environ Microbiol* 2012;78:4702–4714.
- Nunney L, Hopkins DL, Morano LD, Russell SE, Stouthamer R. Intersubspecific recombination in *Xylella fastidiosa* strains native to the United States: Infection of novel hosts associated with an unsuccessful invasion. *Appl Environ Microbiol* 2014;80:1159–1169.
- Gomila M, Moralejo E, Busquets A, Seguí G, Olmo D, et al. Draft genome resources of two strains of *Xylella fastidiosa* XYL1732/17 and XYL2055/17 isolated from Mallorca vineyards. *Phytopathology* 2019;109:222–224.
- Landa BB, Castillo AI, Giampetruzzi A, Kahn A, Román-Écija M, et al. Emergence of a plant pathogen in Europe associated with multiple intercontinental introductions. *Appl Environ Microbiol* 2020;86:e01521–19.
- Olmo D, Nieto A, Adrover F, Urbano A, Beidas O, et al. First detection of *Xylella fastidiosa* infecting cherry (*Prunus avium*) and *Polygala myrtifolia* plants, in Mallorca Island, Spain. *Plant Dis* 2017;101:1820.
- Saponari M, Giampetruzzi A, Loconsole G, Boscia D, Saldarelli P. *Xylella fastidiosa* in olive in Apulia: Where we stand. *Phytopathology* 2019;109:175–186.

43. Su C-C, Chang CJ, Chang C-M, Shih H-T, Tzeng K-C, et al. Pierce's disease of grapevines in Taiwan: Isolation, cultivation and pathogenicity of *Xylella fastidiosa*. *J Phytopathol* 2013;161:389–396.
44. Castillo AI, Tuan S-J, Retchless AC, Hu F-T, Chang H-Y, et al. Draft whole-genome sequences of *Xylella fastidiosa* subsp. *fastidiosa* strains TPD3 and TPD4, isolated from grapevines in Hou-li, Taiwan. *Microbiol Resour Announc* 2019;8:e00835-19.
45. Castillo AI, Bojanini I, Chen H, Kandel PP, De La Fuente L, et al. Allopatric plant pathogen population divergence following disease emergence. *Appl Environ Microbiol* 2021;87:e02095-20.
46. Lin YS, Chang YL. The insect vectors of Pierce's disease on grapevines in Taiwan. *Formos Entomol* 2012;32:155–167.
47. Tuan S-J, Hu F-T, Chang H-Y, Chang P-W, Chen Y-H, et al. *Xylella fastidiosa* transmission and life history of two Cicadellinae sharpshooters, *Kolla paulula* and *Bothrogonia ferruginea* (Hemiptera: Cicadellidae), in Taiwan. *J Econ Entomol* 2016;109:1034–1040.
48. Su C-C, Chang C-M, Chang C-J, Su W-Y, Chu J-C, et al. Occurrence of Pierce's disease of grapevines and its control strategies in Taiwan. *Plant Pathol Bull* 2013;22:245–258.
49. Wistrom C, Purcell AH. The fate of *Xylella fastidiosa* in vineyard weeds and other alternate hosts in California. *Plant Dis* 2005;89:994–999.
50. Leu L-S, Su C-C. Isolation, cultivation, and pathogenicity of *Xylella fastidiosa*, the causal bacterium of pear leaf scorch disease in Taiwan. *Plant Dis* 1993;77:642-646.
51. Su C-C, Deng W-L, Jan F-J, Chang C-J, Huang H, et al. Draft genome sequence of *Xylella fastidiosa* pear leaf scorch strain in Taiwan. *Genome Announc* 2014;2:e00166-14.
52. Su C-C, Deng W-L, Jan F-J, Chang C-J, Huang H, et al. *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. *Int J Syst Evol Microbiol* 2016;66:4766–4771.
53. Weng L-W, Lin Y-C, Su C-C, Huang C-T, Cho S-T, et al. Complete genome sequence of *Xylella taiwanensis* and comparative analysis of virulence gene content with *Xylella fastidiosa*. *Front Microbiol* 2021;12:684092.
54. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
55. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–1760.
56. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–26.
59. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
60. Wingett SW, Andrews S. FastQ screen: A tool for multi-genome mapping and quality control. *F1000Res* 2018;7:1338.
61. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;17:10.
62. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
63. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 2013;20:714–737.
64. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, et al. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* 2009;25:2071–2073.
65. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
66. Scally M, Schuenzel EL, Stouthamer R, Nunney L. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Appl Environ Microbiol* 2005;71:8491–8499.
67. Yuan X, Morano L, Bromley R, Spring-Pearson S, Stouthamer R, et al. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. *Phytopathology* 2010;100:601–611.
68. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
69. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
70. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
71. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, et al. Efficient inference of recent and ancestral recombination within bacterial populations. *Mol Biol Evol* 2017;34:1167–1182.
72. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
73. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: An efficient Swiss Army knife for population genomic analyses in R. *Mol Biol Evol* 2014;31:1929–1936.
74. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–595.
75. Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975;7:256–276.
76. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 2010;26:419–420.
77. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* 2016;2:e000056.
78. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res* 2018;3:93.
79. LeBlanc N, Cubeta MA, Crouch JA. Population genomics trace clonal diversification and intercontinental migration of an emerging fungal pathogen of boxwood. *Phytopathology* 2021;111:184–193.
80. Moralejo E, Gomila M, Montesinos M, Borrás D, Pascual A, et al. Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe. *Commun Biol* 2020;3:560.
81. Soubeyrand S, de Jerphanion P, Martin O, Saussac M, Manceau C, et al. Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent. *New Phytol* 2018;219:824–836.
82. Vieira A, Silva DN, Várzea V, Paulo OS, Batista D. Novel insights on colonization routes and evolutionary potential of *Colletotrichum kahawae*, a severe pathogen of *Coffea arabica*. *Mol Plant Pathol* 2018;19:2488–2501.
83. Robert S, Ravigne V, Zapater M-F, Abadie C, Carlier J. Contrasting introduction scenarios among continents in the worldwide invasion of the banana fungal pathogen *Mycosphaerella fijiensis*. *Mol Ecol* 2012;21:1098–1114.
84. Giampetruzzi A, Saponari M, Loconsole G, Boscia D, Savino VN, et al. Genome-wide analysis provides evidence on the genetic relatedness of the emergent *Xylella fastidiosa* genotype in Italy to isolates from Central America. *Phytopathology* 2017;107:816–827.
85. Nunney L, Ortiz B, Russell SA, Ruiz Sánchez R, Stouthamer R. The complex biogeography of the plant pathogen *Xylella fastidiosa*: Genetic evidence of introductions and subspecific introgression in Central America. *PLoS ONE* 2014;9:e112463.
86. Nunney L, Schuenzel EL, Scally M, Bromley RE, Stouthamer R. Large-scale intersubspecific recombination in the plant-pathogenic

- bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Appl Environ Microbiol* 2014;80:3025–3033.
87. Nunney L, Vickerman DB, Bromley RE, Russell SA, Hartman JR, et al. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. *Appl Environ Microbiol* 2013;79:2189–2200.
88. Nunney L, Yuan X, Bromley R, Hartung J, Montero-Astúa M, et al. Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the U.S. *PLOS ONE* 2010;5:e15488.
89. Iranzo J, Wolf YI, Koonin EV, Sela I. Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nat Commun* 2019;10:5376.
90. Vanhove M, Retchless AC, Sicard A, Rieux A, Coletta-Filho HD, et al. Genomic diversity and recombination among *Xylella fastidiosa* subspecies. *Appl Environ Microbiol* 2019;85:e02972-18.
91. Almeida RPP, Nascimento FE, Chau J, Prado SS, Tsai C-W, et al. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Appl Environ Microbiol* 2008;74:3690–3701.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.