Project title:	What are the factors that make *Xylella fastidiosa* pathogenic and host-specific?
Project number:	3114 0052
Project leader:	Prof. Robert W. Jackson, University of Reading and University of Birmingham
Report:	Annual report, October 2019
Previous report:	N/A
Key staff:	Dr Michelle Hulin, NIAB EMR Dr Emma Cascant López, NIAB EMR Dr Glyn Barrett, University of Reading
Location of project:	NIAB EMR, East Malling ME19 6BJ
Industry Representative:	Mr Brian Fraser, Oakover Nurseries, Ashford TN26 1AR Mr Jamie Dewhurst, J&A Growers, Warwick CV35 8EB
Date project commenced:	01.10.2018

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

Michelle Hulin

Research Leader

NIAB EMR

Signature ..... Date .....10/02/2020.....

Robert Jackson

Project leader

University of Reading and University of Birmingham

Signature ..... Date 11/02/20.....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature ..... Date .....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

CONTENTS

GROWER SUMMARY

Headline

Fundamental knowledge of the bacterium Xylella fastidiosa – detected in over 350 different plant species and causing diseases in many crops – can lead to targeted treatment plans, instead of destroying acres of valuable crops.

Background

What is *Xylella fastidiosa*?

Xylella fastidiosa (*Xf*) is a bacterium that lives in the foregut of insects and the xylem of plants and causes diseases in several economically significant crops, including Pierce's disease (PD) of grapevine, phony peach disease (PPD), oak leaf scorch (OLS) and olive quick-decline syndrome (OQDS). *Xf* has been detected in over 350 different plant species in Europe alone, but detection of the bacterium in a plant does not necessarily lead to disease. However, these asymptomatic hosts may act as a reservoir for insect vectors to spread the bacteria to susceptible plants. Very little is known about the mechanisms behind what makes *Xf* cause symptoms in some plants but not in others. This research project investigates what makes *Xf* host-specific and pathogenic (disease-causing) by using molecular and computational biology. More specifically, the genes that encode effector proteins which regulate biological activity. Effector proteins are secreted by bacteria and interact with a host plant's immune system, the importance of which is explained later in this grower summary.

Currently, there is no treatment available for diseases caused by *Xf*. Management measures are restricted to insect vector control, pruning of infected plant tissue and destruction of the infected host. All surrounding potential plant hosts in a 100 m radius are destroyed and a demarcation order of a 5 km radius is set up banning the movement of any plant material from within this area.

Plant hosts of *X. fastidiosa*

There are over 350 different potential host plants in Europe alone, many of which are economically important crops and could devastate a country's economy if affected. The bacterium has not yet been detected in the UK, but the threat is very high as many of its host plants are grown here. These include, but are not limited, to plants grown in forest nurseries, as protected crops and ornamental garden plants. Some of the most significant crops are:

> alfalfa, bay, blueberry, *Brassica*, *Cercis* (redbuds), *Chionanthus* (fringe tree), *Cytisus* (broom), elderberry, elm, fig, grapevine, *Hedera* (ivy), *Hypericum* (St. John's Wort), magnolia, maple, mulberry, *Nandina domestica* (sacred bamboo), lavender, oak, olive, pear, *Prunus* (e.g. apricot, cherry, plum), *Rubus* (e.g. raspberries, blackberries), *Rosa*, rosemary, strawberry, *Trifolium* (e.g. clover), walnut, willow.

Xf's large host range, its long incubation period (which may be up to six months in some plants), and rapid spread, makes it a highly threatening pathogen. However, in order to come up with effective treatment plans for affected plants, it is important to understand the fundamental biology of the disease-causing bacteria.

How do plant bacteria cause disease?

Phytopathogens, or plant pathogens, have the ability to invade the host, avoid host defence mechanisms and ensure disease progression by secreting virulence factors. Virulence factors are proteins, lipids and carbohydrates produced by the pathogen. One of the best-characterised virulence factors include effector proteins, which are secreted through secretion systems or channels in the bacterial cell. Those secreted through type 3 and type 4 secretion

systems (T3SS and T4SS, respectively) are among the most extensively studied in structure and function. The majority of bacterial phytopathogens have been found to secrete effectors through the T3SS, for example *Pseudomonas syringae*, *Erwinia* spp. and *Xanthomonas* spp. However, *Xf* lacks the T3SS, meaning its strategy to cause disease may be quite different. Effectors may have several functions. In *Xf*, for example, a number of effector proteins are found to be involved in biofilm formation. A biofilm is an adhesive state of bacteria, where they collect in clusters. In the case of *Xf*, biofilm formation often leads to the blocking of the plant's xylem, which stops the flow of water and minerals in the vessels and thus disease symptoms appear.

X. fastidiosa is an increasing threat in the European continent

Xf is believed to be native to the Americas and outbreaks of diseases caused by the bacterium within Europe have only been discovered in 2013. The first outbreak of *Xf* in Europe was detected in Italy, followed by France and Spain, and isolated cases in the Netherlands, Belgium, Switzerland, Germany and Portugal. *Xf* spread has been connected with humans moving infected plants, resulting in distribution of *Xf* across large geographical distances. In Europe, four *Xf* subspecies have been identified: *fastidiosa, multiplex, pauca and sandyi*. Subspecies *fastidiosa* originated in Central America, *multiplex* in North America and *pauca* in South America. The origin of *sandyi* is still under debate. A fifth subspecies, *tashke*, has only been found in the Americas. And a sixth subspecies (*morus*) has been proposed but is still under review.

The importance of studying *X. fastidiosa*

Xf is an increasing threat to British agriculture. Climate change makes the environment more suitable for *Xf* which is known to favour warmer regions, but international plant trade is also growing every year, meaning the bacteria have many ways to enter the country.

This research will provide a better understanding of the evolutionary history of *Xf* and the molecules involved in disease progression, which can ultimately help with the generation of targeted treatments for plants infected by *Xf*. There is currently no treatment solution for plants infected by *Xf*, with the only option for an outbreak being destruction of the host, its surroundings and a quarantine order. This research may gain more insight into the complex host range of the bacterium, its yet unknown mode of action within the plant, and determine why the pathogen causes disease in some hosts but remains asymptomatic in others. Ultimately, the outcome of the study could pave the way to implementing further control measures and creating diagnostic tools for the prevention of an outbreak. Genomics can create diagnostics, and understanding how the pathogen causes disease could lead to a potential treatment rather than having to destroy bacteria-carrying hosts and face huge economic loss.

Summary

Currently, the only control measure of *Xf* is prevention and destruction of plant hosts. There are several reasons why there is still very little known about *Xf*, some of which include its long incubation period in the plant, difficulty to cultivate *in vitro* (in test tubes) and thus study in the laboratory, and the many asymptomatic host plants which the bacterium lives in without causing any disease. The ability of computational methods to investigate an organism's genome has become very powerful, allowing a better understanding of the organism. Investigating genes that are involved in disease development will help with the understanding of the bacterium's molecular biology. Understanding how the bacterium works and causes disease on a molecular level could bring us a step closer to establishing a targeted treatment plan for this devastating bacterium.

A number of known effector proteins are promising – but this is on-going research. The first detection of *Xf* in coffee plants (*Coffea arabica*) in Colombia is also described in the science

section. The detection of *Xf* in the country opens more questions about this fascinating bacterium. Colombia does not appear to have an outbreak of *Xf*, unlike its neighbouring country Brazil, where coffee leaf scorch due to *Xf* is rampant. It would be interesting to find out why *Xf* appears to be more pathogenic in one country/host than another.

Financial Benefits

The financial impact of *Xf* is difficult to accurately estimate due to its large host range. However, the arrival of *Xf* in a country has a significant economic impact in many sectors, as the detection of *Xf* would not only affect farms, but also nurseries, retailers, and importers/exporters. Pierce's disease caused an annual loss of US$ 104 million in California by 2014. Approximately US$ 50 million is also spent on preventative measures every year. In Brazil, 40% of citrus plants are affected by Citrus Variegated Chlorosis, which caused an annual loss of US$ 120 million by 2005. In Europe, the first *Xf* epidemic was identified in Apulia in the South of Italy, where the bacterium was found to be the cause of Olive quick-decline syndrome. 40% of olive trees are grown in Apulia for the production of olive oil in Italy, and over 10 ha of olive trees have since been destroyed.

Action Points

Xf has not yet been detected in the UK. However, remain vigilant of symptoms and report any potential ones. If an outbreak is suspected, contact the APHA Plant Health and Seeds Inspector or PHSI Headquarters for England and Wales (planthealth.info@apha.gsi.gov.uk), the Scottish Government's Horticulture and Marketing Unit (hort.marketing@gov.scot), or the DAERA Plant Health Inspection Branch for Northern Ireland (planthealth@dardni.gov.uk). Be aware of *Xf* disease symptoms, as these can vary between different plants. Visit the EPPO website (https://gd.eppo.int/taxon/XYLEFA/photos) for disease pictures and the European Commission website for an extensive list of susceptible *Xf* plant hosts (https://ec.europa.eu/food/plant/plant_health_biosecurity/legislation/emergency_measures/x

ylella-fastidiosa/susceptible_en). It is also advised to keep up-to-date with plant health news. Most importantly, avoid importation of plants from areas affected by *Xf* (https://www.cabi.org/isc/datasheet/57195#todistributionDatabaseTable and https://www.cabi.org/isc/datasheet/57195#toDistributionMaps) and/or ensure the imported material holds appropriate plant passports and phytosanitary certificates.

SCIENCE SECTION

Introduction

Background

Xylella fastidiosa (*Xf*) is a Gram-negative, rod-shaped bacterium that is most notoriously known to cause olive quick decline syndrome (OQDS) in the South of Italy, citrus variegated chlorosis (CVC) in Brazil, and Pierce's disease (PD) of grapevine in the USA. *Xf* is one of today's most devastating plant pathogens, disrupting international trade and causing huge economic loss for affected countries.

The bacterium is believed to originate from the Americas, where it appears to be a generalist endophyte to native plant species (Hopkins, and Purcell, 2002; Chatterjee, Almeida, and Lindow, 2008). *Xf* spread has been connected with human-mediated movement of infected plants, resulting in distribution of the bacteria across large geographical distances. The first disease associated with the bacterium was detected by Newton B. Pierce in 1892 in the USA. Previously confined to the Americas, the first outbreak of *Xf* in Europe was detected in 2013, largely affecting olive trees in the South of Italy (Saponari, *et al*., 2013). Today, *Xf* outbreaks have been rampant in Italy, France and Spain. In two countries – Netherlands (Bergsma-Vlami, *et al*., 2015) and Belgium (AVBS, 2018) – *Xf* was detected in imported plants that were intercepted at ports, and in another three countries – Switzerland (EPPO, 2015), Germany (EPPO, 2016a), and Portugal (EPPO, 2019) – the bacterium was detected in isolated cases only and is currently under eradication or has since been eradicated.

To date, six *Xf* subspecies have been described. *Xf* subsp. *fastidiosa* originated in Central America, *multiplex* in North America and *pauca* in South America (Sicard, et al., 2018b). The origin of *Xf* subsp. *sandyi* and *morus* are still under debate (Schaad, et al., 2004; Scally, et al., 2005; Nunney, et al., 2014b; Marcelletti, and Scortichini, 2016a). Lastly, subspecies

tashke has only been found in North America. In Europe, only subspecies *fastidiosa*, *multiplex*, *pauca*, *sandyi*, and *morus* have been identified. Therefore, *Xf* subsp. *tashke* has not been included in this study.

Xf is transferred between plants through xylem-sap feeding insects, such as Aphrophoridae (spittlebugs) and Cicadallinae (sharpshooters; Cavalieri and Porcelli, 2017). These vectors remain unaffected by the bacteria. In Europe, *Philaenus spumarius* (meadow spittlebug) appears to be the main vector responsible for most of *Xf* spread (Cornara, et al., 2017; Rapicavoli, et al., 2018). The bacterium has been detected in over 350 different botanical taxa, and dozens of economically significant crops are susceptible to *Xf*. In the United Kingdom, the most significantly cultivated crops include, but are not limited to: alfalfa, bay, blueberry, *Brassica*, *Cercis* (redbuds), *Chionanthus* (fringe tree), *Cytisus* (broom), elderberry, elm, fig, grapevine, *Hedera* (ivy), *Hypericum* (St. John's Wort), magnolia, maple, mulberry, *Nandina domestica* (sacred bamboo), lavender, oak, olive, pear, *Prunus* (e.g. almond, apricot, cherry), *Rubus* (e.g. raspberries, blackberries), *Rosa*, rosemary, strawberry, *Trifolium* (e.g. clover), walnut, and willow. *Xf* poses a major risk to British plant species if the bacteria were to enter our flora. See **Appendix Table A** for an extensive list of *Xf* host plants.

Most plant hosts do not develop disease symptoms when infected by *Xf*. However, they can act as a reservoir for vectors to further spread the bacteria, which poses a threat to those host plants highly susceptible to *Xf* diseases. Due to its long asymptomatic period, *Xf* remains undetected in susceptible crops – often until it is too late. The bacteria could be unknowingly spread to other plants in asymptomatic material across Europe. Over a dozen diseases have been associated with *Xf*-infection (**Table 1**). Often, symptoms of affected hosts resemble nutrient deficiencies, drought stress or infections caused by other pathogens. This complicates the association of a disease with *Xf* and only molecular techniques – such as by amplifying species-specific genes by PCR – can confirm the presence of the bacterium.

Table 1: Diseases caused by *Xylella fastidiosa* (*Xf*). Diseases caused by the bacterium result from the colonisation of the plant's xylem and blocking the flow of water and soluble nutrients. This leads to leaf scorch and stunt in many different plants. The following diseases have been associated with *Xf*.

Leaf scorch	Stunt
almond leaf scorch	alfalfa dwarf
coffee leaf scorch	citrus variegated chlorosis
elm leaf scorch	Lucerne dwarf
mulberry leaf scorch	periwinkle wilt
oak leaf scorch	phony peach disease
oleander leaf scorch	
olive quick-decline syndrome	
pear leaf scorch	
pecan leaf scorch	
Pierce's disease of grapevine	
plum leaf scald	

Plant pathogens release molecules, known as virulence factors, which interact with the host to invade a cell, evade host defence and ensure disease progress. These virulence factors may appear in the form of proteins, carbohydrates and lipids. Effector proteins are a type of virulence factor and play a major role in pathogenicity. Gram-negative bacteria secrete effectors into their surroundings or translocate them into a host cell through secretion systems (SS), of which six types are known to date. *Xf* lacks genes that make up the type 3 secretion system (T3SS; Simpson, *et al*., 2000), one of the most extensively studied secretion systems (Dow, and Daniels, 2000), and instead encodes essential genes that make up components of type 1, 2, 4 and 5 secretion systems (T1SS, T2SS, T4SS and T5SS, respectively) either within the bacterial chromosome (Simpson, et al., 2000; Sluys, et al., 2003) or on plasmids (Rogers and Stenger, 2012). The lack of a T3SS makes *Xf* both interesting and challenging to study as most model plant pathogenic bacteria, e.g. *Pseudomonas syringae* and *Xanthomonas* spp. rely heavily on this secretion system and its effectors.

A number of virulence factors have been identified in *Xf*. Cell wall-degrading enzymes (CWDEs), such as endo-polygalacturonase (endo-PG) in combination with endoglucanase

(EGase) which give *Xf* the ability to digest plant cell wall polymers (Zhang, *et al*., 2015). Lipopolysaccharides (LPS') are a structural component in Gram-negative bacterial cell envelopes. LPS' are a type of pathogen-associated molecular pattern (PAMP), which allow the plant immune system to recognise the presence of a pathogen and induce an immune response. It has been found that *Xf* is able to modify its terminal O-antigen polysaccharide chain of its LPS', allowing the delay of recognition by the plan immune system (Rapicavoli, et al., 2018). A putative CWDE, LipA – a lipase secreted through the T2SS – was also found to be abundantly secreted in PD symptomatic leaves (Nascimento, *et al*., 2016). A number of haemagglutinin and haemagglutinin-like proteins have also been shown to play a major role in biofilm formation, a key virulence factor in *Xf* pathogenesis (De Souza, *et al*., 2003; Guilhabert, and Kirkpatrick, 2005).

There are no treatments for plants infected by *Xf*. Currently in Europe, when a host plant displays symptoms, is found to carry *Xf*, and an outbreak is declared, the host and all neighbouring potential hosts in a 100 m radius are destroyed (Commission Implement Decision (EU) 2015/789). A 5-10 km demarcation order is also implemented, preventing the movement of plant material outside this area and thus greatly affecting a region's economy. Some preventative measures, especially by targeting vectors, have been implemented to reduce the risk of an *Xf* outbreak (Dongiovanni, *et al*., 2018). However, this does not aid those plants already affected by *Xf*. The lack of fundamental knowledge of the molecular biology of *Xf* makes it difficult to truly understand the mode of pathogenicity of the bacterium. Understanding *Xf* molecular biology could help with the development of a targeted treatment plan for infected plants, for example, by directly targeting molecules involved in disease progression.

Objectives

It is important to study the various factors that make this bacterium pathogenic, and by investigating its molecular biology, genetics and community analyses of affected hosts, a better understanding of diseases caused by *Xf* can be gained. The aim of this research project is to understand the different factors that enable *Xf* to become pathogenic and host-specific. Several questions will be considered: why is the bacterium pathogenic in some plants but remains asymptomatic in others? Do effectors play a role in symptomatic versus asymptomatic cases? Are there any effectors that are specific to symptomatic plants only? Does the microbiome play a role in *Xf* pathogenicity? Finding answers to these questions may give us a better understanding as to why *Xf* causes disease in some plants but not in others. In order to develop an effective control measure, or better yet a treatment plan for diseased hosts, research must be conducted to understand how the bacteria cause disease within a plant. Understanding the fundamental biology of this organism can stop the enormous economic and even cultural loss that is caused by the bacterium around the world.

Materials and methods

Genomics

Determining X. fastidiosa host range. An extensive list of documented hosts where *Xf* has been detected was curated. Sources included international organisations (EC, 2018; EFSA, 2018; EPPO, n.d.). Wherever subspecies information was available for a host, a visualisation of the host range of each *Xf* subspecies was created using the R package *VennDiagram* (Chen, and Boutros, 2011).

Phylogeny inference. A total of 55 *Xf* complete and draft genomes and one *Xylella taiwanensis* complete genome was obtained from NCBI's GenBank database (see **Appendix Table B** for details of each genome). *X. taiwanensis* was used as an outgroup for the phylogeny inference. Genomes were annotated with Prokka (Seemann, 2014), and filtered based on N50 statistics and contig number according to a paper published by Levy *et al*. in 2018, and CheckM to remove contaminated and/or incomplete genomes (Parks, *et al*., 2015). The core genome was determined by identifying orthologous sequence groups, descendants of the same ancestral sequence that were separated due to speciation, between the genomes with OrthoFinder (Emms, and Kelly, 2015) using default parameters. Protein sequences were subsequently aligned with ClustalW (Thompson, Higgins, and Gibson, 1994), corrected with GBlocks (Castresana, 2000; Talavera, and Castresana, 2007), and protein models tested using ProtTest (Abascal, Zardoya, and Posada, 2005). Finally, a phylogeny was inferred by maximum-likelihood with IQ-Tree (Nguyen, *et al*., 2015) on concatenated protein sequence alignments of single-copy orthologous groups and visualised using the R package *ape* (Paradis, Claude, and Strimmer, 2004). The detailed pipeline can be found at GitHub (https://github.com/mirloupa/prokka_and_orthofinder).

Multi-locus sequence typing (MLST). Multi-locus sequence typing (MLST) is a method by which bacteria are characterised by the sequence variations in housekeeping genes, which highly conserved sequences in the genome essential for the bacteria to survive (Maiden, *et al*., 1998; Maiden, 2006). In *Xf*, seven housekeeping genes – *leuA, perC, malF, cysG, holC,*

nuoL and *gltT* (see **Appendix Table E** for function and primer sequences of each gene) – have been previously selected for MLST, which is important for the identification of the subspecies of a strain (Scally, *et al.*, 2005). For some *Xf* genomes that were available at NCBI, no subspecies information was provided. Therefore, to identify the sequence type and thus the subspecies of those strains, the seven housekeeping genes of strains of interest were extracted from the genome using NCBI's BLAST, and the sequence type was determined by database search on PUBMLST (https://pubmlst.org/bigsdb?db=pubmlst_xfastidiosa_isolates).

Effector prediction. Putative effectors were identified using the PREFFECTOR programmeme released by Dhroso, Eidson and Korkin (2018). The programme requires protein sequences of interest in FASTA format, which are uploaded to the PREFFECTOR web-server (http://korkinlab.org/preffector). Effector prediction was performed on all coding sequences for 55 *Xf* genomes, one *X. taiwanensis* genome and two *Xanthomonas* genomes.

Survey of *X. fastidiosa* in Colombia

Sampling of plant leaves. Plant leaves were collected from seven different locations within the Antioquia province of Colombia (see **Figure 1**; Kahle, and Wickham, 2013). These included one coffee farm, one citrus farm, two research stations, one location within a rainforest, one university campus and one botanical garden in an urban area. Leaves of 15 different plant species from three families were collected: Malvaceae, Rubiaceae and Rutaceae as *Xf* had previously been detected in several species of these families. See **Appendix Table D** for full details of collected samples. Whenever possible, samples were taken from three plants of each plant species. Of each plant, at least three branches were selected and at least three leaves of each branch were collected (see **Appendix Figure C**) using scissors disinfected with 70% ethanol prior to use. Sufficient leaves were collected per sample plant so that three batches of DNA extractions could be done per sample if needed. Each leaf was surface cleaned with 70% ethanol and air-dried before being placed into a

clear polyethylene bag. This polyethylene bag was placed in an additional two polyethylene bags to prevent contamination and accidental *Xf* spread. All sample bags were stored until shipment to the United Kingdom for processing.

DNA extraction. All Colombian leaf samples were processed in a licensed pathogen laboratory within the National Institute of Agricultural Botany – East Malling Research (NIAB EMR) in Kent. The samples were surface cleaned with 70% ethanol, followed by distilled water and subsequently left to air-dry. Once dried, leaves were cut as only the midrib and basal parts were required for DNA extraction. Cut leaf parts were placed in 2.0mL Eppendorf tubes and frozen in liquid nitrogen and stored in a -20°C freezer until further processing. Total DNA was extracted using a cetyltrimethylammonium bromide (CTAB) method designed by EPPO (EPPO Bulletin, 2016) and modified in this research. The detailed protocol can be found in **Appendix Figure G**. All total DNA extracts are stored at -20°C. The remaining two batches of each sample plant were stored at -80°C for future use.

PCR to detect X. fastidiosa. Three separate PCRs were prepared which are referred to as 16S, XF1 and XF2 hereafter. In the 16S PCR, primers 27F and 1492R targeted the 16S region of a genome to detect the presence of bacteria (Muyzer, De Waal, and Uitterlinden, 1993). The XF1 PCR was a primary mean to determine the presence of *Xf*. In this PCR, primers RST31 and RST33 are *Xf*-specific and target the 3' end of *rpoD*, which encodes an RNA polymerase sigma-70 factor in the bacterium (Minsavage *et al*., 1994). The XF2 PCR, using *Xf*-specific primers 16S-23F and 16S-23R, was a secondary mean and control to confirm the presence of *Xf* in a sample. This targets a 16S-23S intergenic spacer region of the bacterium (Martinati *et al*., 2005). See **Appendix Table E** for complete sequences of each primer pair and PCR conditions for each reaction. 16S PCR was repeated thrice per sample to determine consistency of results. Only samples that showed positive for 16S at least twice were tested for *Xf*. Sigma-Aldrich's *redTaq* polymerase was used for all PCRs.

***Sequencing of positive* samples.** XF1 PCR was repeated on all positive Colombian samples using ThermoFisher Scientific's Platinum *Taq* polymerase, high-fidelity polymerase. Amplicons of samples RUBCA03001, RUBCA03002, RUBCA03003, RUBCA03005,

RUBCA03006, RUBCA03007, RUBCA03008, RUBCA03010, RUBCA03011, RUBCA03013, RUBCA03015, RUBCA05001 and the positive control *Xf* subsp. *fastidiosa* strain Temecula-1 were selected for sequencing to refute contamination. PCR amplicons were purified using the Biolabs Monarch DNA gel extraction kit and Sanger sequenced using the Eurofins LightRun GATC service. Consensus sequences of sequencing data were acquired using DNASTAR's Sanger Sequence Assembly option, using the *rpoD* gene sequence of *Xf* subsp. *fastidiosa* strain 9a5c (downloaded from NCBI) as the reference sequence. Multiple sequence alignment (MSA) by progressive strategy was performed on the consensus sequences using the programme T-Coffee (Notredame, Higgins, and Heringa, 2000). The alignment was finally visualised using JalView (Waterhouse, *et al*., 2009).

Figure 1: Map of collection sites in Colombia. GPS coordinates of all samples were collected during the survey. Using the R package ggmap (Kahle, and Wickham, 2013), the collection sites were mapped to the map of Colombia. Only samples in the Antioquia province of Colombia were collected. See **Appendix Table D** for full details of each collected sample.

Results

Host range of *X. fastidiosa* subspecies

A list of hosts from which *Xf* was isolated was compiled using information collected from the European Commission (EC; 2018), the European Food Safety Authority (EFSA; 2018) and the European and Mediterranean Plant Protection Organisation (EPPO; n.d.). The list includes details such as the *Xf* subspecies that was found in the plant, whether this was found in a natural or experimental setting, in which European country these were found, and diseases found in each plant host (see **Appendix Table A**). From this information, the presence of the subspecies most prevalent in different European countries was identified. The subspecies *fastidiosa* was found in Germany and Spain but is most prevalent in Spain. Subspecies *multiplex* was found in France, Portugal and Spain, and is most prevalent in France. Subspecies *pauca* was found in France, Italy and Spain, and is most prevalent in Italy. The subspecies *sandyi* has only been identified in France.

A Venn diagram was then produced to visualise the number of hosts shared between the four subspecies of interest (Chen, and Boutros, 2011). Information of subspecies isolated from different hosts (see **Appendix Table A**) was used to create this Venn diagram of shared hosts (see **Figure 2**). *Xf* subsp. *multiplex* has the largest host range, with 88 hosts only affected by the subspecies. 48 hosts are affected by *Xf* subsp. *fastidiosa* only, 20 hosts by *pauca* and 3 by *sandyi* alone. Four plant hosts are shared among all four subspecies: *Coffea* sp. (coffee), *Nerium oleander* (oleander), *Polygala myrtifolia* (myrtle-leaf milkwort) and *Prunus dulcis* (almond). No shared plant hosts exist between the following subspecies:

- multiplex vs sandyi
- pauca vs sandyi
- fastidiosa vs multiplex vs sandyi
- fastidiosa vs pauca vs sandyi
- multiplex vs pauca vs sandyi

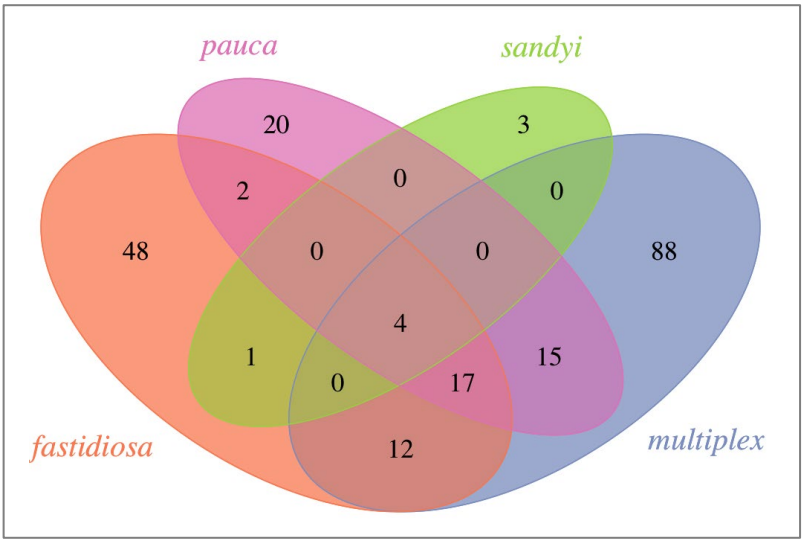

Figure 2: Venn diagram of shared host plants between *Xylella fastidiosa* (*Xf*) subspecies. In total, information of 206 plant hosts wherein *Xf* was detected was collected from EC (2018), EFSA (2018) and EPPO (n.d.). This diagram, created using the R package *VennDiagram* (Chen, and Boutros, 2011), depicts the number of hosts solely found in each subspecies and shared among other subspecies. A detailed list of the different subspecies detected in these plant hosts is found in **Appendix Table A**. Subspecies *multiplex* has the largest host range, whereas *sandyi* has the smallest. Four hosts are shared among all four subspecies: *Coffea* sp. (coffee), *Nerium oleander* (oleander), *Polygala myrtifolia* (myrtle-leaf milkwort) and *Prunus dulcis* (almond).

A phylogenetic tree *X. fastidiosa* strains

A phylogeny using whole genome information of 55 *Xf* and one *X. taiwanensis* was generated (see **Figure 3**). The subspecies information of a number of *Xf* strains was available on GenBank, where the genomes were obtained from. Strains with known subspecies grouped together in the phylogenetic tree, thus allowing the inference of the subspecies of the other strains. The subspecies of these strains were later confirmed as described in the methods section above and included in the phylogeny. Origin information was present for most of the genomes which are coloured according to the continent where the strain was found. All South American strains appear to be of the subspecies *pauca*. Subspecies clearly group together in the phylogeny with the exception of *Xf* subsp. *fastidiosa* strain 6c grouping within the *Xf* subsp. *pauca* clade. Hosts spread across the entire phylogeny with some convergence of different subspecies. For example, the plant family Vitaceae (e.g. grapevine) appears to only be infected by *Xf* subsp. *fastidiosa*, but the plant family Rosaceae is infected by *Xf* subsp.

fastidiosa, *multiplex* and *pauca*. Bootstrap values are predominantly high with five exceptions where bootstrap values are below 70. This is most likely due to the assembly level of the some of the genomes, some of which have only been assembled to the contig level.

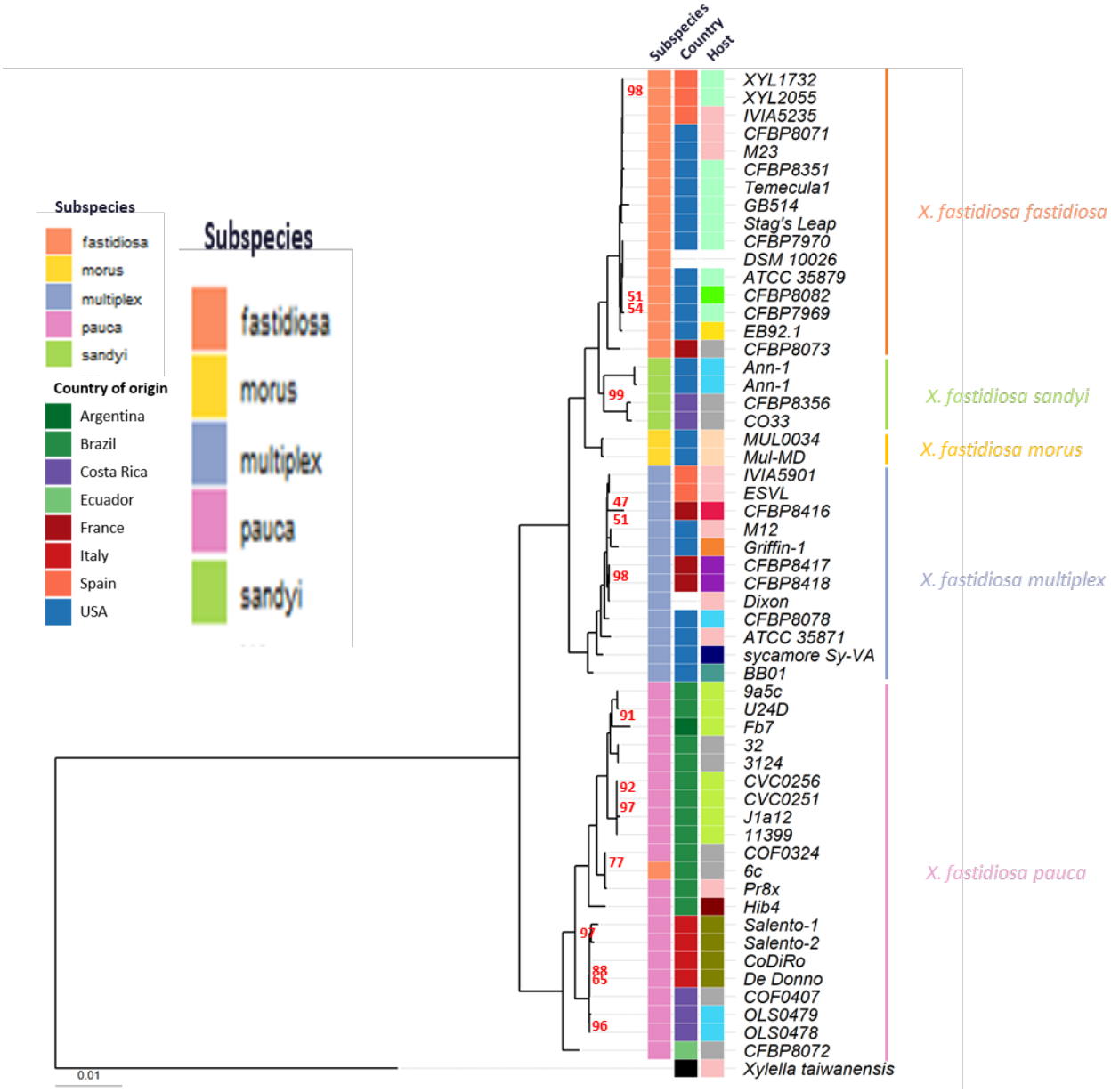

Figure 3: Phylogenetic tree of *Xylella fastidiosa* (*Xf*). A phylogenetic tree of 55 *Xf* and the *Xylella taiwanensis* genome (used outgroup) was created. This tree was generated using IQ-Tree's multiple sequence alignment by maximum-likelihood (Nguyen, et al., 2015). Bootstrap values below 100 are indicated in red. The tree was visualised using the *ape* package on R (Paradis, Claude and Strimmer, 2004). Location and host origin (where available), and subspecies information are highlighted in different colours.

Prediction of effector proteins in *X. fastidiosa*

Amino acid sequences of 55 *Xf* genomes, two *Xanthomonas* genomes and one *X. taiwanensis* genome were acquired from NCBI's GenBank database and uploaded to the PREFFECTOR webserver. As an output, a table was produced for each genome, listing the following information: a database ID generated by PREFFECTOR, a sequence ID identifying the sequence number within the original FASTA input file, the default minimum probability threshold of 0.9, the predicted probability calculated by PREFFECTOR, the effector categorisation, and the original FASTA sequence header of the predicted effector. In total, 3,440 putative effecters were predicted by PREFFECTOR across the 58 genomes of interest. Interestingly, *Xf* strain EB92.1, a strain that appears to be less pathogenic than other *Xf* strains (Hopkins, 1951), has the largest number of predicted effectors (see **Figure 4**).

A first glance of the type or proteins predicted by PREFFECTOR shows that the majority of sequences have not been characterised yet (see **Figure 5**; Fellows, 2012), which is not uncommon as the function of the majority of the genome is unknown. Many predicted effectors of which the sequences have been previously described include various enzymes, transport proteins, membrane proteins, receptors, and haemagglutinins – which have previously been shown to be crucial in biofilm formation (De Souza, *et al*., 2003; Guilhabert, and Kirkpatrick, 2005)

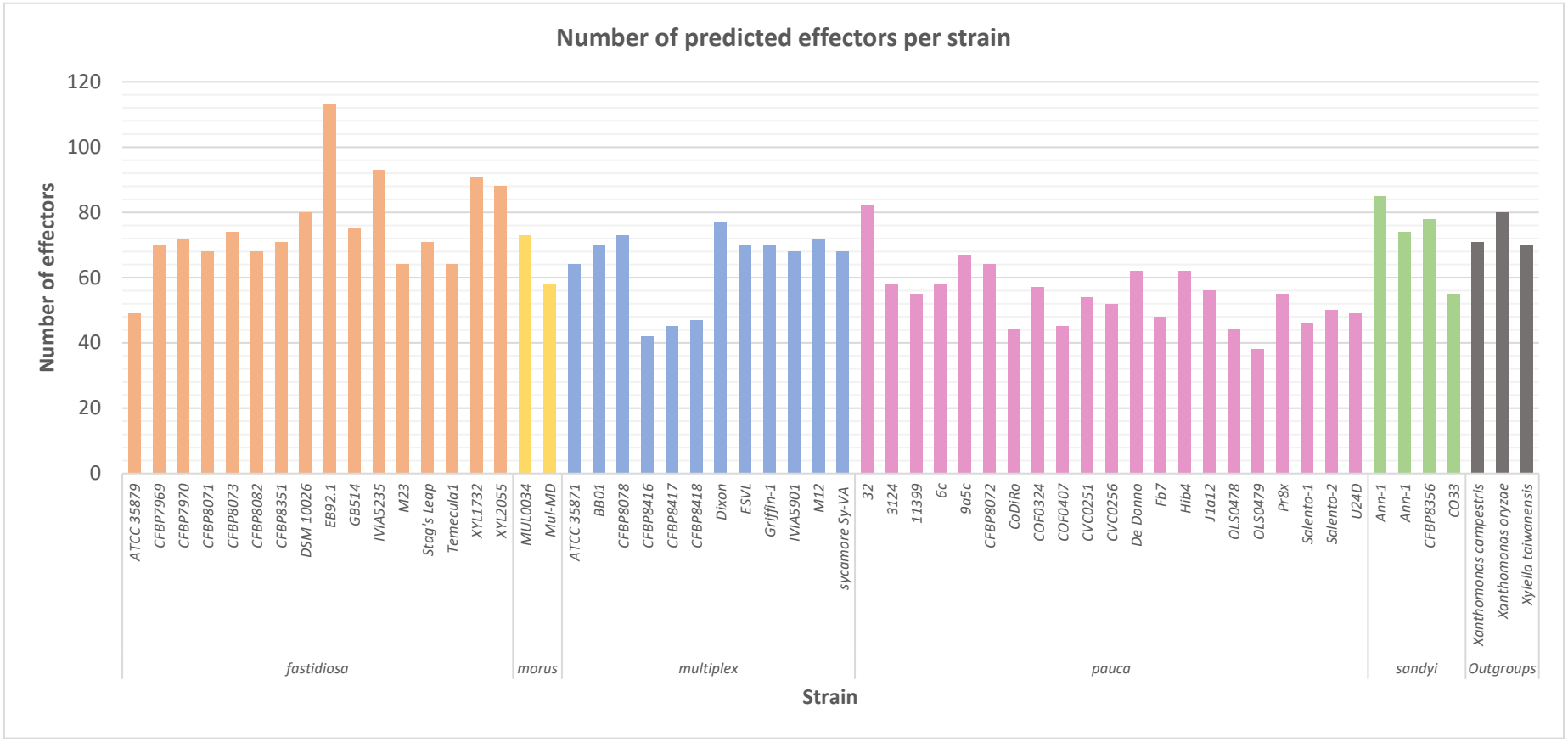

Figure 4: Number of predicted effectors per strain. Effector prediction was performed using the PREFFECTOR software (Dhroso, Eidson and Korkin, 2018), which predicts effectors across all six secretion systems. Protein sequences of 55 *Xylella fastidiosa* (*Xf*) genomes, one *Xylella taiwanensis* genome and two *Xanthomonas* genomes were obtained from NCBI for analysis. A graphical interpretation of the number of effectors predicted per input genome. The colours indicate the different *Xf* subspecies. *Xf* subsp. *fastidiosa* strain EB92.1 which is associated with less pathogenic symptoms, interestingly has the highest number of predicted effectors.

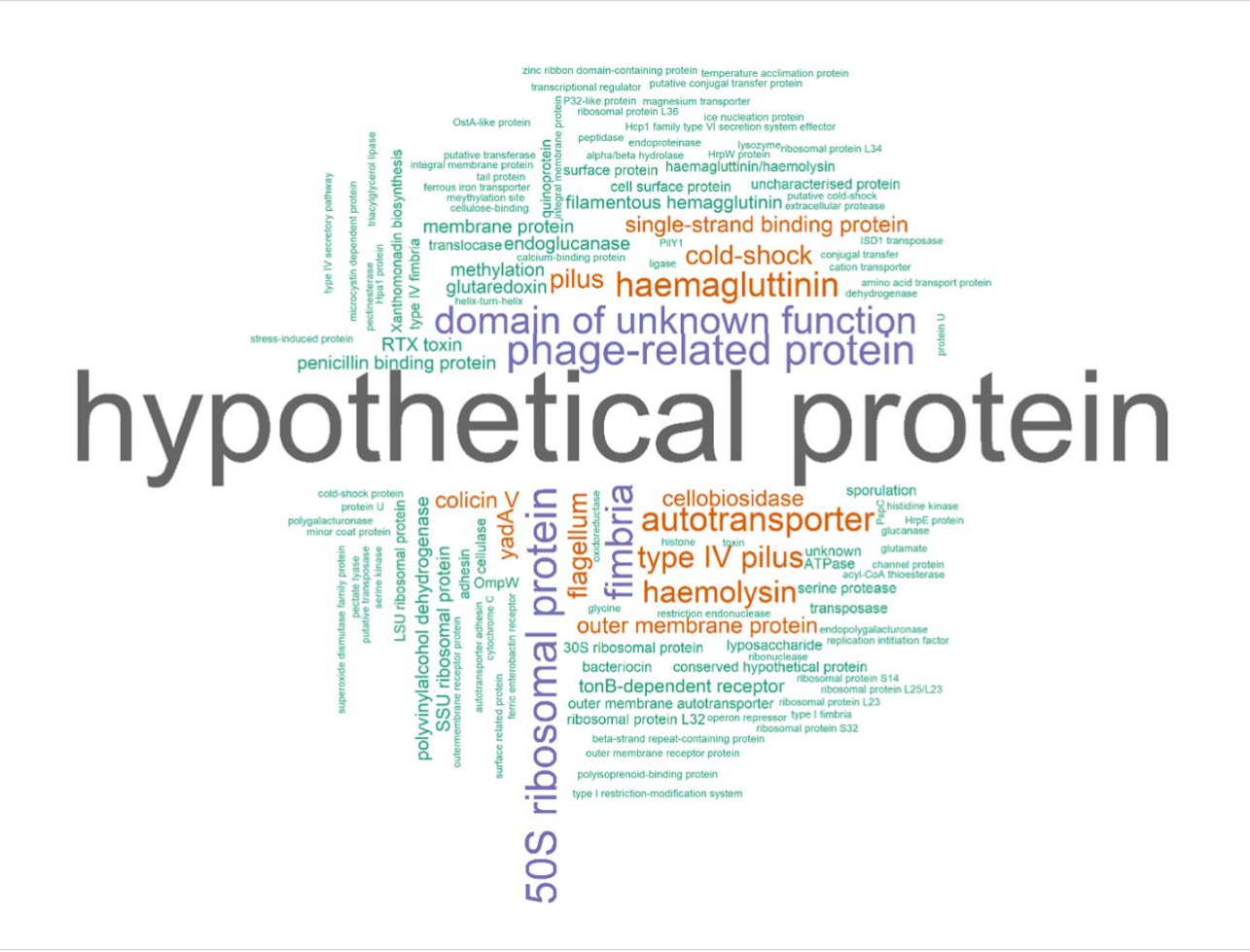

Figure 5: Word cloud of predicted effector sequences. This word cloud shows the resulting predicted effectors from the program PREFFECTOR (Dhroso, Eidson and Korkin, 2018) which predicts effectors across all six secretion systems. The resulting sequences were categorised and finally visually summarised using the R package wordcloud **(**Fellow, 2012). The majority of predicted effector sequences are hypothetical and uncharacterised proteins. Further analyses will be done to predict the functions of these by looking at the motifs, structure and similarity search. The font colours have no meaning and merely serve as visual aid.

Survey of *X. fastidiosa* in Colombia

Xf is believed to originate from the Americas (Hopkins, and Purcell, 2002; Chatterjee, Almeida, and Lindow, 2008). The opportunity to travel and collect plant samples from Colombia, a country where *Xf* has not yet been identified, allowed the first survey to be conducted in the country. Several locations in the Antioquia province of Colombia were chosen to collect leaves of different plant families in which *Xf* had previously been detected. These included Malvaceae (e.g. hibiscus), Rubiaceae (e.g. coffee) and Rutaceae (citrus; see **Appendix Table D** for a full list of collected samples). The goal was to sample as many plants as possible from these families, both in both natural and cultivated environments. The goal was to detect *Xf* in wild plant species, the hypothesis being that *Xf* is an endophytic organism in South American endemic plants. The EPPO standard CTAB DNA extraction protocol (EPPO, 2016b) was optimised for *Coffea* species, as this was one of the main hosts of interest. *Coffea* is an especially interesting *Xf* host, as every subspecies of interest – *fastidiosa, multiplex, pauca* and *sandyi* – have been identified in this plant.

EPPO has published a standard protocol for the extraction of total DNA from plant leaves for subsequent identification of *Xf* in a sample by molecular methods, which was by PCR in this research. Here, the CTAB-based DNA extraction protocol was modified in order to get high concentrations of DNA from *C. arabica* samples. All leaves were surface sterilised with 70% ethanol, washed in distilled water and air-dried to avoid the DNA extraction of epiphytes, microbes living on the surfaces of plants. The leaves freeze-dried and ground with a pestle and mortar instead of a mechanical homogeniser (EPPO, 2016b). Pre-heated CTAB buffer was added after grinding. Due to the high concentrations of RNA in the first trials of the EPPO standard protocol, RNAse A was added after the CTAB step and incubated overnight to allow the RNAse to digest the RNA in the sample. For the precipitation of DNA, room temperature 2-propanol instead of cold 2-propanol, as suggested in the standard protocol, was added to prevent excess salts of being precipitated with the DNA and thus get better concentrations.

Lastly, each sample was washed three times in 70% ethanol ensure all contaminants are removed from the sample.

A total of 51 plant samples were collected in triplicates during the Colombian survey. Thirteen samples collected in Colombia tested positive for *Xf* (see **Table 2**). Twelve of these were collected from a single coffee farm in Fredonia and one from the EAFIT University campus in Medellín, both of which are located in the Antioquia province of the country. All positive samples originated from *Coffea arabica* plants. No *Xf* was identified using the XF1 PCR protocol in any of the Malvaceae and Rutaceae samples. From the coffee farm, plants of three positive samples did not display any *Xf*-specific symptoms. These were samples RUBCA03001 and RUBCA03002 (*C. arabica* cv. Geisha); and RUBCA03005 (*C. arabica* cv. Colombia). The positive sample collected from the EAFIT University, RUBCA05001, was asymptomatic for *Xf*, but was affected by coffee rust, a fungal disease caused by *Hemileia vastatrix*. The cultivar of this plant is unknown. The remaining nine samples that tested positive for *Xf* originated from the same coffee farm and displayed leaf scorch symptoms similar to *Xf*-affected *C. arabica* plants found in Brazil and Costa Rica. These included samples RUBCA03003 (*C. arabica* cv. Geisha); RUBCA03006 (*C. arabica* cv. Colombia); RUBCA03007, RUBCA03008 (*C. arabica* cv. Caturra); RUBCA03010, RUBCA03011, RUBCA03012 (*C. arabica* cv. Pajarito); RUBCA03013 and RUBCA03015 (*C. arabica* cv. Castillo). All thirteen positive samples were tested by PCR using *Xf*-specific primers targeting two different regions in the genome: XF1 PCR amplified the 3' end of *rpoD*, a gene encoding an RNA polymerase sigma-70 factor (see **Figure 4**; Minsavage *et al*., 1994), whereas XF2 PCR amplified the 16S-23S intergenic spacer region (Martinati *et al*., 2005).

Amplicons of XF1 PCR of twelve positive samples – RUBCA03001, RUBCA03002, RUBCA03003, RUBCA03005, RUBCA03006, RUBCA03007, RUBCA03008, RUBCA03010, RUBCA03011, RUBCA03013, RUBCA03015 and RUBCA05001 – and positive control *Xf* subsp. *fastidiosa* strain Temecula-1 were selected for initial sequencing. However, only eight

samples – RUBCA03001, RUBCA03003, RUBCA03005, RUBCA03006, RUBCA03007, RUBCA03011, RUBCA0315 and RUBCA05001 – plus the positive control were returned with sufficient quality sequencing data. Amplicon sequences of RUBCA03002, RUBCA03008, RUBCA03010, and RUBCA03013 were only partially sequenced and were too short when consensus sequences were acquired. An MSA of the positive samples and the positive control reveals nucleotide differences in some sites of the sequences (**Appendix Figure H**). This confirms that the positive samples were not contaminated with the positive *Xf* control.

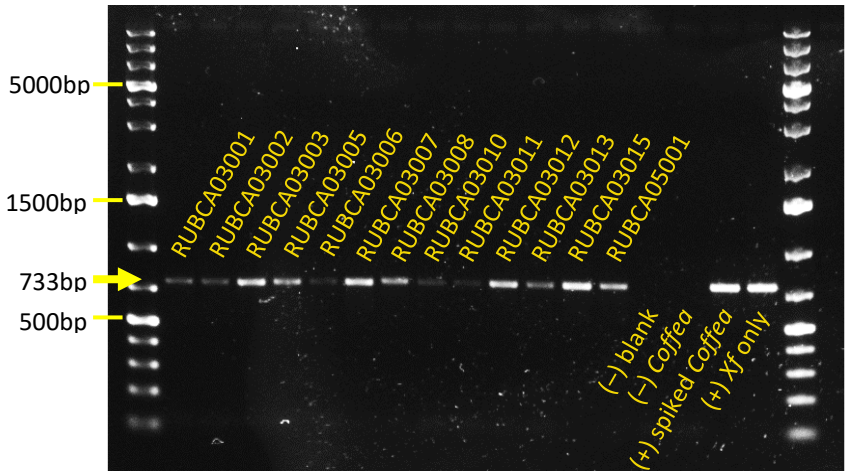

Figure 6: Gel image of XF1 PCR of all positive Colombian samples. A PCR targeting a *Xylella fastidiosa* (*Xf*) specific gene was performed on all Colombian samples. This gel depicts all samples where *Xf* was identified, which comprise of several *Coffea arabica* cultivars from a single farm, and one *C. arabica* plant from a university campus. The resulting amplicon is 733bp long. Two negative controls (one blank and one *C. arabica* total DNA extract) and two positive controls (one *C. arabica* extract spiked with *Xf* DNA and one pure *Xf* DNA sample) were included.

Table 3: A list of all Colombian samples that tested positive for *Xylella fastidiosa* (*Xf*). Different parameters were measured during the collection of leave samples in Colombia. Below are details of the samples that tested positive for *Xf* by PCR. All samples underwent two *Xf*-specific PCRs amplifying different regions in the genome: XF1 PCR amplified the 3' end of rpoD, a gene encoding an RNA polymerase sigma-70 factor (Minsavage et al., 1994), and XF2 PCR amplified the 16S-23S intergenic spacer region (Martinati et al., 2005). Four samples that tested positive for *Xf* were collected from plants that did not display *Xf*-specific symptoms. However, one of these (RUBCA05001) was affected by coffee rust. The remaining nine samples that tested positive for *Xf* originated from plants that displayed leaf scorch, a symptom that has been observed in *Xf*-affected *Coffea arabica* plants in Brazil and Costa Rica.

ID	Date	Time	Family	Species	Cultivar	Symptoms	location	Location	MAMSL	GPS (dd)	Temp (°C)	Humidity (%)
RUBCA03001	20190627	15:25	Rubiaceae	*Coffea arabica*	Geisha	A	Coffee farm	Fredonia	1423	5.970375, -75.670041	24	59
RUBCA03002	20190627	15:30	Rubiaceae	*Coffea arabica*	Geisha	A	Coffee farm	Fredonia	1423	5.9703, -75.6701	24	59
RUBCA03003	20190627	15:45	Rubiaceae	*Coffea arabica*	Geisha	S	Coffee farm	Fredonia	1423	5.9704, -75.6704	24	59
RUBCA03005	20190627	16:07	Rubiaceae	*Coffea arabica*	Colombia	A	Coffee farm	Fredonia	1423	5.9730, -75.6700	24	59
RUBCA03006	20190627	16:12	Rubiaceae	*Coffea arabica*	Colombia	S	Coffee farm	Fredonia	1423	5.9730, -75.6701	24	59
RUBCA03007	20190627	16:42	Rubiaceae	*Coffea arabica*	Caturra	S	Coffee farm	Fredonia	1786	5.99748, -75.6644	24	59
RUBCA03008	20190627	16:46	Rubiaceae	*Coffea arabica*	Caturra	S	Coffee farm	Fredonia	1786	5.9749, -75.6643	24	59
RUBCA03010	20190627	16:54	Rubiaceae	*Coffea arabica*	Pajarito	S	Coffee farm	Fredonia	1786	5.9748, -75.6644	24	59
RUBCA03011	20190627	16:59	Rubiaceae	*Coffea arabica*	Pajarito	S	Coffee farm	Fredonia	1786	5.9747, -75.6644	24	59
RUBCA03012	20190627	17:07	Rubiaceae	*Coffea arabica*	Pajarito	S	Coffee farm	Fredonia	1786	5.9746, -75.6643	24	59
RUBCA03013	20190627	17:10	Rubiaceae	*Coffea arabica*	Castillo	S	Coffee farm	Fredonia	1786	5.9748, -75.6645	24	59
RUBCA03015	20190627	17:20	Rubiaceae	*Coffea arabica*	Castillo	S	Coffee farm	Fredonia	1786	5.9740, -75.6645	24	59
RUBCA05001	20190703	11:15	Rubiaceae	*Coffea arabica*	N/A	A	University campus	Medellín	1504m	6.2002, -75.5785	23	64

Discussion

Host range of *X. fastidiosa*

A comprehensive literature search was conducted to assess the host-range of *Xf*. It was found that the different subspecies of interest share some hosts, but can also be very host-specific. However, sampling bias – i.e. plants with *Xf*-symptoms are more likely to be tested for the bacterium – as well as a concentration too small for any molecular test to detect any bacteria could be limiting the knowledge of the true host range of *Xf*.

As the subspecies *sandyi* was only established in 2005 by Schuenzel, *et al*., it might explain why only a limited number of hosts have been associated with the *sandyi*. A complete list of hosts affected by each subspecies is found in **Appendix Table A**. As the list of plants affected by *Xf* is incomplete, it is difficult to know whether multiplex really does have the largest host range, or whether this is just due to sampling bias. Also, the majority of plants *Xf* has been isolated from are crops and ornamentals. Very limited research is done on *Xf* found on native plants, therefore it is very likely that *Xf* is present in such plants but just has not been detected yet.

X. fastidiosa phylogeny

Phylogenies are a helpful way to understand the relationships between different strains of bacteria and how they might have diverged. However, the high instances of recombination between bacterial strains make it difficult to find a 'true' phylogenetic tree of a bacterial species. To create bacterial phylogenetic trees, one must look at the core genome instead, as these are usually more conserved between strains. This was done by implementing the OrthoFinder programme on available *Xf* genomes (Emms, and Kelly, 2015). OrthoFinder finds orthologous genes, which are sequences that are descendants of the same ancestral sequence that were separated due to speciation, between the genomes of interest. Creating a phylogeny of all *Xf* genomes currently available and mapping the hosts where each strain

was isolated from allows the visualisation of any possible patterns of host specificity within and between subspecies.

All South American strains appear to be of the subspecies *pauca*. This supports research that shows the *Xf* CoDiRo strain (Marcelletti, and Scortichini, 2016b), associated with the first European outbreak of OQDS in Italy, belongs to the subspecies *pauca*, as the strain clusters in the same group. Unfortunately, genomes of only two strains of the subspecies *sandyi* were available. This is interesting, as *sandyi* is also the subspecies with the most limited host range. Further research is needed to determine whether this limited host range is a result of the sporadic instances of *sandyi* in plants of interest, sampling bias, or whether *sandyi* is more prevalent in plants as a generalist endophyte. As very limited research is available on generalist microbes, it might not be clear if the majority of subspecies *sandyi* strains are actually non-disease-causing. The heterogeneous location of origin (North America and Europe) of subspecies *fastidiosa* and *multiplex* strains implies that European strains were introduced from North America. Subspecies information for *Xf* strains MUL-MD and MUL0034 were not available. Nunney, *et al*. (2014) have proposed the subspecies *morus* for strain MUL0034, however this novel subspecies is still under review. Further research is required to validate this or determine if the two strains belong to other subspecies, as they do not clearly group in any of the clades in this phylogeny.

From the curation of *Xf* hosts it is known that Rubiaceae is a plant family affected by all four *Xf* subspecies of interest, however the phylogeny does not show this. This is because there is no genome of *Xf* subsp. *multiplex* affecting Rubiaceae available. Unfortunately in the sciences, it is often the case to be working with incomplete data. This uncovers many questions: How many more plant hosts does *Xf* have? Is host-specificity between subspecies even more blurred than previously thought? In other words, how many more plant hosts are out there that can be affected by all four subspecies? A phylogeny will not be able to answer those questions, but it does provide a good visualisation of the signatures of host-specificity for *Xf*. For example, this phylogeny will be of great support when analysing putative effectors of *Xf* to determine if there are any host-specific or subspecies-specific effectors.

Putative effectors of *X. fastidiosa*

There are several methods by which bacterial effectors can be predicted. However, most of the available programmes focus on effectors secreted by the T3SS. An issue with effector prediction is that the majority of proteins have not been characterised and the function of most proteins is unknown. Instead, one could look at protein motifs, the structure or detect similarity with proteins in other bacteria to find out the function of a similar protein in that bacterial strain. Over 3,000 effectors have been predicted using the PREFFECTOR software (Dhroso, Eidson, and Korkin, 2018) and the analysis of these results are still being carried out. Moreover, predicted effectors will be mapped to an up-to-date phylogeny to determine possible patterns across subspecies and/or hosts.

Presence of *X. fastidiosa* in Colombia

Xf is a familiar plant pathogen in the Americas. In Central and South America in particular, *Xf* is known to cause disease symptoms in citrus, coffee and *Prunus* spp. In Brazil, *Xf* is especially devastating as it is known to be the cause of CVC, a disease resulting in smaller and lower quality fruits, directly impacting the country's economy. In South America, *Xf* is also known to affect *C. arabica*, where it is known to cause leaf scorch symptoms. *C. arabica* is a particularly interesting host as all four subspecies of interest – *fastidiosa, multiplex, pauca* and *sandyi* – have been detected in the plant (EFSA, 2018). Colombia is known for its high-quality coffee production. Interestingly, in many nearby countries, *Xf* has been detected in *C. arabica*, e.g. Venezuela, Brazil, Paraguay and Puerto Rico. However, no report of either the presence, nor absence of *Xf* in Colombia is available, and this is interesting because of Colombia's trade in coffee. In this research, the first *Xf* is detected for the first time in a *C. arabica* farm in Colombia. Samples of five different cultivars of *C. arabica* of a coffee farm in Fredonia, which lies in the Antioquia province, have been collected and *Xf* was detected in

plants of all five cultivars. XF1 PCR was repeated on all thirteen positive Colombian samples and the positive control *Xf* subsp. *fastidiosa* strain Temecula-1 with Platinum *Taq* polymerase, a high-fidelity polymerase. High-fidelity polymerases provide better specificity during the replication process in PCR. Only amplicons of twelve samples – RUBCA03001, RUBCA03002, RUBCA03003, RUBCA03005, RUBCA03006, RUBCA03007, RUBCA03008, RUBCA03010, RUBCA03011, RUBCA03013, RUBCA03015 and RUBCA05001 – and the positive control were selected for sequencing. The PCR for RUBCA03012 did not give any amplicons and was therefore omitted. The twelve samples and control were sent for Sanger sequencing, however, only eight samples – RUBCA03001, RUBCA03003, RUBCA03005, RUBCA03006, RUBCA03007, RUBCA03011, RUBCA0315 and RUBCA05001 – and the positive control were returned with sufficient quality sequencing data. Amplicon sequences of RUBCA03002, RUBCA03008, RUBCA03010, and RUBCA03013 were only partially sequenced and were too short when consensus sequences were acquired. MSA of consensus sequences of all samples and the positive control show several differences between the sequences (**Appendix Figure H**). This confirmed that the positive amplification of the Colombian samples were not in fact contamination from the positive control *Xf* subsp. *fastidiosa* strain Temecula-1, which was used throughout the PCR process. An initial BLAST on NCBI of the XF1 amplicons suggest that the two sequenced samples RUBCA03005 and RUBCA05001 are subspecies *pauca*, which may indicate a relation with *Xf* coffee strains in Brazil or Costa Rica. Interestingly, one coffee plant that was found to harbour *Xf* in Costa Rica was of the cultivar Caturra (Rodríguez, *et al*., 2001). This cultivar was also sampled in Colombia and *Xf* was detected in two plants with weak *Xf*-like symptoms. MLST will be performed on all positive samples to identify the subspecies.

Samples from different Malvaceae and Rutaceae have also been collected, but no *Xf* could be detected in those samples. However, this could be false negatives, as the detection of *Xf* by PCR can be very limiting. *Xf* might not have been detected in these samples due to a too low of a concentration of bacteria in the sample, and the PCR not being powerful enough to detect these concentrations.

Conclusions

Even though *Xf* was the first plant pathogen to be sequenced (Simpson, *et al*., 2000), there are still many aspects in its genome that need to be further explored. This study is attempting to understand the role of effector proteins in the pathogenicity and host-range of the bacterium. More specifically, this study attempts to determine if certain effectors are subspecies-specific, host-specific and/or only expressed in pathogenic strains. A number of putative effectors appear to have very promising links to *Xf* virulence and further analyses are required. It would also be interesting to investigate the expression of effectors in symptomatic and asymptomatic hosts of *Xf*. Furthermore, other factors, such as community dynamics will be explored to determine whether these play a role in *Xf* virulence (see **Table 4** for detailed future plans). Lastly, the first detection of *Xf* in *C. arabica* in Colombia shows that the biogeography of the bacteria is still underexplored. How many more countries harbour *Xf* with no reports of outbreaks? How do the Colombian *Xf* strains differ from the strains currently known? What impact will the discovery of *Xf* in Colombia have to the country? These are questions that will be investigated further.

Table 4: Future plan for this research project. This list details a number of analyses to be explored in this project and papers planned to be published in the near future.

	Analysis	**Time allocation (mm/yy)**	**Notes**
Genomics	Try different programmes to predict bacterial effector proteins	03/20 – 05/20	Several programmes and databases are available to identify effector proteins. These programmes will be applied to *Xf* genomes and those resulting predicted effector proteins that are shared among the different programmes will be further used in this project. If a protein sequence is predicted to be an effector by multiple programmes, it is more likely that this protein is in fact an effector and not just a false positive.
	Identify small secreted non-annotated proteins	03/20 – 04/20	A paper published by Shindo, *et al*. (2016) explores effector proteins that share no homology with annotated proteins.
	Characterisation of putative effector proteins	02/20 – 06/20	Predicted effector proteins of interested that have not been characterised yet will be explored by looking at proteins of similar structure, comparison searches and databases of protein motifs.
	Create pipeline of genetic gain and loss of effectors across *Xf* strains	02/20 – 06/20	A pipeline of the genetic gain and loss across *Xf* strains will be established to be applied to all predicted effectors of interest later in the project. This will be in the form of a heatmap of the presence/absence of predicted effector proteins will be added to a phylogeny. This could be in an interactive manner, such as by using R shiny, to create a software so the user can apply different thresholds themselves.
	Explore expressed predicted effector proteins in available RNAseq data	06/20 – 09/20	The expression of predicted effectors of interest will be explored by comparing sequences with publicly available RNAseq data of pathogenic *Xf* strains.
	Metagenomic analysis of asymptomatic vs symptomatic Colombian strains	06/20 – 10/20	Scientific skills to be acquired during an EMBL metagenomics course in April 2020.
	Analysis of neighbouring sequences of effectors	09/20 – 11/20	Explore the neighbouring sequences of predicted effectors to determine if any transposable elements are present and whether these are shared across any of the *Xf* strains.
	Comparative genomics of Colombian vs European *Xf* strains	12/20 – 03/21	Determine presence/absence of effector sequences in Colombian strains and compare with European strains.
	Identify pathogenicity islands	12/20 – 02/21	It would be interesting to determine whether any of the predicted effector proteins of interest are encoded in pathogenicity islands, if these pathogenicity islands have

			neighbouring transposable elements, and if these islands of sequences are shared across different pathogenic *Xf* strains.
	Create an up-to-date phylogeny	01/21	There are now 91 available *Xf* genomes on NCBI and many more are expected to become publicly available. A more recent phylogeny will be created using the same pipeline as above.
Colombian samples	MLST of positive Colombia samples	02/20 – 04/20	MLST of the 13 positive Colombian samples following Yuan, *et al.*'s (2010) sequence typing of *Xf* to identify the subspecies of each strain.
	Isolation and sequencing of Colombian *Xf* strain(s)	04/20 – 12/20	To be done at a laboratory where *Xf* has previously been isolated (e.g. Fera Science Ltd., collaborators in Italy or Spain).
Papers	First report of *Xf* in Colombia	03/20 – 05/20	To include MLST of positive Colombian samples.
	Genome(s) of Colombian *Xf* strain(s)	02/21 – 05/21	To include genome of Colombian strain(s) and a comparative analysis of Colombian vs European *Xf* sequences

Knowledge and Technology Transfer

Table 5. List of attended knowledge and technology transfer events.

Date	Event	Activity
02/2020	NIAB EMR Seminars East Malling, UK	seminar
02/2020	NIAB EMR PhD student meeting East Malling, UK	poster presentation
02/2020	The Linnean Society Student Conference 2020 London, UK	oral presentation
01/2020	AHDB Crop PhD Conference 2020 Nottingham, UK	poster presentation
11/2019	AHDB Soft Fruit Day 2019 NIAB EMR, East Malling, UK	poster presentation
10/2019	2nd European Conference on Xylella fastidiosa Ajaccio, France	poster presentation
10/2019	National Fruit Show 2019 Maidstone, UK	'Bacterial Diseases' co-exhibitor
10/2019	University of Nottingham Doctoral Training Programme student visit NIAB EMR, East Malling, UK	oral presentation
10/2019	The Worshipful Company of Gardeners' Association visit NIAB EMR, East Malling, UK	oral presentation
07/2019	Tropical Microbiology Course 2019 EAFIT University, Medellín, Colombia	seminar
06/2019	Soapbox Science 2019 Canterbury, UK	oral presentation
05/2019	Biosecurity and Xylella training RHS Garden Wisley UK	training
05/2019	AHDB industry visit and meeting with growers J&A Growers, Warwick, UK	industry visit
03/2019	Weekly Genetics, Genomics & Breeding department meeting NIAB EMR, East Malling, UK	oral presentation
03/2019	MBPP conference 2019 JIC Conference Centre, Norwich, UK	poster presentation
03/2019	NIAB Poster Day 2019 NIAB, Cambridge, UK	poster presentation
03/2019	Monthly PhD student meeting NIAB EMR, East Malling, UK	oral presentation
02/2019	Weekly Genetics, Genomics & Breeding department meeting NIAB EMR, East Malling, UK	oral presentation
11/2018	AHDB PhD Studentship Conference 2018 Solihul, UK	oral presentation
11/2018	Genetics, Genomics and Breeding Department Research Symposium 2018 Maidstone, UK	oral presentation

Glossary

AVBS	[Belgian Nurserymen and Growers' Federation]
BLAST	basic local alignment search tool
bp	base pair(s)
CTAB	cetyltrimethylammonium (cetrimonium) bromide
cv.	cultivar
CVC	citrus variegated chlorosis
CWDE	cell-wall degrading enzyme
EC	European Commission
EFSA	European Food Safety Authority
EGase	endoglucanase
endo-PG	endo-polygalacturonase
EPPO	European and Mediterranean Plant Protection Organisation
EPS	extracellular polymer substance(s)
LPS	lipopolysaccharide
MAMSL	metres above median sea level
MSA	multiple sequence alignment
NCBI	National Centre for Biotechnology Information
n.d.	no date
OLS	oak leaf scorch
OQDS	olive quick-decline syndrome
PAMP	pathogen-associated molecular pattern
PCR	polymerase chain reaction
PD	Pierce’s disease
sp. / spp.	species (singular / plural)
subsp.	subspecies
T[1-6]SS	type [1-6] secretion system
Xf	*Xylella fastidiosa*

References

Abascal, F, Zardoya, R, and Posada, D, 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 21(9), pp.2104–2105.

Almeida, RPP, and Nunney, L, 2015. How do plant diseases caused by Xylella fastidiosa emerge? Plant Disease. [online] 99(11), pp.1457–1467. Available at: <http://dx.doi.org/10.1094/PDIS-02-15-0159-FE>.

AVBS, 2018. First discovery of Xylella fastidiosa in Belgium. [online] Available at: <https://www.avbs.be/actualiteit/eerste-detectie-van-xylella-fastidiosa-belgië>.

Bergsma-Vlami, M, van de Bilt, JLJM, Tjou-Tam-Sin, NNA, van de Vossenberg, BTLH, and Westenberg, M, 2015. Imported From Costa Rica and Honduras in the Netherlands Xylella Fastidiosa in Coffea. Journal of Plant Pathology. 97(2), pp.391–403.

Buttner, D, and He, SY, 2009. Type III Protein Secretion in Plant Pathogenic Bacteria. Plant Physiology. [online] 150(4), pp.1656–1664. Available at: <http://www.plantphysiol.org/cgi/doi/10.1104/pp.109.139089>.

Castresana, J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution. 17(4), pp.540–552.

Cavalieri, V, and Porcelli, F, 2017. Main insect vectors of Xylella fastidiosa worldwide & in Italy. In: A.M. D'Onghia, S. Brunel and F. Valentini, eds., *Xylella fastidiosa & the Olive Quick Decline Syndrome (OQDS). A serious worldwide challenge for the safeguard of olive trees.* Bari: CIHEAM, pp.31–32.

Chatterjee, S, Almeida, RPP, and Lindow, S, 2008. Living in two Worlds: The Plant and Insect Lifestyles of Xylella fastidiosa. Annual Review of Phytopathology. 46, pp.243–271.

Chen, H, and Boutros, PC, 2011. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 12, p.35.

Cornara, D, Cavalieri, V, Dongiovanni, C, Altamura, G, Palmisano, F, Bosco, D, Porcelli, F, Almeida, RPP, and Saponari, M, 2017. Transmission of Xylella fastidiosa by naturally infected Philaenus spumarius (Hemiptera, Aphrophoridae) to different host plants. Journal of Applied Entomology. 141(1–2), pp.80–87.

Dhroso, A, Eidson, S, and Korkin, D, 2018. Genome-wide prediction of bacterial effector candidates across six secretion system types using a feature-based statistical framework. Scientific reports. [online] 8(1). Available at: <http://dx.doi.org/10.1038/s41598-018-33874-1>.

Didelot, X, Lawson, D, Darling, A, and Falush, D, 2010. Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*.

Dongiovanni, C, Di Carolo, M, Fumarola, G, Tauro, D, Altamura, G, and Cavalieri, V, 2018. Evaluation of Insecticides for the Control of Juveniles of Philaenus spumarius L., 2015–2017. Arthropod Management Tests. 43(1), pp.1–2.

Dow, JM, and Daniels, MJ, 2000. Xylella Genomics and Bacterial Pathogenicity to Plants. Yeast. 1(4), pp.263–271.

Emms, DM, and Kelly, S, 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. [online] 16(157). Available at: <https://genomebiology.biomedcentral.com/track/pdf/10.1186/s13059-015-0721-2>.

EPPO, 2015. Xylella fastidiosa detected in Coffea spp. plants imported into Switzerland. [online] Available at: <https://gd.eppo.int/reporting/article-5128>.

EPPO, 2016a. First report of Xylella fastidiosa subsp. fastidiosa on Nerium oleander in Germany. [online] Available at: <https://gd.eppo.int/reporting/article-5878>.

EPPO, 2016b. PM 7/24 (2) Xylella fastidiosa. In: *EPPO Bulletin*. [online] pp.463–500. Available at: <https://doi.org/10.1111/epp.12327>.

EPPO, 2019. First report of Xylella fastidiosa subsp. multiplex in Portugal. [online] Available at: <https://gd.eppo.int/reporting/article-6447>.

Feil, H, and Purcell, AH, 2007. Temperature-Dependent Growth and Survival of Xylella fastidiosa in Vitro and in Potted Grapevines . Plant Disease. 85(12), pp.1230–1234.

Fellows, I, 2012. Wordcloud: Word clouds. *R package version*.

Guilhabert, MR, and Kirkpatrick, BC, 2005. Identification of Xylella fastidiosa Antivirulence Genes: Hemagglutinin Adhesins Contribute to X. fastidiosa Biofilm Maturation and Colonization and Attenuate Virulence. Molecular Plant-Microbe Interactions. 181, pp.856–868.

Henneberger, TS, 2003. Effects of Low Temperature on Populations of Xylella fastidiosa in Sycamore. [online] University of Georgia. Available at: <https://getd.libs.uga.edu/pdfs/henneberger_tiffany_s_200308_ms.pdf>.

Hopkins, DL, 1951. Biological Control of Pierce' s Disease in the Vineyard with Strains of Xylella fastidiosa Benign to Grapevine. Plant Disease. [online] 89(12), pp.1348–1352. Available at: <http://apsjournals.apsnet.org/doi/full/10.1094/PHYTO-07-18-0245-FI?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3Dpubmed>.

Hopkins, DL, and Purcell, AH, 2002. Xylella fastidiosa : Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. *Plant Disease*.

Janse, JD, and Obradovic, A, 2010. Xylella fastidiosa: its biology, diagnosis, control and risks. Journal of Plant Pathology. 92, p.S1.35-S1.48.

Kahle, D, and Wickham, H, 2013. ggmap: Spatial visualization with ggplot2. *R Journal*.

Levy, A, Salas Gonzalez, I, Mittelviefhaus, M, Clingenpeel, S, Herrera Paredes, S, Miao, J, Wang, K, Devescovi, G, Stillman, K, Monteiro, F, Rangel Alvarez, B, Lundberg, DS, Lu, TY, Lebeis, S, Jin, Z, Mcdonald, M, Klein, AP, Feltcher, ME, Rio, TG, Grant, SR, Doty, SL, Ley, RE, Zhao, B, Venturi, V, Pelletier, DA, Vorholt, JA, Tringe, SG,

Woyke, T, Dangl, JL, Gonzalez, IS, Mittelviefhaus, M, Clingenpeel, S, Paredes, SH, Miao, J, Wang, K, Devescovi, G, Stillman, K, Monteiro, F, Alvarez, BR, Lundberg, DS, Lu, TY, Lebeis, S, Jin, Z, Mcdonald, M, Klein, AP, Feltcher, ME, Rio, TG, Grant, SR, Doty, SL, Ley, RE, Zhao, B, and Venturi, V, 2018. Genomic features of bacterial adaptation to plants. Nature Genetics. [online] 50, pp.138–150. Available at: <http://dx.doi.org/10.1038/s41588-017-0012-9>.
Lindow, S, 2019. Money Matters: Fueling Rapid Recent Insight Into *Xylella fastidiosa*— An Important and Expanding Global Pathogen. Phytopathology. [online] p.PHYTO-09-18-032. Available at: <https://apsjournals.apsnet.org/doi/10.1094/PHYTO-09-18-0325-PER>.
Maiden, MCJ, 2006. Multilocus Sequence Typing of Bacteria. *Annual Review of Microbiology*.
Maiden, MCJ, Bygraves, JA, Feil, E, Morelli, G, Russell, JE, Urwin, R, Zhang, Q, Zhou, J, Zurth, K, Caugant, DA, Feavers, IM, Achtman, M, and Spratt, BG, 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Sciences of the United States of America. 95(6), pp.3140–3145.
Marcelletti, S, and Scortichini, M, 2016a. Genome-wide comparison and taxonomic relatedness of multiple Xylella fastidiosa strains reveal the occurrence of three subspecies and a new Xylella species. Archives of Microbiology. [online] 198(8), pp.803–812. Available at: <https://dx.doi.org/10.1007/s00203-016-1245-1>.
Marcelletti, S, and Scortichini, M, 2016b. Xylella fastidiosa CoDIRO strain associated with the olive quick decline syndrome in southern Italy belongs to a clonal complex of the subspecies pauca that evolved in Central America. Microbiology (United Kingdom). 162(12), pp.2087–2098.
Martelli, GP, Boscia, D, Porcelli, F, and Saponari, M, 2016. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. European Journal of Plant Pathology. [online] 144(2), pp.235–243. Available at: <https://link.springer.com/article/10.1007/s10658-015-0784-7>.
Martinati, JC, Hansen Pacheco, FT, Oliveira De Miranda, VF, and Siu, MT, 2005. Phylogenetic relationships of Xylella fastidiosa strains based on 16s-23s rDNA sequences. Current Microbiology. 50(4), pp.190–195.
Meyer, MM, and Kirkpatrick, BC, 2008. Examining the Effects of Cold Therapy on Pierce ' s Disease-infected Grapevines and on the Viability of Xylella fastidiosa Cells in vitro. In: *Proceedings of the 2nd Annual National Viticulture Research Conference*.
Minsavage, G V, Thompson, CM, Hopkins, DL, Leite, RMVBC, and Stall, RE, 1994. Development of a Polymerase Chain Reaction Protocol for Detection of Xylella fastidiosa in Plant Tissue. Phytopathology. 84, pp.456–461.
Muyzer, G, De Waal, EC, and Uitterlinden, AG, 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology. 59(3), pp.695–700.
Nascimento, R, Gouran, H, Chakraborty, S, Gillespie, HW, Almeida-Souza, HO, Tu, A, Rao, BJ, Feldstein, PA, Bruening, G, Goulart, LR, and Dandekar, AM, 2016. The Type II Secreted Lipase/Esterase LesA is a Key Virulence Factor Required for Xylella fastidiosa Pathogenesis in Grapevines. Scientific Reports. [online] 6. Available at: <http://dx.doi.org/10.1038/srep18598>.
Nguyen, LT, Schmidt, HA, Von Haeseler, A, and Minh, BQ, 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 32(1), pp.268–274.
Notredame, C, Higgins, DG, and Heringa, J, 2000. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*.
Nunney, L, Ortiz, B, Russell, SA, Sánchez, RR, Stouthamer, R, Nunney, L, Ortiz, B, Russell, SA, and Sa, RR, 2014a. The complex biogeography of the plant pathogen xylella fastidiosa: Genetic evidence of introductions and subspecific introgression in central America. PLoS ONE. [online] 9(11). Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112463>.
Nunney, L, Schuenzel, EL, Scally, M, Bromley, RE, and Stouthamerc, R, 2014b. Large-scale intersubspecific recombination in the plant-pathogenic bacterium xylella fastidiosa is associated with the host shift to mulberry. Applied and Environmental Microbiology. [online] 80(10), pp.3025–3033. Available at: <http://aem.asm.org/cgi/pmidlookup?view=long&pmid=24610840>.
Paradis, E, Claude, J, and Strimmer, K, 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20(2), pp.289–290.
Parks, DH, Imelfort, M, Skennerton, CT, Hugenholtz, P, and Tyson, GW, 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 25(7), pp.1043–1055.
Pierce, N.B. 1892. The California vine disease: a preliminary report of investigations. US Government Printing Office, Washington D.C.
Randall, JJ, Goldberg, NP, Kemp, JD, Radionenko, M, French, JM, Olsen, MW, and Hanson, SF, 2009. Genetic analysis of a novel Xylella fastidiosa subspecies found in the Southwestern United States. *Applied and Environmental Microbiology*.
Rapicavoli, J, Ingel, B, Blanco-Ulate, B, Cantu, D, and Roper, C, 2018. Xylella fastidiosa: an examination of a re-emerging plant pathogen. Molecular Plant Pathology. 19(4), pp.786–800.
Rodríguez, CM, Obando, JJ, Villalobos, W, Moreira, L, and Rivera, C, 2001. First Report of Xylella fastidiosa Infecting Coffee in Costa Rica. Plant Disease. 85(9), p.1027.
Rogers, EE, and Stenger, DC, 2012. A Conjugative 38 kB Plasmid Is Present in Multiple Subspecies of Xylella fastidiosa. PLoS ONE. [online] 7(12). Available at: <http://dx.plos.org/10.1371/journal.pone.0052131>.
Saponari, M, Boscia, D, Nigro, F, and Martelli, GP, 2013. Identification of DNA sequences related to Xylella fastidiosa in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia. Journal of Plant Pathology. 95(3), p.668.

Scally, M, Schuenzel, EL, Stouthamer, R, and Nunney, L, 2005. Multilocus sequence type system for the plant pathogen Xylella fastidiosa and relative contributions of recombination and point mutation to clonal diversity. Applied and Environmental Microbiology. 71(12), pp.8491–8499.
Schaad, NW, Postnikova, E, Lacy, G, Fatmi, M, and Chang, CJ, 2004. Xylella fastidiosa subspecies: X. fastidiosa subsp piercei, subsp. nov., X. fastidiosa subsp. multiplex subsp. nov., and X. fastidiosa subsp. pauca subsp. nov. Systematic and Applied Microbiology. 27(3), pp.290–300.
Schuenzel, EL, Scally, M, Stouthamer, R, and Nunney, L, 2005. A Multigene Phylogenetic Study of Clonal Diversity and Divergence in North American Strains of the Plant Pathogen Xylella fastidiosa. Applied and Environmental Microbiology. [online] 71(7), pp.3832–3839. Available at: <https://aem.asm.org/content/71/7/3832>.
Seemann, T, 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 30(14), pp.2068–269.
Shindo, T, Kaschani, F, Yang, F, Kovács, J, Tian, F, Kourelis, J, Hong, TN, Colby, T, Shabab, M, Chawla, R, Kumari, S, Ilyas, M, Hörger, AC, Alfano, JR, and van der Hoorn, RAL, 2016. Screen of Non-annotated Small Secreted Proteins of Pseudomonas syringae Reveals a Virulence Factor That Inhibits Tomato Immune Proteases. PLoS Pathogens. 12(9), pp.1–24.
Sicard, A, Zeilinger, AR, Vanhove, M, Schartel, TE, Beal, DJ, Daugherty, MP, and Almeida, RP, 2018a. Xylella fastidiosa: Insights into an Emerging Plant Pathogen. Annual Review of Phytopathology. (June), pp.1–22.
Sicard, A, Zeilinger, AR, Vanhove, M, Schartel, TE, Beal, DJ, Daugherty, MP, and Almeida, RP, 2018b. Xylella fastidiosa: Insights into an Emerging Plant Pathogen. Annual Review of Phytopathology. (56), pp.181–202.
Simpson, AJG, Reinach, FC, Arruda, P, Abreu, FA, Acencio, M, Alvarenga, R, Alves, LMC, Araya, JE, Baia, GS, Baptista, CS, Barros, MH, Bonaccorsi, ED, Bordin, S, Bové, JM, Briones, MRS, Bueno, MRP, Camargo, AA, Camargo, LEA, Carraro, DM, Carrer, H, Colauto, NB, Colombo, C, Costa, FF, Costa, MCR, Costa-Neto, CM, Coutinho, LL, Cristofani, M, Dias-Neto, E, Docena, C, El-Dorry, H, Facincani, AP, Ferreira, AJS, Ferreira, VCA, Ferro, JA, Fraga, JS, França, SC, Franco, MC, Frohme, M, Furlan, LR, Garnier, M, Goldman, GHS, Goldman, MHS, Gomes, SL, Gruber, A, Ho, PL, Hoheisel, JD, Junqueira, ML, Kemper, EL, Kitajima, JP, Krieger, JE, Kuramae, EE, Laigret, F, Lambais, MR, Leite, LCC, Lemos, EGM, Lemos, MVF, Lopes, SA, Lopes, CR, Machado, JA, Machado, MA, Madeira, AMBN, Madeira, HMF, Marino, CL, Marques, M V, Martins, EAL, Martins, EMF, Matsukuma, AY, Menck, CFM, Miracca, EC, Miyaki, CY, Monteiro-Vitorello, CB, Moon, DH, Nagai, MA, Nascimento, ALTO, Netto, LES, Nhani, A, Nobrega, FG, Nunes, LR, Oliveira, MA, de Oliveira, MC, de Oliveira, RC, Palmieri, DA, Paris, A, Peixoto, BR, Pereira, GAG, Pereira, HA, Pesquero, JB, Quaggio, RB, Roberto, PG, Rodrigues, V, Rosa, AJM, de Rosa, VE, de Sá, RG, Santelli, R V, Sawasaki, HE, da Silva, ACR, da Silva, AM, da Silva, FR, Silva, WA, da Silveira, JF, Silvestri, MLZ, Siqueira, WJ, de Souza, AA, de Souza, AP, Terenzi, MF, Truffi, D, Tsai, SM, Tsuhako, MH, Vallada, H, Van Sluys, MA, Verjovski-Almeida, S, Vettore, AL, Zago, MA, Zatz, M, Meidanis, J, and Setubal, JC, 2000. The genome sequence of the plant pathogen Xylella fastidiosa: The Xylella fastidiosa consortium of the organization for nucleotide sequencing and analysis, Sao Paulo, Brazil. Nature. 406(6792), pp.151–157.
Sluys, MA Van, Oliveira, MC De, Miyaki, CY, Furlan, LR, Camargo, LEA, Silva, ACR, Moon, DH, Takita, MA, Lemos, EGM, Machado, MA, Ferro, MIT, Silva, FR, Goldman, MHS, Goldman, GH, Lemos, MVF, Tsai, SM, Carrer, H, Carraro, DM, Oliveira, RC De, Nunes, LR, Siqueira, WJ, Coutinho, LL, Kimura, ET, Ferro, ES, Harakava, R, Kuramae, EE, Marino, CL, Giglioti, E, Abreu, IL, Alves, LMC, Amaral, AM, Baia, GS, Blanco, SR, Brito, MS, Cannavan, FS, Celestino, A V, Cunha, AF, Fenille, RC, Ferro, JA, Formighieri, EF, Kishi, LT, Leoni, SG, Oliveira, AR, Jr, VER, Sassaki, FT, Sena, JAD, Souza, AA De, Truffi, D, Tsukumo, F, Yanai, GM, Zaros, LG, Civerolo, EL, Simpson, AJG, Jr, NFA, Setubal, JC, and Kitajima, JP, 2003. Comparative Analyses of the Complete Genome Sequences of Pierce ' s Disease and Citrus Variegated Chlorosis Strains of Xylella fastidiosa. Journal of Bacteriology. [online] 185(3), pp.1018–1026. Available at: <http://jb.asm.org/cgi/pmidlookup?view=long&pmid=12533478>.
De Souza, AA, Takita, MA, Coletta-Filho, HD, Caldana, C, Goldman, GH, Yanai, GM, Muto, NH, De Oliveira, RC, Nunes, LR, and Machado, MA, 2003. Analysis of gene expression in two growth states of Xylella fastidiosa and its relationship with pathogenicity. Molecular Plant-Microbe Interactions. 16(10), pp.867–875.
Strona, G, Carstens, CJ, and Beck, PSA, 2017. Network analysis reveals why Xylella fastidiosa will persist in Europe. Scientific Reports. [online] 7(1). Available at: <http://dx.doi.org/10.1038/s41598-017-00077-z>.
Talavera, G, and Castresana, J, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology. 56(4), pp.564–577.
Thompson, JD, Higgins, DG, and Gibson, TJ, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 22(22), pp.4673–4680.
Waterhouse, AM, Procter, JB, Martin, DMA, Clamp, M, and Barton, GJ, 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 25(9), pp.1189–1191.
Yuan, X, Morano, L, Bromley, R, Spring-pearson, S, Stouthamer, R, and Nunney, L, 2010. Multilocus sequence typing of Xylella fastidiosa causing Pierce's disease and oleander leaf scorch in the United States. Phytopathology. 100(6), pp.601–611.
Zhang, S, Chakrabarty, PK, Fleites, LA, Rayside, PA, Hopkins, DL, and Gabriel, DW, 2015. Three new pierce's disease pathogenicity effectors identified using xylella fastidiosa biocontrol strain EB92-1. PLoS ONE. [online] 10(7), pp.1–17. Available at: <http://dx.doi.org/10.1371/journal.pone.0133796>.

Appendices

Additional Figures

Additional Tables

Table A: List of *Xylella fastidiosa* (*Xf*) host plants. A list of host plants wherein *Xf* was detected was compiled using data from EC (2018), EFSA (2018) and EPPO (n.d.). The list includes the *Xf* subspecies found in each host plant (if available). N, *Xf* detected in a natural setting; E, *Xf* detected in an experimental setting; U, no information available in which setting *Xf* was detected. The list also includes information whether *Xf* was found in the European countries France, Spain, Germany, Italy and Portugal. *Xf* was also detected in olive in Belgium, but no subspecies information has yet been published. No information could be found of the presence of host plants in Europe of rows highlighted in orange.

Hosts	Common names	*fastidiosa*	*multiplex*	*pauca*	*sandyi*	France	Spain	Germany	Italy	Belgium	Portugal	Disease	Reference
Acacia dealbata	silver wattle, blue wattle, mimosa		N			*multiplex*							EC (2018), EFSA (2018)
Acacia saligna	coojong, golden wreath wattle, orange wattle, blue-leafed wattle, Western Australian golden wattle		N	N		*multiplex*	*pauca*		*pauca*				EC (2018), EFSA (2018)
Acacia sp.			N	N			*multiplex, pauca*						EFSA (2018)
Acer griseum			N										EFSA (2018)
Acer platanoides			N										EFSA (2018)
Acer pseudoplatanus	sycamore		N			*multiplex*							EC (2018), EFSA (2018)
Acer rubrum			EN										EFSA (2018)
Acer sp.		N											EFSA (2018)
Alnus rhombifolia			N										EFSA (2018)
Amaranthus blitoides		E											EFSA (2018)
Ambrosia acanthicarpa		E											EFSA (2018)
Ambrosia psilostachya			N										EFSA (2018)
Ambrosia psilostachya var. texana			N										EFSA (2018)
Ambrosia trifida			N			*multiplex*							EFSA (2018)
Ampelopsis cordata			N										EFSA (2018)
Anthyllis hermanniae	Maltese yellow kindey vetch, Maltese shrubby kidney vetch		N			*multiplex*							EC (2018), EFSA (2018)

Artemisia arborescens	tree wormwood		N		*multiplex*			EC (2018), EFSA (2018)
Asparagus acutifolius	wild asparagus		N	N	*multiplex*		*pauca*	EC (2018), EFSA (2018)
Calicotome spinosa	thorn broom	N	U	U		*fastidiosa*		EC (2018), EFSA (2018)
Calicotome villosa	hairy thorny broom		N		*multiplex*			EC (2018), EFSA (2018)
Carya illinoinensis			EN					EFSA (2018)
Carya sp.			N					EFSA (2018)
Catharanthus roseus		E		EN			*pauca*	EFSA (2018)
Catharanthus sp.	periwinkles			U			*pauca*	EC (2018)
Celtis occidentalis			N					EFSA (2018)
Cercis canadensis			N					EFSA (2018)
Cercis occidentalis		N	N					EFSA (2018)
Cercis siliquastrum	Judas tree	N	N		*multiplex*			EC (2018), EFSA (2018)
Chenopdoium album	fat hen, lamb's quarters, melde, goosefoot (weed)			N			*pauca*	EC (2018), EFSA (2018)
Chenopodium quinoa		E						EFSA (2018)
Chionanthus sp.			N					EFSA (2018)
Cistus albidus	white leaved rock rose, grey-leaved cistus	U	U	U				EC (2018)
Cistus creticus	Cretan rock rose, pink rock-rose, hoary rock-rose		N	N	*multiplex*		*pauca*	EC (2018), EFSA (2018)
Cistus monspeliensis	Montpellier cistus	N	N		*multiplex*	*fastidiosa*		EC (2018), EFSA (2018)
Cistus salviifolius	sage-leaved rock-rose, salvia cistus, Gallipoli rose		U		*multiplex*			EC (2018), EFSA (2018)
Cistus sp.			N	N	*multiplex*			EFSA (2018)
Citroncirus sp.								EPPO (n.d.)
Citrus sp.				EN				EFSA (2018)

Citrus x sinensis	sweet orange	N	N	EN		*multiplex*	citrus-variegated chlorosis (CVC)	EFSA (2018)
Coffea arabica		N		N				EFSA (2018)
Coffea canephora		N			N			EFSA (2018)
Coffea sp.	coffee	U	U	N	N		coffee leaf scorch (CLS)	EC (2018), EFSA (2018)
Conium maculatum		E						EFSA (2018)
Convolvulus cneorum	shrubby bindweed, silverbush		U					EC (2018)
Convulus arvensis		E						EFSA (2018)
Coronilla glauca	scorpion vetch, shrubby scorpion-vetch		U					EC (2018)
Coronilla valentina	bastard senna, shrubby scorpion-vetch, scorpion vetch		N			*multiplex*		EC (2018), EFSA (2018)
Coronilla valentina ssp. glauca			N			*multiplex*		EFSA (2018)
Cyperaceae sp.								EPPO (n.d.)
Cyperus esculentus		E						EFSA (2018)
Cytisus racemosus			N					DEFRA (2016)
Cytisus scoparius	common broom, Scotch broom		N			*multiplex*		EC (2018), EFSA (2018)
Cytisus sp.			N			*multiplex*		EFSA (2018)
Cytisus villosus	hairy broom		N			*multiplex*		EC (2018), EFSA (2018)
Datura wrightii		E						EFSA (2018)
Dendranthema x grandiflorum		E						EFSA (2018)

Dodonaea viscosa	hopbush		N			*pauca*	EC (2018), EFSA (2018)
Echinochloa crus-galli		E					EFSA (2018)
Encelia farinosa			N				EFSA (2018)
Eremophila maculata	spotted fuchsia-bush, spotted emu bush		N			*pauca*	EC (2018), EFSA (2018)
Erigeron bonariensis	hairy fleabane, flax-leaf fleabane, wavy-leaf fleabane, Argentine fleabane (weed)		N			*pauca*	EC (2018), EFSA (2018)
Erigeron canadensis		E					EFSA (2018)
Erigeron sumatrensis	Guernsey fleabane (weed)		N			*pauca*	EC (2018), EFSA (2018)
Eriochloa graciis		E					EFSA (2018)
Erodium moschatum		E					EFSA (2018)
Erysimum hybrids		N					EFSA (2018)
Erysimum sp.	wallflower	U			*fastidiosa*		EC (2018)
Eucalyptus camaldulensis		E					EFSA (2018)
Eucalyptus globulus		E					EFSA (2018)
Euphorbia terracina	false caper, coastal spurge, Geraldton carnation weed		N			*pauca*	EC (2018), EFSA (2018)
Euryops chrysanthemoides	African bush daisy, bull's-eye		N	*multiplex*			EC (2018), EFSA (2018)
Fallopia japonica		N					EFSA (2018)
Ficus carica	common fig		N		*multiplex*		EC (2018), EFSA (2018)
Fortunella sp.							EPPO (n.d.)
Fraxinus americana			N				EFSA (2018)
Fraxinus angustifolia	narrow-leafed ash		N		*multiplex*		EC (2018), EFSA (2018)

Fraxinus sp.			N					EFSA (2018)
Genista corsica	broom		N		*multiplex*			EC (2018), EFSA (2018)
Genista ephedroides	broom		N		*multiplex*			EC (2018), EFSA (2018)
Genista lucida	broom	N	U	U		*fastidiosa*		EC (2018), EFSA (2018)
Genista sp.			N		*multiplex*			EFSA (2018)
Genista x spachiana (syn. Cytisus racemosus Broom)	sweet broom		N		*multiplex*			EC (2018), EFSA (2018)
Ginkgo biloba			N					EFSA (2018)
Gleditsia triacanthos			N					EFSA (2018)
Grevillea juniperina	juniper-leaf grevillea, juniper grevillea, prickly spider-flower		U	N			*pauca*	EC (2018), EFSA (2018)
Hebe sp.	shrubby veronica		N	N	*multiplex*		*pauca*	EC (2018)
Helianthus annuus		E	N					EFSA (2018)
Helianthus sp.			N					EFSA (2018)
Helichrysum italicum	curry plant, Italian strawflower, immortelle		N		*multiplex*			EC (2018), EFSA (2018)
Helicrysum stoechas	shrubby everlasting	U	U	U				EC (2018)
Heliotropium europaeum	common heliotrope, European heliotrope, European turn-sole			N			*pauca*	EC (2018), EFSA (2018)
Hemerocallis sp.					N			EFSA (2018)
Hibiscus rosa-sinensis			N					EFSA (2018)
Ipomoea purpurea		E						EFSA (2018)
Iva annua			N					EFSA (2018)
Jacaranda mimosifolia					N			EFSA (2018)

Juglans regia	common walnut, Persian walnut, English walnut, Circassian walnut	N				*fastidiosa*					EC (2018), EFSA (2018)
Koelreuteria bipinnata			N								EFSA (2018)
Lactuca serriola		E									EFSA (2018)
Lagerstroemia indica			N								EFSA (2018)
Lagerstroemia sp.			N								EFSA (2018)
Laurus nobilis	bay, bay laurel, sweet bay, true laurel, Grecian laurel, laurel tree, laurel		U	N			*pauca*				EC (2018), EFSA (2018)
Lavandula angustifolia	English lavender, lavender, true lavender		N	N	*multiplex*		*pauca*				EC (2018), EFSA (2018)
Lavandula dentata	French lavender, fringed lavender	U	N	N	*multiplex*	*multiplex, pauca*	*pauca*	*multiplex*			EC (2018), EFSA (2018)
Lavandula sp.			N		*multiplex*				asymptomatic		EFSA (2018)
Lavandula stoechas	French lavender, Spanish lavender, topped lavender		N	N	*multiplex*	unknown	*pauca*				EC (2018), EFSA (2018)
Lavandula x allardii (syn. Lavandula x heterophylla)	Allards lavender		U		*multiplex*						EC (2018)
Lavandula x chaytoriae	velvet lavender, Sawyers, lavender 'Sawyers'	U	U	U							EC (2018)
Lavandula x heterophylla			N		*multiplex*						EFSA (2018)
Lavandula x intermedia	fat lavender, hybrid lavender		N		*multiplex*						EC (2018), EFSA (2018)
Liquidambar styraciflua			EN								EFSA (2018)
Liriodendron tulipifera			N								EFSA (2018)

Lonicera japonica	Japanese honeysuckle, golden-and-silver honeysuckle		U								EC (2018)
Lupinus aridorum		N									EFSA (2018)
Lupinus villosus			N								EFSA (2018)
Magnolia grandiflora		N									EFSA (2018)
Malva parviflora		E									EFSA (2018)
Medicago sativa	alfalfa, lucerne	EN	N			*multiplex*				lucerne dwarf	EC (2018), EFSA (2018)
Metrosideros excelsa	pōhutukawa, New Zealand pohutukawa, New Zealand Christmas tree, New Zealand Christmas bush, iron tree		N			*multiplex*					EC (2018), EFSA (2018)
Metrosideros sp.		N									EFSA (2018)
Morus alba											EPPO (n.d.)
Morus rubra											EPPO (n.d.)
Myoporum insulare	blueberry tree, common boobialla, native juniper			N					*pauca*		EFSA (2018)
Myrtus communis	common myrtle		N	N		*multiplex*			*pauca*		EC (2018), EFSA (2018)
Nerium oleander	oleander	N	N	EN	EN		unknown	*fastidiosa*	*pauca*	oleander leaf scorch (OLS)	EC (2018), EFSA (2018)
Nicotiana clevelandii				E							EFSA (2018)
Nicotiana glauca		E									EFSA (2018)
Nicotiana tabacum		E	E	E							EFSA (2018)
Olea europaea	olive	E	EN	EN			*multiplex, pauca*		*pauca*	olive-quick-decline syndrome (OQDS)	EC (2018), EFSA (2018)

Olea europaea ssp. sylvestris	wild olive		N	N			*multiplex, pauca*			olive-quick-decline syndrome (OQDS)	EFSA (2018)
Olea sp.				N				*pauca*	NA		EFSA (2018)
Pelargonium graveolens	sweet scented geranium, rose geranium, old fashion rose geranium, rose-scent geranium		N			*multiplex*					EC (2018)
Pelargonium sp.			N			*multiplex*					EFSA (2018)
Pelargonium x fragrans	nutmeg pelargonium			N				*pauca*			EFSA (2018)
Persea americana											EPPO (n.d.)
Phagnalon saxatile			N			*multiplex*					EC (2018), EFSA (2018)
Phillyrea latifolia	green olive tree, mock privet			N				*pauca*			EC (2018), EFSA (2018)
Platanus occidentalis			EN								EFSA (2018)
Pluchea odorata		N									EFSA (2018)
Polygala moleracea		E									EFSA (2018)
Polygala myrtifolia	myrtle-leaf milkwort	N	EN	EN	N	*multiplex, pauca, sandyi*	*fastidiosa, multiplex, pauca*	*pauca*			EC (2018), EFSA (2018)
Polygala sp.			N			*multiplex*					EFSA (2018)
Polygala x dalmaisiana			N			*multiplex*					EFSA (2018)
Polygala x grandiflora nana			N			*multiplex*					EFSA (2018)
Portulaca oleracea		E									EFSA (2018)
Prunus angustifolia											EPPO (n.d.)
Prunus armeniaca			N								EFSA (2018)

Prunus avium	wild cherry, sweet cherry, gean	N	N	EN		*multiplex*	*fastidiosa*	*pauca*		EC (2018), EFSA (2018)
Prunus cerasifera	cherry plum, myrobalan plum		EN			*multiplex*				EC (2018), EFSA (2018)
Prunus cerasus	morello cherry, sour cherry, tart cherry, dwarf cherry		N							EC (2018), EFSA (2018)
Prunus domestica	common plum		N	EN			*multiplex*		plum leaf scald (PLS)	EFSA (2018)
Prunus dulcis	almond	EN	EN	EN	E	*multiplex, pauca*	*fastidiosa, multiplex, pauca*	*pauca*	almond leaf scorch (ALS)	EC (2018), EFSA (2018)
Prunus persica x P. Webbii		E	E							EFSA (2018)
*Prunus persica**	peach	N	N	EN		*pauca*			phony peach disease (PPD)	EFSA (2018)
Prunus salicina				E						EFSA (2018)
Prunus sp.		E	EN							EFSA (2018)
Prunus x amygdalo-persica				E						EFSA (2018)
Quercus coccinea			N							EFSA (2018)
Quercus falcata			N							EFSA (2018)
*Quercus ilex**	holm oak			EN		*pauca*				EFSA (2018)
Quercus laevis			N							EFSA (2018)
Quercus macrocarpa			N							EFSA (2018)
Quercus nigra			N							EFSA (2018)
Quercus palustris			N							EFSA (2018)
Quercus phellos			N							EFSA (2018)
Quercus pubescens				E						EFSA (2018)

Quercus robur			N						EFSA (2018)
Quercus rubra			N						EFSA (2018)
Quercus shumardii			N						EFSA (2018)
Quercus sp.			N						EFSA (2018)
Quercus suber	cork oak		N		*multiplex*				EC (2018), EFSA (2018)
Ratibida columnifera			N						EFSA (2018)
Rhamnus alaternus	Italian buckthorn, Mediterranean buckthorn	N	N	N		*fastidiosa, multiplex*		*pauca*	EC (2018), EFSA (2018)
Rosa floribunda	dog rose		N						DEFRA (2016)
Rosa canina			N		*multiplex*				EC (2018), EFSA (2018)
Rosa hybrids									EPPO (n.d.)
Rosa multiflora									EPPO (n.d.)
Rosa sp.			N						EFSA (2018)
Rosmarinus officinalis	rosemary	N	N	N	*multiplex*	*multiplex*	*fastidiosa*	*pauca*	EC (2018), EFSA (2018)
Rubus sp.			N						EFSA (2018)
Rubus ursinus		E	E						EFSA (2018)
Rumex crispus		E							EFSA (2018)
Salvia mellifera			N						EFSA (2018)
Sambucus canadensis		N							EFSA (2018)
Sambucus sp.		N	N						EFSA (2018)
Sapindus saponaria			N						EFSA (2018)
Simmondsia chinensis		E							EFSA (2018)
Solanum lycopersicum		E							EFSA (2018)
Solanum melongena		E							EFSA (2018)
Solidago virgaurea			N						EFSA (2018)
Sonchus oleraceus		E							EFSA (2018)

Sorghum halepense		E						EFSA (2018)
Spartium junceum	Spanish broom weaver's broom	N	N	N	*multiplex*		*pauca*	EC (2018), EFSA (2018)
Spartium sp.			N		*multiplex*			EFSA (2018)
Streptocarpus hybrids		N						EFSA (2018)
Streptocarpus sp.	Cape primrose	U				*fastidiosa*		EC (2018)
Teucrium capitatum	cat-thyme germander, felty germander	U	U	U				EC (2018)
Ulmus americana			N					EFSA (2018)
Ulmus crassifolia		N						EFSA (2018)
Vaccinium corymbosum		E	EN					EFSA (2018)
Vaccinium corymbosum x V. angustifolium hybrid			E					EFSA (2018)
Vaccinium sp.		E	EN					EFSA (2018)
Vaccinium virgatum								EPPO (n.d.)
Veronica elliptica	shore hebe, speedwell	U	U	U				EC (2018)
Vicia faba		E						EFSA (2018)
Vicia sativa		E						EFSA (2018)
Vinca major				E				EFSA (2018)
Vinca minor				N			*pauca*	EFSA (2018)
Vinca sp.	periwinkle		N	N			*pauca*	EC (2018), EFSA (2018)
Vitis aestivalis		N						EFSA (2018)
Vitis aestivalis hybrid		N						EFSA (2018)
Vitis candicans		N						EFSA (2018)
Vitis cinerea var. helleri x. V. vulpina		N						EFSA (2018)

Vitis girdiana		N							EFSA (2018)
Vitis labrusca									EPPO (n.d.)
Vitis rotundifolia		N							EFSA (2018)
Vitis sp.		N				*fastidiosa*			EFSA (2018)
Vitis vinifera	common grape vine	EN	E	E		*fastidiosa*		Pierce's disease (PD)	EC (2018), EFSA (2018)
Westringia fruticosa	coastal/Australian rosemary		N	N	*multiplex*		*pauca*		EC (2018), EFSA (2018)
Westringia glabra	violet westringia			N			*pauca*		EC (2018), EFSA (2018)
Xanthium strumarium		E	N						EFSA (2018)

Table B: Complete list of genomes used in the project so far. 46 Xylella fastidiosa genomes and two *Xanthomonas* genomes were obtained from NCBI's GenBank. Details on the genome size, sequencing information and origin are listed below.

Subspecies	Strain	Size (Mb)	Date added	Last updated	Submitted by	Host	Assembly ID	Assembly level	Assembly method	Genome coverage	Sequencing technology	Collection date	Location of origin	Plasmids
NA	9a5c	2.73175	02/06/2000	29/03/2017	Sao Paulo state (Brazil) Consortium	CVC-affected Valencia sweet orange	GCA_000006725.1	Complete Genome	NA	NA	NA	21/05/1992	Macaubal, Sao Paulo, Brazil	pXF1.3, pXF51
NA	BB01	2.72975	10/07/2002	11/04/2017	DOE Joint Genome Institute	*Vaccinium corymbosum* (blueberry)	GCA_000166855.2	Contig	ALLPATHS v. R37654	NA	Sanger	01/10/2016	Georgia, USA	NA
NA	Dixon	2.62233	10/07/2002	30/03/2017	DOE Joint Genome Institute	almond tree	GCA_000166835.1	Scaffold	NA	NA	NA	NA	NA	NA
NA	Temecula1	2.52115	29/01/2003	29/03/2017	Sao Paulo state (Brazil) Consortium	PD-affected *Vitis vinifera* (grapevine)	GCA_000007245.1	Complete Genome	NA	NA	NA	1998	Temecula, California, USA	pXFPD1.3
NA	M12	2.47513	19/02/2008	30/03/2017	US DOE Joint Genome Institute	ALSD-affected almond	GCA_000019325.1	Complete Genome	NA	NA	NA	2003	San Joaquin Valley, California, USA	NA
NA	M23	2.57399	11/04/2008	30/03/2017	US DOE Joint Genome Institute	ALSD-affected almond	GCA_000019765.1	Complete Genome	NA	NA	NA	2003	San Joaquin Valley, California, USA	pXFAS01
fastidiosa	GB514	2.51738	23/09/2010	11/04/2017	Research and Testing Laboratory	*Vitis vinifera* (grapevine)	GCA_000148405.1	Complete Genome	NA	NA	NA	NA	Texas, USA	unnamed

NA	EB92.1	2.47543	24/06/2011	22/11/2017	University of Florida	asymptomatic *Sambucus canadensis* (eldeberry; grapevine)	GCA_000219235.2	Contig	Newbler v. 2.3	194X	454 GS Titanium	1992	Leesburg, USA	NA
multiplex	ATCC 35871	2.41626	15/07/2013	01/04/2017	DOE Joint Genome Institute	hybrid plum	GCA_000428665.1	Scaffold	NA	NA	Illumina HiSeq 2000	NA	Georgia, USA	NA
multiplex	Griffin-1	2.38731	12/09/2013	11/04/2017	USDA	OLSD-affected *Quercus rubra* (red oak tree)	GCA_000466025.1	Contig	Newbler v. v2.6	30.0x	454	summer 2006	Griffin, Georgia, USA	NA
NA	32	2.60755	11/12/2013	02/04/2017	Universidade de Mogi das Cruzes	CLSD-affected coffee plants	GCA_000506405.1	Contig	GS de novo Assembler v. 2.5.3	70x	454	NA	Sao Paulo, Brazil	NA
NA	6c	2.60398	11/12/2013	06/04/2017	Universidade de Mogi das Cruzes	CLSD-affected coffee plants	GCA_000506905.2	Contig	Bowtie2 v. 2.2.9	900x	Illumina MiSeq	NA	Sao Paulo, Brazil	pXF6c
NA	Mul-MD	2.52055	10/02/2014	02/04/2017	FNPRU-USNA-ARS-USDA	leaf-scorch-affected mulberry plant	GCA_000567985.1	Contig	Newbler v. 08-06-2012	5.0x	454	2011	Beltsville, Maryland, USA	NA
sandyi	Ann-1	2.78091	06/06/2014	02/04/2017	University of California (LANL Genome	leaf-scorch-affected *Nerium oleander*	GCA_000698805.1	Complete Genome	Velvet v. 1.0.13	22.3X	454; Illumina	1993	Palm Springs, California, USA	unnamed1

					Science Group)									
(*morus*)	MUL0034	2.66658	06/06/2014	02/04/2017	University of California (LANL Genome Science Group)	leaf-scorch-affected mulberry plant	GCA_000698825.1	Complete Genome	Newbler v. 2.3; VELVET v. 0.7.63	NA	454; Illumina	NA	USA	unnamed2
NA	sycamore Sy-VA	2.47588	22/07/2014	02/04/2017	Beltsville Agricultural Research Center	leaf-scorch-affected sycamore tree	GCA_000732705.1	Contig	Newbler v. 2.7	70.0x	454	10/2002	Virginia, USA	NA
NA	ATCC 35879	2.52233	21/10/2014	02/04/2017	Crop Diseases, Pests, Genetics Research Unit, San Joaquin Valley Agricultural Sciences Center, USDA	*Vitis vinifera* (grapevine)	GCA_000767565.1	Contig	CLC Genomic Workbench v. 7.0.3	1380.0x	Illumina MiSeq	1987	Florida, USA	NA
NA	CoDiRO	2.54293	29/12/2014	03/04/2017	National Research Council (C.N.R.),	OQDS-affected olive trees	GCA_000811965.1	Contig	Velvet v. 1.2.08; SOAPdenovo v. 2.04;	345.0x	llumina HiSeq	NA	Apulia, Italy	unnamed

					Institute for Sustainable Plant Protection				EDENA v. 0.3; post-assembly SSPACE v. 1.0.7					
NA	CO33	2.68193	28/10/2015	04/04/2017	National Research Council (C.N.R.), Institute for Sustainable Plant Protection	CLSD-affected coffee plants	GCA_001417925.1	Contig	Velvet v. 1.2.8; SOAPdenovo v. 2.04; Edena v. 0.3; post-assembly SSPACE v. 1.0.7	310.0x	Illumina HiSeq	10/2014	imported from Costa Rica through Netherlands and to northern Italy	NA
NA	3124	2.74859	03/12/2015	28/06/2017	Universidade de Sao Paulo	CLSD-affected coffee plants	GCA_001456195.1	Complete Genome	Newbler v. 2.3; CROSSMATCH	267x	454 GS FLX Titanium	01/11/2009	Matao, Sao Paulo, Brazil	NA
NA	Fb7	2.69932	03/12/2015	22/05/2018	Universidade de Sao Paulo	citrus	GCA_001456335.3	Complete Genome	NA	NA	NA	01/11/2009	Corrientes, Argentina	unnamed
NA	Hib4	2.87755	03/12/2015	28/06/2017	Universidade de Sao Paulo	hibiscus	GCA_001456315.1	Complete Genome	Newbler v. 2.3; CROSSMATCH	100x	454 GS FLX Titanium	01/11/2009	Jarinu, Sao Paulo, Brazil	pXF64-HB
NA	J1a12	2.86724	03/12/2015	28/06/2017	Universidade de Sao Paulo	citrus	GCA_001456235.1	Complete Genome	Newbler v. 2.3; CROSSMATCH	65x	454 GS FLX Titanium	01/11/2009	Jales, Sao Paulo, Brazil	pXF27-J1, pXF51-J1

NA	Pr8x	2.70582	03/12/2015	28/06/2017	Universidade de Sao Paulo	plum	GCA_001456295.1	Complete Genome	Newbler v. 2.3; CROSSMATCH	63x	454 GS Titanium	01/11/2009	Jarinu, Sao Paulo, Brazil	pXF39
NA	U24D	2.73249	03/12/2015	28/06/2017	Universidade de Sao Paulo	*Citrus x sinensis*	GCA_001456275.1	Complete Genome	Newbler v. 2.3; CROSSMATCH	81x	454 GS FLX Titanium	01/11/2009	Ubarana, Sao Paulo, Brazil	pXF51ud
pauca	CFBP8072	2.49666	18/12/2015	04/04/2017	INRA	CLSD-affected *Coffea arabica*	GCA_001469345.1	Scaffold	Velvet v. 1.2.02	700.0x	Illumina HiSeq	21/05/2012	imported from Ecuador to France	NA
NA	CFBP8073	2.58215	18/12/2015	04/04/2017	INRA	*Coffea canephora*	GCA_001469395.1	Scaffold	Velvet v. 1.2.02; SOAPdenovo v. 1.05	800.0x	Illumina HiSeq	27/09/2012	France	NA
pauca	COF0324	2.77256	05/02/2016	04/04/2017	cBio Corp	CLSD-affected *Coffea*	GCA_001549815.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v. 1.1.1; Pilon v. 1.8; Quast v.	736.432x	Illumina MiSeq	2006	Varginha, Minas Gerais State, Brazil	pXF-BHR-COF0324, pXF-P1.COF0324, pXF-PC_COF0324, pXF-RC.COF0324

									2.2; Prokka v. 1.7					
pauca	COF0407	2.53847	05/02/2016	04/04/2017	cBio Corp	CLSD-affected *Coffea*	GCA_001549825.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v. 1.1.1; Pilon v. 1.8; Quast v. 2.2; Prokka v. 1.7	612.211x	Illumina MiSeq	06/2009	Curridabat, San Jose Province, Costa Rica	pXF-P1.OLS0479, pXF-P4.OLS0479, pXF-PS.OLS0479, pXF-RC.OLS0479
pauca	CVC0251	2.74025	05/02/2016	04/04/2017	cBio Corp	CVC-affected *Citrus sinensis*	GCA_001549765.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v.	944.475x	Illumina MiSeq	1999	Bebedouro, Sao Paulo State, Brazil	pXF-BHR.CVC0251, pXF-P1.CVC0251, pXF-P4.CVC0251, pXF-PS.CVC0251

									1.1.1; Pilon v. 1.8; Quast v. 2.2; Prokka v. 1.7					
pauca	CVC0256	2.70214	05/02/2016	04/04/2017	cBio Corp	CVC-affected *Citrus sinensis*	GCA_001549745.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v. 1.1.1; Pilon v. 1.8; Quast v. 2.2; Prokka v. 1.7	691.101x	Illumina MiSeq	1999	Colina, Sao Paulo State, Brazil	pXF-BHR.CVC0256, pXF-P1.CVC0256, pXF-P4.CVC0256, pXF-PS.CVC0256
pauca	OLS0478	2.55541	05/02/2016	04/04/2017	cBio Corp	leaf-scorch-affected *Nerium oleander*	GCA_001549755.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v.	788.469x	Illumina MiSeq	02/2011	Sabanilla, San Jose Province, Costa Rica	pXF-P1.OLS0478, pXF-P4.OLS0478

									1.1.1; Pilon v. 1.8; Quast v. 2.2; Prokka v. 1.7					
pauca	OLS0479	2.53996	05/02/2016	04/04/2017	cBio Corp	leaf-scorch-affected *Nerium oleander*	GCA_001549735.1	Contig	Trimmomatic v. 0.32; SGA v. 0.10.13; iMetAMOS v. 1.5; samtools v. 1.1; FastQC v. 0.10.0; Spades v. 3.1.1; idba v. 1.1.1; Pilon v. 1.8; Quast v. 2.2; Prokka v. 1.7	844.258x	Illumina MiSeq	02/2011	Sabanilla, San Jose Province, Costa Rica	pXF-P1.COF0407, pXF-P4.COF0407, pXF-PS.COF0407, pXF-RC.COF0407
fastidiosa	Stag's Leap	2.5108	24/02/2016	04/04/2017	USDA-ARS	PD-affected *Vitis vinifera* (grapevine)	GCA_001572105.1	Contig	Bowtie 2 v. 2.2.6	750.0x	Illumina MiSeq	NA	Napa Valley, California, USA	NA
pauca	11399	2.73606	13/07/2016	11/04/2017	IAC - Centro de citricultura	orange tree	GCA_001684415.1	Contig	CLC NGS Cell v. 6.0	70.0x	Illumina HiSeq	1996	Brazil	pXF51
sandyi	Ann-1	2.51152	25/11/2016	05/04/2017	USDA-ARS	poisonous evergreen	GCA_001886315.1	Scaffold	CLC Genomics	1271.0x	Illumina MiSeq	NA	USA	NA

						shrub (oleander)			Workbench v. 7.5					
NA	DSM 10026	2.43165	02/12/2016	06/04/2017	DOE Joint Genome Institute	NA	GCA_900129695.1	Scaffold	NA	416x	NA	NA	NA	NA
multiplex	CFBP8416	2.46675	25/01/2017	25/01/2017	INRA	*Polygala myrtifolia*	GCA_001971475.1	Contig	Velvet v. 1.2.07; SOAPdenovo v. 2.04	125.0x	Illumina MiSeq	2015	Propriano, Corse, France	NA
multiplex	CFBP8417	2.50498	25/01/2017	06/04/2017	INRA	leaf-scorch-affected *Spartium junceum*	GCA_001971505.1	Contig	Velvet v. 1.2.07; SOAPdenovo v. 2.04	125.0x	Illumina MiSeq	2015	Alata, Corse, France	NA
multiplex	CFBP8418	2.51397	25/01/2017	06/04/2017	INRA	leaf-scorch-affected *Spartium junceum*	GCA_001971465.1	Contig	Velvet v. 1.2.07; SOAPdenovo v. 2.04	125.0x	Illumina MiSeq	2015	Alata, Corse, France	NA
pauca	De Donno	2.54374	04/05/2017	10/05/2017	POnTE (Pest Organisms Threatening Europe)	OQDS-affected *Olea europaea*	GCA_002117875.1	Complete Genome	SPAdes v. 3.9.0	636.0x	PacBio; Illumina HiSeq	01/06/2014	Apulia, Italy	pXF-De_Donno
NA	Salento-1	2.54337	27/02/2018	04/03/2018	CNR	OQDS-affected *Olea europaea*	GCA_002954185.1	Complete Genome	HGAP v.2 + Circlator v. 1.2.1	402.7x	PacBio	2015	Taviano, Lecce, Apulia, Italy	pSal1

NA	Salento-2	2.54357	27/02/2018	04/03/2018	CNR	OQDS-affected *Olea europaea*	GCA_002954205.1	Complete Genome	HGAP v.2 + Circlator v. 1.2.1	349.25x	PacBio	2015	Ugento, Lecce, Apulia, Italy	pSal2
fastidiosa	IVIA5235	2.49157	10/09/2018	12/09/2018	Spanish National Research Council (CSIC), Institute for Sustainable Agriculture	leaf-scorch-affected *Prunus avium*	GCA_003515915.1	Contig	SPAdes v. 3.9.0	450.0x	Illumina HiSeq 4000	2016	Mallorca Island, Spain	pXFAS_5235
fastidiosa	XYL1732/17	2.444109	27/12/2018	04/01/2018	University of Balearic Islands	PD-affected *Vitis vinifera* (grapevine; white grape cultivar Paradella)	GCA_003973705.1	Contig	Newbler v. 2.9	102.0x	Illumina MiSeq	07/2017	Manacor, Mallorca, Spain	pXFAS01, pXFAS_5235
fastidiosa	XYL2055/17	2.45678	27/12/2018	04/01/2018	University of Balearic Islands	PD-affected *Vitis vinifera*	GCA_003973695.1	Contig	Newbler v. 2.9	151.0x	llumina HiSeq	08/2017	Manacor, Mallorca, Spain	pXFAS01, pXFAS_5235

Table C: Details of each strain displayed in the phylogenetic tree. This table lists details of the genomes from which a phylogenetic tree was created. Information includes GenBank accession number of each strain and location and host from which the isolate was obtained from.

Accession Number	Species	Strain	Continent	Country	Detailed location	Genus	Host	Common host name
AAAL02000032.1	*Xylella fastidiosa subsp. unknown*	Dixon	NA	NA	NA	*Prunus*	*Prunus dulcis*	almond
AAAM04000275.1	*Xylella fastidiosa subsp. sandyi*	Ann-1	North America	USA	Palm Springs, California	*Nerium*	*Nerium oleander*	oleander
AE003849.1	*Xylella fastidiosa subsp. unknown*	9a5c	South America	Brazil	Macaubal, Sao Paulo	*Citrus × sinensis*	*Citrus × sinensis pummelo x mandarin orange*	Valencia sweet orange
AE009442.1	*Xylella fastidiosa subsp. unknown*	Temecula1	North America	USA	Temecula, California	*Vitis*	*Vitis vinifera*	grapevine
AFDJ01000168.1	*Xylella fastidiosa subsp. unknown*	EB92.1	North America	USA	Leesburg	*Sambucus*	*Sambucus canadensis*	common elderberry
AVGA01000001.1	*Xylella fastidiosa subsp. multiplex*	Griffin-1	North America	USA	Griffin, Georgia	*Quercus*	*Quercus rubra*	red oak tree
AWYH01000001.1	*Xylella fastidiosa subsp. unknown*	32	South America	Brazil	Sao Paulo	*Coffea*	*Coffea*	coffee
AXDP01000001.1	*Xylella fastidiosa subsp. unknown*	Mul-MD	North America	USA	Beltsville, Maryland	*Morus*	*Morus*	mulberry
CM003178.1	*Xylella fastidiosa subsp. unknown*	CoDiRo	Europe	Italy	Apulia	*Olea*	*Olea europaea*	common olive
CM003743.1	*Xylella fastidiosa subsp. unknown*	OLS0479	North America	Costa Rica	Sabanilla, San Jose Province	*Nerium*	*Nerium oleander*	oleander
CM003748.1	*Xylella fastidiosa subsp. unknown*	CVC0256	South America	Brazil	Colina, Sao Paulo	*Citrus × sinensis*	*Citrus x sinensis*	sweet orange
CM003752.1	*Xylella fastidiosa subsp. unknown*	OLS0478	North America	Costa Rica	Sabanilla, San Jose Province	*Nerium*	*Nerium oleander*	oleander
CM003754.1	*Xylella fastidiosa subsp. unknown*	CVC0251	South America	Brazil	Bebedouro, Sao Paulo	*Citrus × sinensis*	*Citrus x sinensis*	sweet orange
CM003758.1	*Xylella fastidiosa subsp. unknown*	COF0324	South America	Brazil	Varginha, Minas Gerais	*Coffea*	*Coffea*	coffee
CM003762.1	*Xylella fastidiosa subsp. unknown*	COF0407	North America	Costa Rica	Curridabat, San Jose	*Coffea*	*Coffea*	coffee
CM004499.1	*Xylella fastidiosa subsp. pauca*	11399	South America	Brazil	NA	*Citrus × sinensis*	*Citrus x sinensis*	sweet orange
CM007617.1	*Xylella fastidiosa subsp. unknown*	6c	South America	Brazil	Sao Paulo	*Coffea*	*Coffea*	coffee plant
CM010656.1	*Xylella fastidiosa subsp. fastidiosa*	IVIA5235	Europe	Spain	Mallorca Island	*Prunus*	*Prunus avium*	sweet cherry
CP000941.1	*Xylella fastidiosa subsp. unknown*	M12	North America	USA	San Joaquin Valley, California	*Prunus*	*Prunus dulcis*	almond
CP001011.1	*Xylella fastidiosa subsp. unknown*	M23	North America	USA	San Joaquin Valley, California	*Prunus*	*Prunus dulcis*	almond
CP002165.1	*Xylella fastidiosa subsp. fastidiosa*	GB514	North America	USA	Texas	*Vitis*	*Vitis vinifera*	grapevine
CP006696.1	*Xylella fastidiosa subsp. sandyi*	Ann-1	North America	USA	NA	*Nerium*	*Nerium oleander*	oleander
CP006740.1	*Xylella fastidiosa subsp. unknown*	MUL0034	North America	USA	NA	*Morus*	*Morus*	mulberry
CP009790.1	*Xylella fastidiosa subsp. unknown*	U24D	South America	Brazil	Ubarana, Sao Paulo	*Citrus × sinensis*	*Citrus x sinensis*	sweet orange
CP009823.1	*Xylella fastidiosa subsp. unknown*	J1a12	South America	Brazil	Jales, Sao Paulo	*Citrus*	*Citrus*	citrus
CP009826.1	*Xylella fastidiosa subsp. unknown*	Pr8x	South America	Brazil	Jarinu, Sao Paulo	*Prunus*	*Prunus*	plum
CP009829.1	*Xylella fastidiosa subsp. unknown*	3124	South America	Brazil	Matao, Sao Paulo	*Coffea*	*Coffea*	coffee
CP009885.1	*Xylella fastidiosa subsp. unknown*	Hib4	South America	Brazil	Jarinu, Sao Paulo	*Hibiscus*	*Hibiscus*	hibiscus
CP010051.2	*Xylella fastidiosa subsp. unknown*	Fb7	South America	Argentina	Corrientes	*Citrus*	*Citrus*	citrus
CP016608.1	*Xylella fastidiosa subsp. unknown*	Salento-1	Europe	Italy	Taviano, Lecce, Apulia	*Olea*	*Olea europaea*	common olive
CP016610.1	*Xylella fastidiosa subsp. unknown*	Salento-2	Europe	Italy	Ugento, Lecce, Apulia	*Olea*	*Olea europaea*	common olive

Accession Number	Species	Strain	Continent	Country	Detailed location	Genus	Host	Common host name
CP020870.1	*Xylella fastidiosa subsp. pauca*	De Donno	Europe	Italy	Apulia	*Olea*	*Olea europaea*	common olive
FQWN01000063.1	*Xylella fastidiosa subsp. unknown*	DSM 10026	NA	NA	NA	*NA*	*NA*	NA
JMHP01000001.1	*Xylella fastidiosa subsp. unknown*	sycamore Sy-VA	North America	USA	Virginia	*Acer*	*Acer pseudoplatanus*	sycamore tree
JQAP01000001.1	*Xylella fastidiosa subsp. unknown*	ATCC 35879	North America	USA	Florida	*Vitis*	*Vitis vinifera*	grapevine
KE386775.1	*Xylella fastidiosa subsp. multiplex*	ATCC 35871	North America	USA	Georgia	*Prunus*	*Prunus*	hybrid plum
LJZW01000001.1	*Xylella fastidiosa subsp. unknown*	CO33	North America	Costa Rica	imported from Costa Rica through Netherlands and to northern Italy	*Coffea*	*Coffea*	coffee plant
LKDK01000001.1	*Xylella fastidiosa subsp. pauca*	CFBP8072	South America	Ecuador	imported from Ecuador to France	*Coffea*	*Coffea arabica*	Arabica coffee
LKES01000001.1	*Xylella fastidiosa subsp. unknown*	CFBP8073	Europe	France	NA	*Coffea*	*Coffea canephora*	Robusta coffee
LSMJ01000001.1	*Xylella fastidiosa subsp. fastidiosa*	Stag's Leap	North America	USA	Napa Valley, California	*Vitis*	*Vitis vinifera*	grapevine
LUYA01000001.1	*Xylella fastidiosa subsp. multiplex*	CFBP8418	Europe	France	Alata, Corse	*Spartium*	*Spartium junceum*	Spanish broom
LUYB01000001.1	*Xylella fastidiosa subsp. multiplex*	CFBP8417	Europe	France	Alata, Corse	*Spartium*	*Spartium junceum*	Spanish broom
LUYC01000001.1	*Xylella fastidiosa subsp. multiplex*	CFBP8416	Europe	France	Propriano, Corse	*Polygala*	*Polygala myrtifolia*	myrtle-leaf milkwort
MPAZ01000045.1	*Xylella fastidiosa subsp. unknown*	BB01	North America	USA	Georgia	*Vaccinium*	*Vaccinium corymbosum*	blueberry
NC_003902.1	*Xanthomonas*	Xanthomonas campestris	NA	NA	NA	*NA*	*NA*	NA
NC_010717.2	*Xanthomonas*	Xanthomonas oryzae	NA	NA	NA	*NA*	*NA*	NA
QTJS01000001.1	*Xylella fastidiosa subsp. fastidiosa*	XYL2055	Europe	Spain	Manacor, Mallorca	*Vitis*	*Vitis vinifera*	grapevine
QTJT01000001.1	*Xylella fastidiosa subsp. fastidiosa*	XYL1732	Europe	Spain	Manacor, Mallorca	*Vitis*	*Vitis vinifera*	grapevine (white grape cultivar Paradella)

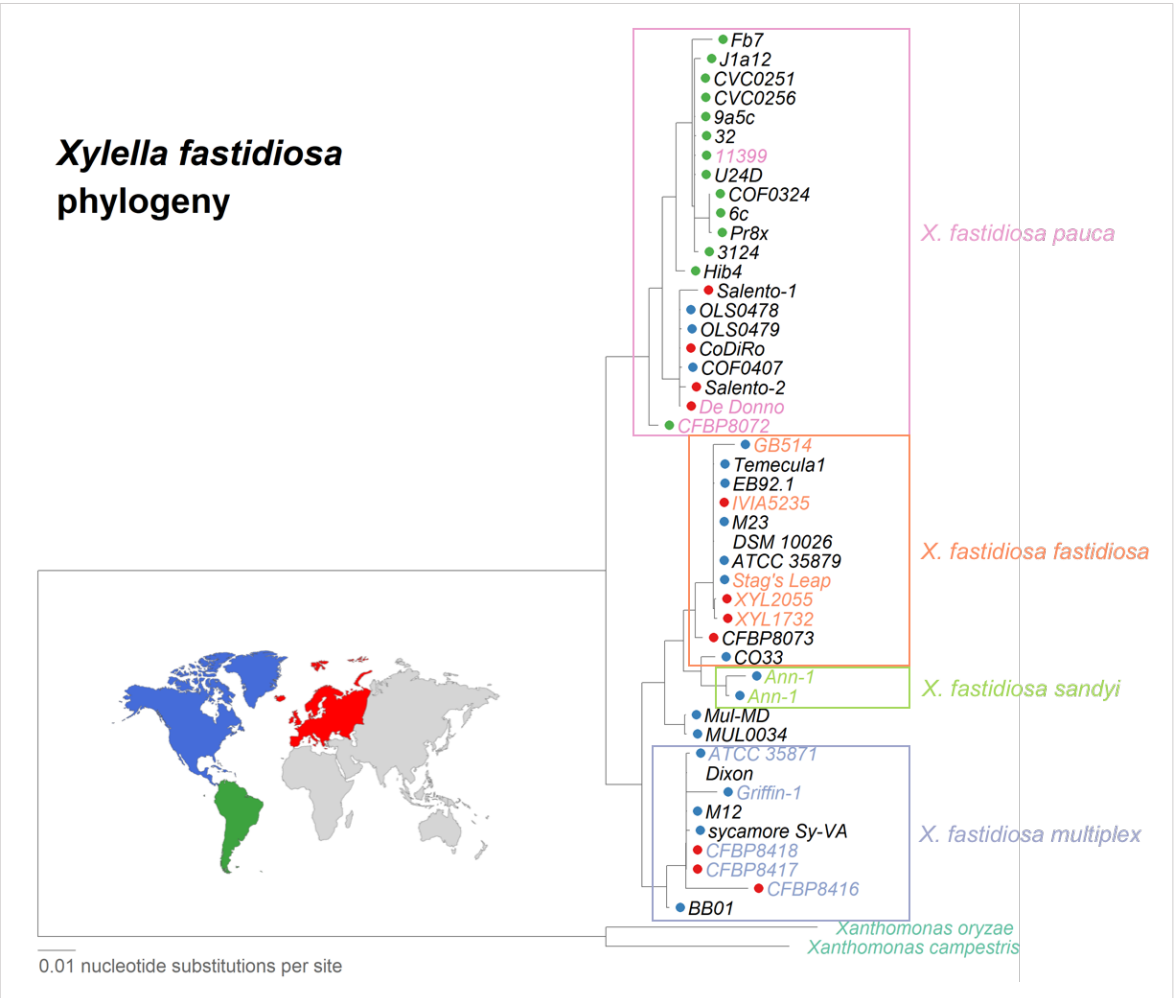

Figure A: First draft of a *Xylella fastidiosa* (*Xf*) phylogeny. A phylogenetic tree of 46 Xf and two *Xanthomonas* genomes (outgroups) was created. This tree was generated using FastTree's multiple sequence alignment by maximum-likelihood. The tree was visualised using the 'ape' package on R. Location of origin is highlighted by coloured circles corresponding to continents in the world map at the bottom left. The subspecies of strains with coloured fonts were confirmed by previous research. The subspecies of Xf strains Mul-MD and MUL0034 are less clear.

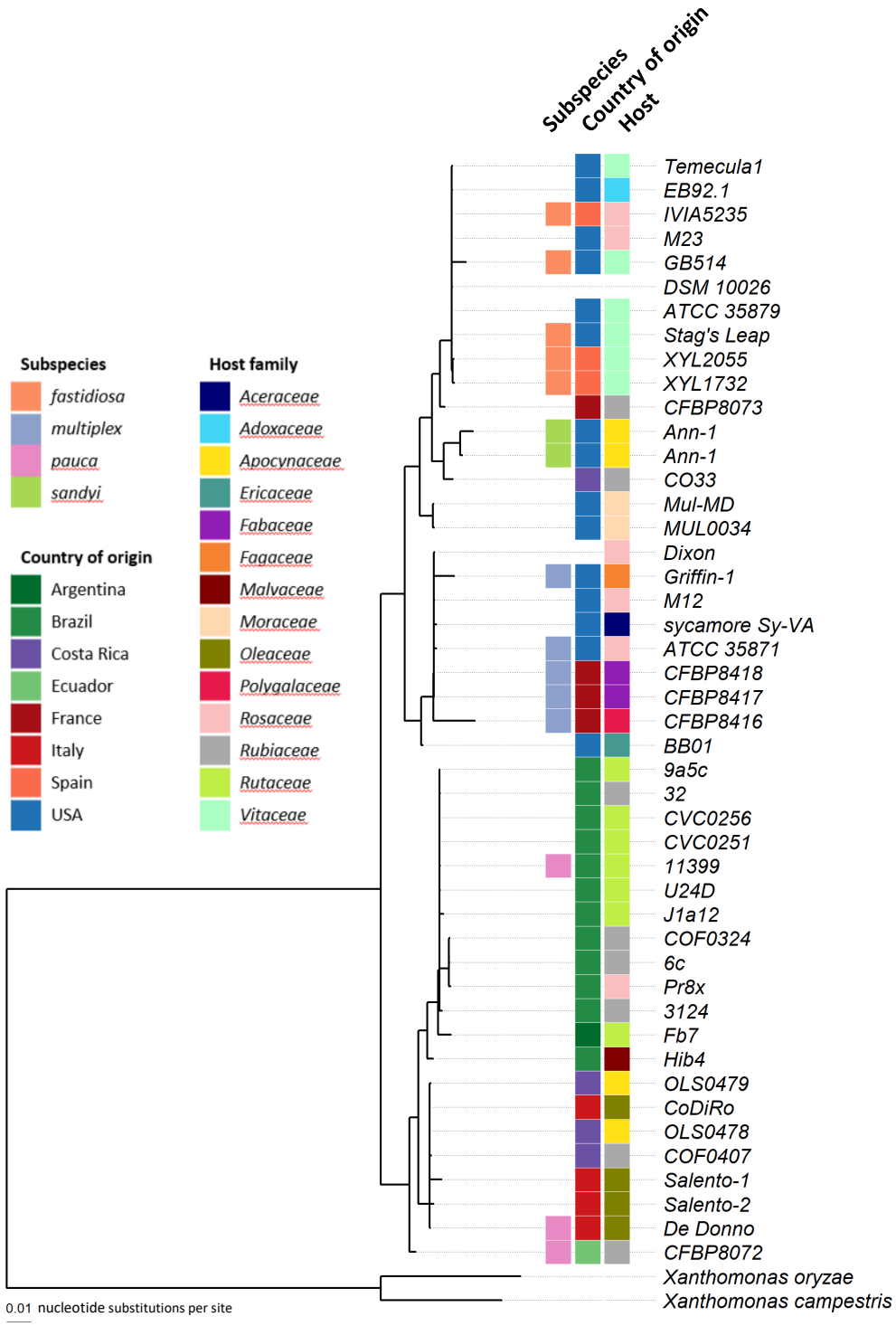

Figure B: Phylogeny of *Xylella fastidiosa* (*Xf*) whole-genome sequencing data and strain traits. A phylogenetic tree of 46 *Xf* and two *Xanthomonas* genomes (outgroups) was created. This tree was generated using FastTree's multiple sequence alignment by maximum-likelihood. The tree was visualised using the 'ggtree' package on R . Traits of each strain are depicted as a heatmap. Different colors represent different subspecies, country of origin and taxonomic family of the plant from which the strains were isolated from.

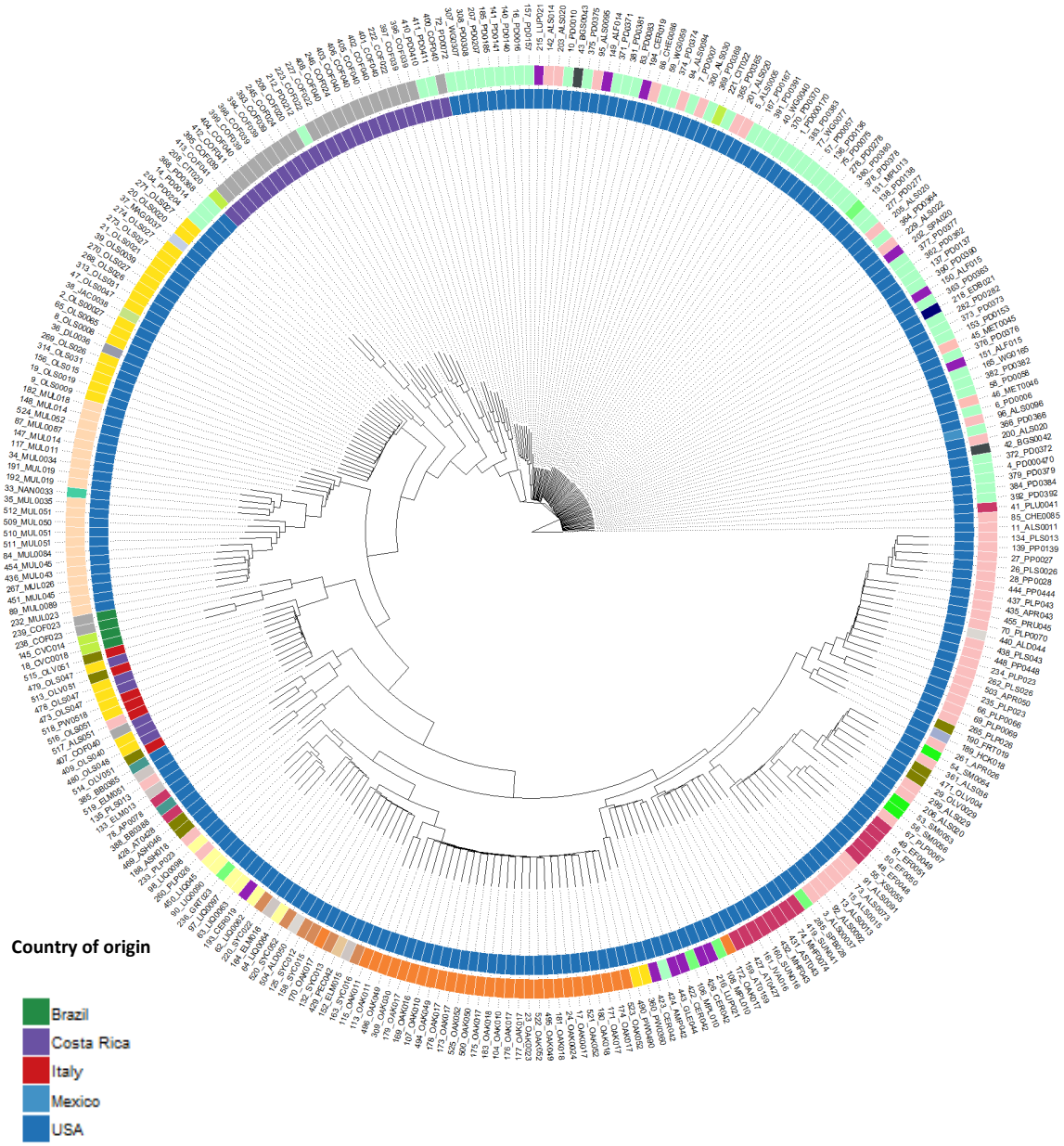

Figure D: Phylogeny of *Xylella fastidiosa* (*Xf*) multilocus sequencing type (MLST) data and strain traits. MLST data of 293 *Xf* isolates are available from the PubMLST database (https://pubmlst.org/xfastidiosa/). *Xf* MLST looks at secen different house-keeping genes: *leuA*, *petC*, *malF*, *cysG*, *holC*, *nuoL* and *gitT*. More details of each of these genes can be found in Appendix Table F. Concatenated nucleotide sequences of all 293 isolates were aligned using ClustalW's progressive alignment algorithm. A Newick tree was created using Phylip's consensus option (steps followed as per http://www.sfu.ca/~carmean/phylip1.html). The tree was then visualised using R's 'ggtree' package. Colours in the inner circle depict the country where each isolate was sampled from. Colours in the outer circle depict the taxonomic family (plant and insect) from which the strain was isolated from.

Table D: Details of collected leaf samples from Colombia.\ This is a list of all 51 plants collected in Colombia. Samples were collected in triplicates for each plant sample (see Figure E for a schematic). Several details were noted and measurements taken, including collection date and time, cultivar (Var) information if given, whether the plant had Xylella-like symptoms (S) or not (A), whether the plant was cultivated (C) or naturally occurring (N), location details, sea level in metres above median sea level (MAMSL), GPS coordinates in decimal degrees (DD; latitude, longitude), median aerial temperature in °C, humidity and notable observations.

ID	Date	Time	Family	Species	Var	Symptoms	Cultivation	Location description	Location	MAMSL	GPS (DD)	°C	Humidity	Notes
MALHR02001	20190625	15:00	Malvaceae	*Hibiscus rosa-sinensis*	N/A	A	C	Tulenapa research station	Urabá	30m	7.774001, -76.664901	29 C	0.88	
MALHR02002	20190625	15:15	Malvaceae	*Hibiscus rosa-sinensis*	N/A	S	C	Tulenapa research station	Urabá	30m	7.774192, -76.664902	29 C	0.88	
MALBCO4001	20190628	11:35	Malvaceae	*Theobroma cacao*	N/A	S	C	Farm	Sopetrán	521m	6.5377, -75.8318	23 C	0.57	leafhopper on tree
MALBC04002	20190628	11:45	Malvaceae	*Theobroma cacao*	N/A	S	C	Farm	Sopetrán	521m	6.5374, -75.8318	23 C	0.57	leafhopper on tree
UNKXX01001	20190625	09:51	N/A	*N/A*	N/A	A	N	Rainforest	Urabá	30m	7.7729, -76.6703	29 C	0.88	
RUBAP02001	20190625	15:40	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulenapa research station	Urabá	30m	7.775482, -76.665425	29 C	0.88	
RUBAP02002	20190625	15:40	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulenapa research station	Urabá	30m	7.775398, -76.665438	29 C	0.88	
RUBAP02003	20190625	15:40	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulenapa research station	Urabá	30m	7.775398, -76.665438	29 C	0.88	
RUBAP02004	20190626	06:30	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulanepa research station	Tulanepa	30m	7.773682, -76.654593	30 C	0.74	
RUBAP02005	20190626	06:30	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulanepa research station	Tulanepa	30m	7.775513, -76.665425	30 C	0.74	
RUBAP02006	20190626	06:30	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulanepa research station	Tulanepa	30m	7.773980, -76.656314	30 C	0.74	

RUBAP02007	20190626	06:30	Rubiaceae	*Alibertia patinoi*	N/A	A	C	Tulanepa research station	Tulanepa	30m	7.773708, -76.654650	30 C	0.74
RUBCA03001	20190627	15:25	Rubiaceae	*Coffea arabica*	Geisha	A	C	Farm	Fredonia	1423m	5.970375, -75.670041	24 C	0.59
RUBCA03002	20190627	15:30	Rubiaceae	*Coffea arabica*	Geisha	A	C	Farm	Fredonia	1423m	5.9703, -75.6701	24 C	0.59
RUBCA03003	20190627	15:45	Rubiaceae	*Coffea arabica*	Geisha	S	C	Farm	Fredonia	1423m	5.9704,-75.6704	24 C	0.59
RUBCA03004	20190627	15:55	Rubiaceae	*Coffea arabica*	Colombia	S	C	Farm	Fredonia	1423m	5.9730, -75.6701	24 C	0.59
RUBCA03005	20190627	16:07	Rubiaceae	*Coffea arabica*	Colombia	A	C	Farm	Fredonia	1423m	5.9730, -75.6700	24 C	0.59
RUBCA03006	20190627	16:12	Rubiaceae	*Coffea arabica*	Colombia	S	C	Farm	Fredonia	1423m	5.9730, -75.6701	24 C	0.59
RUBCA03007	20190627	16:42	Rubiaceae	*Coffea arabica*	Caturra	S	C	Farm	Fredonia	1786m	5.99748, -75.6644	24 C	0.59
RUBCA03008	20190627	16:46	Rubiaceae	*Coffea arabica*	Caturra	S	C	Farm	Fredonia	1786m	5.9749, -75.6643	24 C	0.59
RUBCA03009	20190627	16:49	Rubiaceae	*Coffea arabica*	Caturra	S	C	Farm	Fredonia	1786m	5.9748, -75.6642	24 C	0.59
RUBCA03010	20190627	16:54	Rubiaceae	*Coffea arabica*	Pajarito	S	C	Farm	Fredonia	1786m	5.9748, -75.6644	24 C	0.59
RUBCA03011	20190627	16:59	Rubiaceae	*Coffea arabica*	Pajarito	S	C	Farm	Fredonia	1786m	5.9747, -75.6644	24 C	0.59
RUBCA03012	20190627	17:07	Rubiaceae	*Coffea arabica*	Pajarito	S	C	Farm	Fredonia	1786m	5.9746, -75.6643	24 C	0.59
RUBCA03013	20190627	17:10	Rubiaceae	*Coffea arabica*	Castillo	S	C	Farm	Fredonia	1786m	5.9748, -75.6645	24 C	0.59

RUBCA03014	20190627	17:13	Rubiaceae	*Coffea arabica*	Castillo	S	C	Farm	Fredonia	1786m	5.9749, -75.6645	24 C	0.59	
RUBCA03015	20190627	17:20	Rubiaceae	*Coffea arabica*	Castillo	S	C	Farm	Fredonia	1786m	5.9740, -75.6645	24 C	0.59	
RUBCA05001	20190703	11:15	Rubiaceae	*Coffea arabica*	N/A	S	C	EAFIT Campus	Medellín	1504m	6.2002, -75.5785	23 C	0.64	rust
RUBCA05002	20190703	11:25	Rubiaceae	*Coffea arabica*	N/A	S	C	EAFIT Campus	Medellín	1504m	6.2001, -75.5785	23 C	0.64	rust
RUBTX06001	20190704	14:30	Rubiaceae	*Tocoyena*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2693, -75.5631	28 C	0.51	rust
RUBGA06001	20190704	14:45	Rubiaceae	*Genipa americana*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2698, -75.5625	28 C	0.51	
RUBPL06001	20190704	14:55	Rubiaceae	*Posoqueria latifolia*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2699, -75.5626	28 C	0.51	
RUBPX06001	20190704	15:00	Rubiaceae	*Pogonopus*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2700, -75.5625	28 C	0.51	
2251	20190704	15:15	Rubiaceae	*Cosmibuena grandiflora*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2713, -75.5626	28 C	0.51	
RUBHP06001	20190704	15:20	Rubiaceae	*Hamelia patens*	N/A	A	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2705, -75.5622	28 C	0.51	
RUBHP06002	20190704	15:30	Rubiaceae	*Hamelia patens*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2706, -75.5623	28 C	0.51	
RUBIJ06001	20190704	15:35	Rubiaceae	*Ixora javanica*	N/A	A	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2708, -75.5623	28 C	0.51	
RUBIH06001	20190704	15:55	Rubiaceae	*Isertia haenkeana*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2723, -75.5642	28 C	0.51	
RUTCL02001	20190625	15:55	Rutaceae	*Citrus lemón*	N/A	S	C	Tulenapa research station	Urabá	30m	7.773901, -76.664054	29 C	0.88	

RUTCH05001	20190703	10:30	Rutaceae	*Citrus hystrix*	N/A	S	C	EAFIT Campus	Medellín	1504m	6.2001, -75.5783	23 C	0.64	
RUTCH06001	20190704	15:40	Rutaceae	*Citrus hystrix*	N/A	S	C	Botanic gardens	Botanic Gardens, Medellín	1474m	6.2699, -75.5629	28 C	0.51	
RUTCS07001	20190705	09:10	Rutaceae	*Citrus sinensis*	Valencia	S	C	Farm	La Pintada	729m	5.8284, -75.6082	24 C	0.76	CVC
RUTCS07002	20190705	09:15	Rutaceae	*Citrus sinensis*	Valencia	S	C	Farm	La Pintada	729m	5.8284, -75.6082	24 C	0.76	CVC
RUTCS07003	20190705	09:20	Rutaceae	*Citrus sinensis*	Valencia	S	C	Farm	La Pintada	729m	5.8283, -75.6082	24 C	0.76	CVC
RUTCS07004	20190705	09:40	Rutaceae	*Citrus sinensis*	Salustiana	S	C	Farm	La Pintada	696m	5.8268, -75.6123	24 C	0.76	CVC
RUTCS07005	20190705	09:45	Rutaceae	*Citrus sinensis*	Salustiana	S	C	Farm	La Pintada	696m	5.8269, -75.6124	24 C	0.76	CVC
RUTCS07006	20190705	09:50	Rutaceae	*Citrus sinensis*	Salustiana	S	C	Farm	La Pintada	696m	5.8267, -75.6124	24 C	0.76	CVC
RUTCL07001	20190705	10:40	Rutaceae	*Citrus lemón*	Tahiti	A	C	Farm	La Pintada	774m	5.8235, -75.6076	24 C	0.76	
RUTCL07002	20190705	10:45	Rutaceae	*Citrus lemón*	Tahiti	S	C	Farm	La Pintada	774m	5.8235, -75.6075	24 C	0.76	smaller fruits, lighter leaves
RUTCL07003	20190705	10:55	Rutaceae	*Citrus lemón*	Tahiti	A	C	Farm	La Pintada	774m	5.8235, -75.6072	24 C	0.76	
RUTCL07004	20190706	11:55	Rutaceae	*Citrus lemón*	Tahiti	A	C	Farm	La Pintada	774m	5.8236, -75.6071	24 C	0.76	

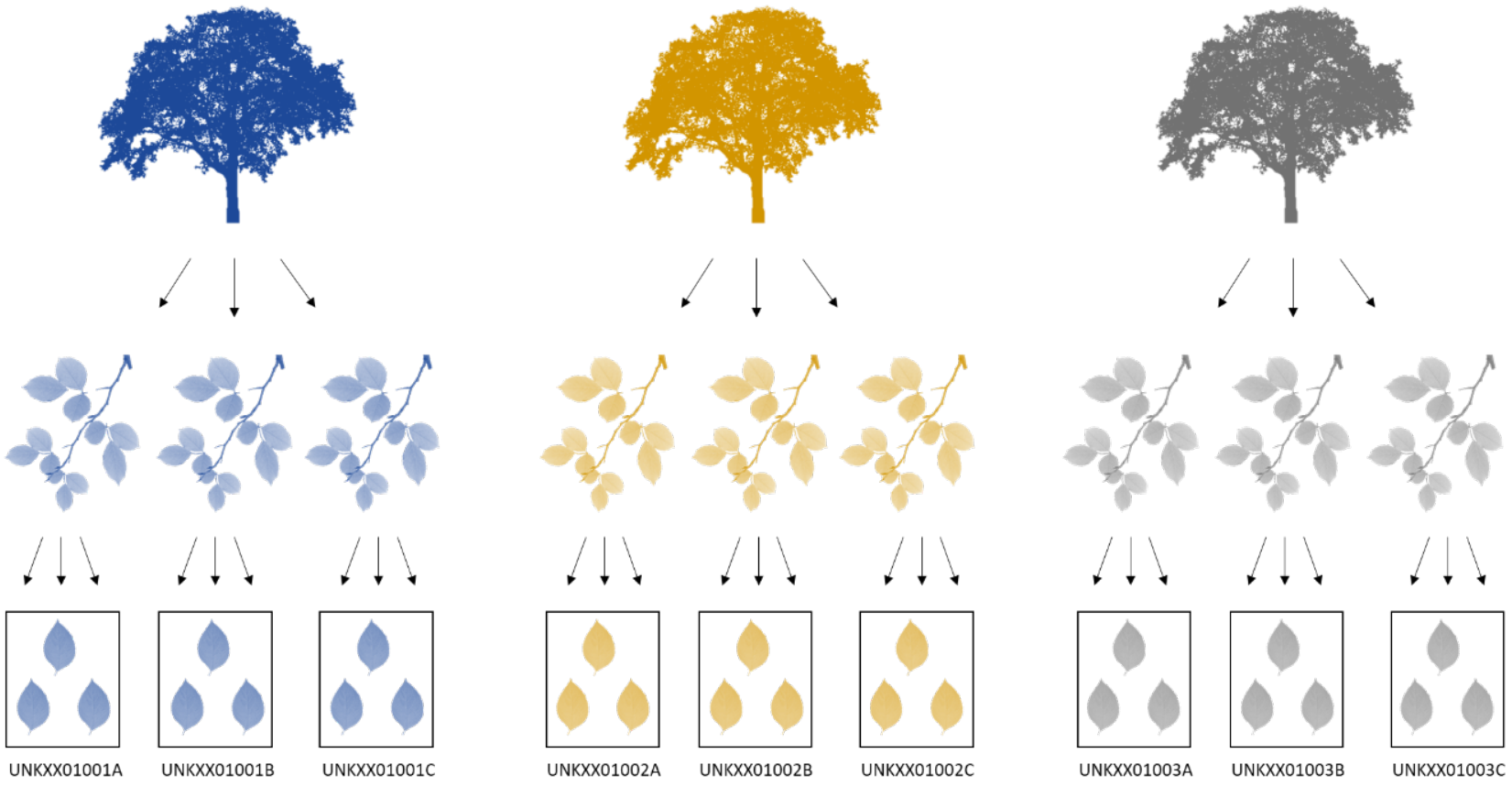

Figure E: A schematic of sampling leaves in Colombia. A total of 51samples were collected in the duration of a two-week survey in Colombia. Sampling was performed as follows: whenever possible, three plants of each species at each location were sampled. Per plant, three branches were selected, from which three leaves were removed using scissors disinfected in 70% ethanol and placed into a polyethylene bag. This would ultimately result in having triplicates of each plant originally sampled. Each sample was given a unique eleven-digit ID comprising of the first three letters of the plant family, the first letter of the genus, the first letter of the species, two digits indicating the location, three digits indicating the sample number, and a letter indicating the replicate (A, B or C).

RUBCA03001

RUBCA03002

RUBCA03003

RUBCA03005

RUBCA03006

RUBCA03006

RUBCA03007

RUBCA03010

RUBCA03008

RUBCA03008

RUBCA03011

RUBCA03013

RUBCA03012

RUBCA03012

RUBCA03015

RUBCA03015

RUBCA05001

RUBCA05001

Figure F: Photographs of Colombian samples that resulted positive for *Xylella fastidiosa* (*Xf*). *Xf* was identified by PCR in samples collected from the plants photographed above. See Appendix Table D: Details of collected leaf samples from Colombia.Table D for full details of each sample. Samples RUBCA03003, -006, -007, -008, -010, -011, -012, -013 and -015 had *Xf*-like symptoms, though difficult to see in some photographs.

1. Turn on water bath at 65°C
2. Clean surface of leaves to be used with ethanol
3. Place 0.5-1.0g of fresh small pieces of midribs, petioles, leaf basal part or twigs (1/4 of amount if lyophilised) into suitable tubes and immediately freeze dry in liquid nitrogen
4. Homogenise leaves using liquid nitrogen in pestle and mortar, or a tissue grinder
5. Add 5ml of CTAB buffer per 0.5-1.0g sample tube
6. Transfer sample and CTAB mix to 15ml falcon tubes and mix well
7. Heat at 65°C for 30min
8. Centrifuge at 16,000g, RT for 5min
9. Transfer 1ml aliquots of supernatant to fresh 2ml microcentrifuge tube (do not transfer any plant debris!)
10. Add 5µl of RNAse A (10mg/ml)
11. Incubate at 37°C overnight
12. Add 1ml of chloroform (isoamyl alcohol [24:1])
13. Mix well by shaking
14. Centrifuge at 16,000g for 10min
15. Transfer 700µl supernatant to new 1.5ml microcentrifuge tube
16. Add 490µl (or ~0.7 of available supernatant volume) of 2-propanol (room temperature)
17. Mix by inverting twice
18. Incubate at RT for 20min
19. Centrifuge at 16,000g, 4°C for 20min (recovery of pellet)
20. Remove supernatant
21. Wash pellet with 1ml of cold 70% ethanol
22. Centrifuge at 16,000g, 4°C for 10min
23. Remove supernatant and wash pellet again in 500µl of RT 70% ethanol
24. Mix by inversion
25. Centrifuge at 16,000g, 4°C for 10min
26. Remove supernatant and wash pellet again in 500µl of RT 70% ethanol
27. Mix by inversion
28. Centrifuge at 16,000g, 4°C for 10min
29. Remove supernatant and air-dry (~20min)
30. Re-suspend pellet in 100-150µl of TE buffer

Figure G: CTAB-based DNA extraction protocol. This is the modified protocol for the total DNA extraction from leaf samples. The original protocol was designed by EPPO (2016). The original protocol was modified as follows: homogenisation of leaf tissue was performed using pestle and mortar, and liquid nitrogen instead of a mechanical homogeniser; an overnight RNAse step was included to degrade unwanted RNA in the sample; room temperature 2-propanol was used for precipitation of DNA instead of cold 2-propanol to reduce the amount of salts being co-precipitated; finally, each sample was washed three times with 70% alcohol to ensure the removal of all contaminants.

Table E: Primer sequences used in this project. This is a list of all primer sequences used in this research, the target sequence and PCR conditions for each reaction. No PCR conditions are available for MLST primers as these have not been performed yet.

Primer name	Amplicon size (bp)	Forward primer sequence	Reverse primer sequence	Target sequence	Reference	PCR conditions (with redTaq polymerase)
27F / 1492R	~1,500	AGAGTTTGATCCTGGCTCAG	CTACGGCTACCTTGTTACGA	Bacteria-specific; 16S rRNA	Muyzer, De Waal and Uitterlinden, 1993	95°C 60s 95°C 30s, 51°C 30s, 72°C 120s (35x) 72°C 5min
RST31 / RST33	733	GCGTTAATTTTCGAAGTGATTCGATTGC	CACCATTCGTATCCCGGTG	*Xylella*-specific; 3' end of the gene *rpoD*, coding for an RNA polymerase sigma-70 factor	Minsavage *et al*., 1994	95°C 60s 95°C 30s, 57.9°C 30s, 72°C 120s (40) 72°C 5min
16S-23SF / 16S-23SR	650	GATGACTGGGGTGAAGTCGT	GACACTTTTCGCAGGCTACC	*Xylella*-specific; 16S-23S intergenic spacer	Martinati *et al*., 2005	95°C 60s 95°C 30s, 57°C 30s, 72°C 120s (40x) 72°C 5min
***Xylella*-specific primers for multi-locus sequence typing (MLST)**						
leuA-F / leuA-R	708	GGTGCACGCCAAATCGAATG	GTATCGTTGTGGCGTACACTG	*leuA*, coding for 2-isopropylmalate synthase	Yuan *et al*., 2010	
petC-F / petC-R	533	GCTGCCATTCGTTGAAGTACCT	GCACGTCCTCCCAATAAGCCT	*petC*, coding for ubiquinol cytochrome c oxidoreductase C1 subunit	Yuan *et al*., 2010	
malF-F / malF-R	730	TTGCTGGTCCTGCGGTGTTG	GACAGCAGAAGCACGTCCCAGAT	*malfF*, coding for ABC transporter sugar permease	Yuan *et al*., 2010	
cysG-F / cysG-R	600	GCCGAAGCAGTGCTGGAAG	GCCATTTTCGATCAGTGCAAAAG	*cysG*, coding for sirohaem synthase	Yuan *et al*., 2010	
holC-F / holC-R	379	ATGGCACGCGCCGACTTCT	ATGTCGTGTTTGTTCATGTGCAGG	*holC*, coding for DNA polymerase III holoenzyme chi subunit	Yuan *et al*., 2010	
nuoL-F / nuoL-R	557	TAGCGACTTACGGTTACTGGGC	ACCACCGATCCACAACGCAT	*nuoL*, coding for NADH ubiquinone oxidoreductase NQO12 subunit	Yuan *et al*., 2010	
gltT-F / gltT-R	654	TCATGATCCAAATCACTCGCTT	ACTGGACGCTGCCTCGTAAACC	*gltT*, coding for glutamate symport protein	Yuan *et al*., 2010	

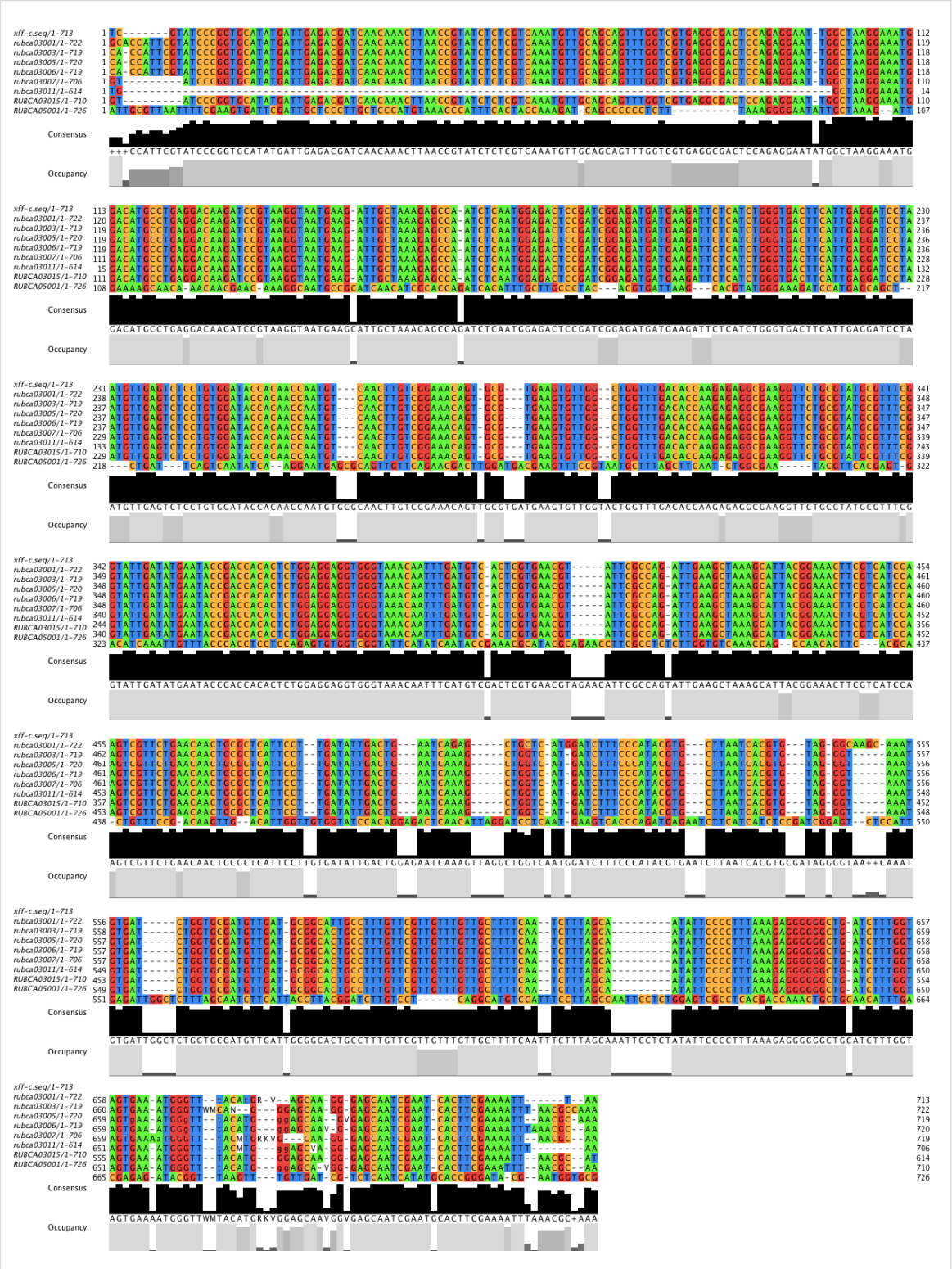

Figure H: Alignment of positive Colombian XF1 PCR amplicons. RUBCA03001, RUBCA03003, RUBCA03005, RUBCA03006, RUBCA03007, RUBCA03011, RUBCA03015 and RUBCA05001 were sent for Sanger sequencing using the Eurofins GATC LightRun service. Consensus sequences acquired using DNASTAR's Sanger Sequence Assembly and the *rpoD* gene sequence of *Xylella fastidiosa* (*Xf*) subsp. *fastidiosa* strain 9a5c, the *Xf* reference genome. Once consensus sequences were obtained, multiple sequence alignment by progressive strategy was performed using the program T-Coffee (Notredame, Higgins and Heringa, 2000). The alignment was finally visualised using JalView (Waterhouse, et al., 2009).